Institutions, Repression and the Spread of Protest*

Mehdi Shadmehr† Raphael Boleslavsky‡

Abstract

We analyze the strategic interactions between a state that decides whether to repress a group of activists and the general public that decides whether to protest following repression. Strategic complementarities between the strategies of the public and the state generate multiple equilibria, suggesting a role for social norms. These results shed light on conflicting empirical findings regarding the determinants of repression. We investigate the effects of exogenous restrictions, imposed by international institutions, showing that weak restrictions can paradoxically increase repression. This result provides a rationale for the puzzling empirical finding that international pressure can increase repression. Finally, we study the effects of endogenous restrictions, imposed by domestic institutions set up by the state to restrict its own subsequent repression. We characterize when the state can benefit from introducing such institutions, offering an explanation for the presence of partially independent judiciaries in authoritarian regimes.

JEL Codes: D74, D83, F5.

Keywords: Repression, Protest, Legitimate Coercion, Social Norms, Human Rights Treaties, Judicial Independence.

---

*We wish to thank Dan Bernhardt, Brandice Canes-Wrone, Odilon Câmara, Charles Cameron, Christopher Cotton, Justin Fox, Camilo Garcia-Jimeno, Soroush Ghazi, Adam Meirowitz, Steve Morris, Michael Peress, Nicola Persico, Andrea Prat, Carlo Prato, Kristopher Ramsay, Luis Rayo, Kevin Reffett, Kai Steverson, Milan Svolik, Scott Tyson, Léonard Wantchékon, and participants and discussants at APSA, MPSA, SPSA, and Miami Economic Theory conferences, and seminar audiences at the University of Utah, NYU-Abu Dhabi, and Princeton.

†Department of Economics, University of Miami, 5250 University Dr., Coral Gables, FL 33146. E-mail: `shad@miami.edu`

‡Department of Economics, University of Miami, 5250 University Dr., Coral Gables, FL 33146. E-mail: `r.boleslavsky@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 spreads 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 the public sometimes joins the protest following repression. This paper studies the subtle interactions between the state's incentive to repress activists and the public's incentive to join the protest, exploring the fundamental role of institutions, both domestic and foreign, in shaping the interaction between the state and the general public. What factors influence the likelihood that protest spreads following repression, and how does this potential for the spread of protest affect the state's response to dissent? How do the efforts of international institutions to protect legitimate dissidents affect the public's incentives to protest? Do authoritarian regimes benefit from setting up institutions that restrict their repression of legitimate dissent?

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.[4] This often-ignored uncertainty is

[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. In Ukraine, following the repression of protesters on November 30, 2013, the scale of protest expanded dramatically, culminating in the ouster of President Yanukovich. Notably, in a survey of over a thousand protesters in Independence Square, conducted on December 7-8, about 70% of respondents chose repression of November 30 as a main reason for joining the protest (Survey conducted by the Kiev International Institute of Sociology and the Ilko Kucheriv Democratic Initiative Foundation. Available at: http://www.kiis.com.ua/?lang=eng&cat=reports&id=216&page=11). Similarly, in a survey of protesters in Gezi Park protest in Turkey between June 6-8, 2013, about half of over 4400 respondents indicated that they joined the protests after observing government's repression (KONDA 2014, p. 20).

[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 contender's cost of collective action" (p. 100).

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

[4] In a survey of Turkish citizens conducted on July 6-7, 2013, with over 2600 participants from 28 provinces, over

at the core of the interaction between the state and citizens in this paper.

This paper develops a model of strategic interactions between a government that must decide whether to repress a group of activists or concede to their demands, 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. 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.

The bystander citizen's uncertainty about the types of the government and activists underlies the citizen's fundamental tradeoff: by supporting activists and toppling the government, he risks implementing harmful social changes, but by supporting the government's decision to repress activists, he risks blocking beneficial reforms. In turn, the potential for the bystander's protest creates an endogenous cost of repression that drives the government's fundamental tradeoff: by repressing the activists, the government prevents costly social changes, but it risks public protests that would topple it. However, by conceding to the activists' demands, the government pays a cost, but it avoids public protests. These two fundamental tradeoffs are linked through the bystander's posterior beliefs about the types of the government and the activists upon observing repression. Because good and bad governments have different incentives to repress different types of activists, observing repression is informative about both the type of the government and the type of the activists.

We show that the bystander's Bayesian updating—the formation of public opinion—has two key features. First, repression is bad news about *both* the government and the activists: upon observing repression, the bystander citizen updates negatively about both the government and the activists. Second, when the good government represses bad activists more often, the bystander up-

55% of the respondents believed that "the [Gezi Park] protests are a plot set up by foreign conspirators which resent Turkey's development" is a "right" or "absolutely right" statement, while about 27% believed that it is "wrong" or "absolutely wrong" (KONDA 2014, p. 59).

dates more negatively about the activists and less negatively about the government.[5] This reduces the public's incentives to join the activists' protest, lowering the (endogenous) cost of repression for the government, thereby raising the good government's incentive to repress bad activists. The strategic complementarity between the public's incentive to protest and the good government's incentive to concede 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.

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.[6] 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 provides 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.[7]

In our initial analysis, the only force that prevents bad governments from repressing activists is the endogenous threat of protest by the public. However, governments often face additional restrictions on the repression of legitimate activists. Some of these restrictions are imposed by the international community and are outside of the government's control. Foreign governments and international institutions can sometimes prevent the repression of legitimate dissidents by using a variety of instruments such as suspending aid, terminating trade benefits, making exclusive memberships contingent on observation of human rights, or threatening military intervention (Simmons 2009; Hafner-Burton 2013; Magesan 2013). Other restrictions are self-imposed: the government may *choose* to delegate repression decisions to security agencies or grant a degree of judicial independence, which serve as commitment devices that allow the government to credibly restrict its subsequent repression choices. We next investigate the effects of these exogenous restrictions (generated by international pressure) and endogenous restrictions (generated by domestic institutions) on the interactions between the government and the public.

Even though international pressure sometimes directly blocks the repression of good activists, we show that the overall likelihood that good activists are repressed may *increase*. This result is

[5]In contrast, when the bad government is more likely to repress good activists, the opposite updating obtains.

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

[7]As the bystander's status quo payoff under the bad ruler falls, multiple equilibria arise in a larger area of the parameter space—see the discussion of corollaries 1 and 2.

driven by the effect of international pressure on the bystander's updating: because international pressure sometimes prevents the bad government from repressing good activists, the bystander citizen updates less negatively about the government and more negatively about the activists upon observing repression. By shifting the bystander's beliefs in favor of the government, international pressure reduces the bystander's incentive to protest, thereby leading the bad government to *attempt* repression of good activists more often. We characterize when this strategic effect dominates, so that international pressure has the perverse effect of increasing the overall likelihood that good activists are repressed. This result provides a rationale for the puzzling finding in the empirical literature that international pressure and human rights treaties are sometimes associated with higher repression.[8] Moreover, we show that even when the overall likelihood that good activists are repressed falls, the likelihood that good reforms are implemented can also fall: the international community faces a tradeoff between protecting legitimate dissidents and promoting good reforms. These results generate a sharp and simple policy implication: in order to protect good activists, international pressure must be sufficiently strong; otherwise, at best it achieves nothing, and at worst it generates the *opposite* of its intended effect.

The analysis of the endogenous restrictions on repression is more subtle. When a bad government can observably commit to a repression strategy, it can gain by manipulating both the public's beliefs and the good government's equilibrium behavior. In the text, we discuss how the absence of transparency in authoritarian regimes enables the bad faction of the regime to commit to a repression strategy without revealing its type. The consequence is that when the bystander citizen observes repression, he cannot distinguish whether a good or a bad faction has repressed the activists, but he knows the repression strategy of the bad faction. Knowing that the bad faction is committed not to raise repression above a particular level, the bystander is less inclined to protest. This raises the good faction's incentive to repress bad activists, which reduces the bystander's incentive to protest even more, and so on. In effect, the bad faction's ability to limit its repression enables it to exploit the strategic complementarity between the strategies of the good faction and the bystander, thereby reducing the likelihood of the bystander's protest and the endogenous costs of repression.

Paradoxically, the regime's commitment to limit the repression of good activists benefits the bad faction. In particular, we identify conditions under which commitment power enables the bad faction to raise repression and yet face a lower likelihood of public protest in equilibrium, or to maintain the same level of repression that it would without those restrictions, but reduce the risk

[8] As Hafner-Burton (2012) discusses, the empirics seem to be robust to the endogeneity problem that international pressure is applied when repression is harsh. See also Hollyer and Rosendorff (2011) and our literature review.

of public protest. Commitment power is crucial: without it, the bad faction would deviate by increasing repression, which would raise the bystander's incentives to protest, thereby upsetting the equilibrium. These results provide a rationale for why authoritarian regimes sometimes establish partially independent judiciaries that restrict the state's use of repression.

We next discuss the literature. Section 2 presents the model and the discussion of its assumptions, and Section 3 provides some preliminary results. Section 4 analyzes the model without restrictions on the ruler's strategy. Section 5 incorporates exogenous restrictions on the ruler's ability to repress good activists (i.e. international pressure). Section 6 analyzes the game with commitment by the bad ruler (i.e. domestic institutions). A conclusion follows. Proofs are in the Online Appendix I.

1.1 LITERATURE

Our paper contributes to a number of literatures. First, there is a large political economy literature on protests and revolutions.[9] The literature has focused on the deterrence effect of repression, modeling repression as state actions that either directly reduce the likelihood of revolution,[10] or indirectly reduce it by raising the costs of dissent.[11] However, empirical evidence shows that repression can be followed by a spread of protest that further destabilizes the regime. Our model integrates this empirical finding by allowing the public to join the activists' protest following repression, thereby endogenizing the costs of repression to the state. Moreover, we make the observation that the public is often uncertain about the nature of the activists' demands and the intentions of the state. This often-ignored uncertainty, embedded in our setting, underlies the informational content of repression. These features, in turn, reveal the complementarities between the public's incentive to protest and the good government's incentive to concede. While the literature has extensively studied complementarities that arise due to coordination aspects of revolution, our paper reveals a new source of strategic complementarities that is important in the study of revolutions. These comple-

[9] 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; Tyson and Smith 2014), the role of vanguards (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; Guriev and Treisman 2015), the effect of elections (Little 2012; Egorov and Sonin 2015), contagion (Chen and Suen 2016), determinants of extremism (Shadmehr 2015), and middle class activism (Chen and Suen 2015).

[10] Examples include Acemoglu and Robinson (2000, 2001, 2006), Boix (2003), Besley and Persson (2011), Conrad and Ritter (2013), Svolik (2013), and Shadmehr (2014).

[11] Examples include Lichbach (1987), Persson and Tabellini (2009), Bueno de Mesquita (2010), Shadmehr and Bernhardt (2011), Boix and Svolik (2013), Casper and Tyson (2014), Chen et al. (2014), Shadmehr (2015), Guriev and Treisman (2015), Ananyev et al. (2015), Chen and Suen (2015), Egorov and Sonin (2015). 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.

mentarities drive the multiplicity of equilibria and the role of social norms in our model, which can shed light on the confounding empirical findings on the determinants of repression. In addition, we investigate how international pressure and domestic institutions influence the interactions between the state and the public, a question that has received little attention in the theoretical literature.

Second, our paper is related to the literature on the effect of international pressure on repression. In particular, Simmons (2009, p. 135-55) argues that international human rights treaties provide domestic activists with opportunities "to mobilize for human rights." However, the empirical evidence is mixed: human rights treaties reduce repression in some cases, but increase repression in others (Hathaway 2002; Hafner-Burton and Tsutsui 2005; Smith-Cannoy 2012; Magesan 2013). As Hafner-Burton (2012) documents in her comprehensive review of the literature, "there is a troubling and recurrent finding that participation in some treaties correlates with worse human rights behavior" (p. 280). Our paper provides a rationale for why international pressure in general, and human rights treaties in particular, may lead to an increase in repression: by altering the nature of updating (public opinion) following repression in favor of the state, international pressure can mitigate the public's incentive to protest, thereby reducing the domestic costs of repression to the state.

Third, our paper is related to the emerging literature on the institutions of authoritarian regimes. This literature has focused on parties and legislatures, military-civilian relationships, and power-sharing in dictatorships (Acemoglu et al. 2008; Gandhi 2008; Cheibub et al. 2010; Svolik 2012; Geddes et al. 2014). Our paper contributes to understanding the role and emergence of (partially) independent judiciaries with limited power to protect dissidents. In his review of the literature on "law and courts in authoritarian regimes," Moustafa (2014, p. 282) reiterates several key questions, including: "What motivates authoritarian rulers to grant nominal or even substantial independence to judicial institutions?...How do courts in authoritarian systems structure conflict and state-society interactions?" The literature suggests three broad answers. Independent judiciaries help the regime by protecting property rights, thereby enhancing economic development. They stabilize policy when there are electoral uncertainties or the prospects of regime transition (Ginsburg 2003; Hirschl 2004). Finally, a more independent judiciary provides the top echelon of the regime with more reliable information about the performance of their officials, reducing excessive corruption and mitigating principal-agent problems (Shapiro 1981; Verner 1984; Rosberg 1995; Peerenboom 2002; Ginsburg and Moustafa 2008). According to these explanations, protecting activists from repression is an unintended byproduct of judicial independence. In contrast, our results suggest that an independent judiciary whose primary purpose is to constrain repression can be beneficial to the regime. By granting a degree of judicial independence, the regime shifts the public's beliefs in favor of the state, stabilizing the regime by reducing the likelihood that repression leads to public protest.

From a theoretical perspective, the structure of our initial model, *without* exogenous or endogenous restrictions, is related to the political economy literature on the interaction between informed politicians and uninformed voters in democratic settings. This literature investigates topics such as the pandering of politicians to voters' opinions (Canes-Wrone et al. 2001), politicians' anti-herding behavior (Levy 2004), politicians' incentives to acquire policy expertise (Prato and Strulovici 2013), optimal institutional design (Maskin and Tirole 2004; Fox and Stephenson 2011), government transparency (Fox 2007), or populism (Acemoglu et al. 2013). In this literature, the authority over actions rests solely with the decision maker. In contrast, 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. Therefore, in contrast to this literature, updating about the activist's type (the state of the world) enters into the bystander's calculations. This, in turn, underlies the strategic complementarities that arise between the actions of the good ruler and the bystander, a key force in our analysis.[12]

Finally, our analysis treats commitment power in a novel way. Unlike the political delegation literature in which the uninformed principal commits to restrict the informed agent's decisions, in our model it is the informed agent—the bad ruler—who commits ex-ante to restrict *his own* subsequent choices. Moreover, to fit the nature of commitment in authoritarian settings, (in the text) only the bad ruler can commit, generating multiple equilibria that require appropriately designed refinements. That the informed agent (the ruler) also has a private conflict of interest, and that only one of the agent types can commit are among the new features of our analysis.[13]

[12] Our model has been developed for authoritarian settings. However, to further highlight its contrasts with the literature, it is useful to describe how it would map into democratic settings. Consider our base model, and relabel "ruler" as "incumbent politician", "activist" as "challenger", and "bystander citizen/public" as a "median voter" who is uncertain about the policy that maximizes his welfare. The challenger proposes a policy and commits to implement it if elected. There are two types of (policy proposals by the) challenger: the good policy improves the median voter's payoff over the status quo, the bad one worsens it. There are also two types of incumbent: the good/congruent incumbent shares the median voter's policy preference, while the bad/incongruent incumbent is biased toward the status quo. The incumbent knows the challenger's type, but the median voter is uncertain about the types of the incumbent and the challenger—he does not know whether the challenger's policy improves or worsens the status quo. After the challenger announces his platform, the incumbent has two options: if he preemptively implements the challenger's policy, he ensures that he will retain office—the median voter is assumed to re-elect him, a feature of the model that fits authoritarian settings (see our Discussion of the Model), but becomes problematic in democratic settings. If he maintains the status quo, then the median voter updates his beliefs, and votes accordingly. Beyond the base model, the exogenous and endogenous restrictions that we consider are difficult to interpret in democratic contexts.

[13] 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

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) demands reforms that would benefit the bystander (relative to the status quo) if implemented, while reforms demanded 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 receive 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 $p \in (0, 1)$, and the activist is bad with probability $q \in (0, 1)$.

The game proceeds as follows. First, nature chooses the ruler's type and the activist's type. The ruler observes the activist's type, and 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.[14] 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 (because 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

principal who then decides which action to take. In addition to the differences delineated above, 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.

[14] 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.

the ruler—even when the ruler is bad. Moreover, protesting against a good ruler has more potential downside than protesting against a bad ruler.[15] When the ruler is bad, the potential downside is $(\beta_b - \beta)$; when the ruler is good, the potential downside is β_b.[16]

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 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 $0 < \alpha_g < \alpha_b < 1$. We focus on the case where $\delta_b < \alpha_g$, so that the bad ruler's incentives to repress the good activist is larger than the good ruler's incentives to repress the bad activist.[17]

In Section 5, we modify the model to investigate the effects of exogenous restrictions on the bad ruler's ability to repress good activists. In Section 6, we modify the model to investigate the effects of the bad ruler's ability to observably commit to a repression 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, 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.

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

[16] 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, further reducing the bystander's payoff by β, for a net reduction of $-(\beta_b - \beta) - \beta = -\beta_b$.

[17] We focus on $\delta_b < \alpha_g$ because it captures the bad ruler's strong incentives to maintain the status quo, and because it leads to richer strategic interactions and more interesting results than the opposite case. Results for the case of $\delta_b > \alpha_g$ are available upon request.

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).[18] 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 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. 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 (McCarthy and Zald 1973, 1977; Gamson 1975; Tilly 1978, 1996,

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

2004; Tarrow 1998; McAdam 1999; McAdam, Tarrow, and Tilly 2001). 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). Therefore, in our model, when the ruler concedes to the activist's demands, the movement ends, and the ruler retains power.[19]

Repression by the good ruler. In line with the literature, 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 an organized movement that will achieve its goals unless stopped by coercive force. 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 persecutes 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 racist policies 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 protest in the low and intermediate repression equilibria in which the ruler may concede or the bystander citizen may protest, and hence these equilibria and their characteristics remain unchanged. Moreover, our modeling choice of having the activist always protest reflects the observation that some activists protest even when they are sure to be repressed as a matter of principle or as a strategy to gain 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 tradeoffs and strategic complementarities that underly our results (including multiple equilibria) also arise with private costs, while the analysis becomes significantly more cumbersome.

[19]This feature of our model captures an important aspect of the nature of interactions between the state, activists, and the public in many authoritarian settings, but this feature is less suitable for democratic settings. In democratic settings, the public can vote the ruler out of office regardless of his policy choices.

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 his (updated) belief that the activist is bad, q', and on his (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, his expected payoff is $-\beta p'$: with probability p' the ruler is bad, and the bystander receives $-\beta$; with the remaining probability the ruler is good, and the bystander receives his payoff that is normalized to 0. Therefore, the bystander's best response 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 best response 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).[20] To see why, note that the bad ruler dislikes conceding to the bad activist more 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. 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 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 remains uncertain about the activist's type following repression.

[20]Proof is in the Online Appendix I.

4 EQUILIBRIUM

If the bystander knew the types of the ruler and activist, he 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, he faces a tradeoff: by supporting the ruler, he risks blocking beneficial changes, but by supporting the activist, he risks implementing harmful changes. The bystander uses all the available information to update his 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:

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

Proposition 1 highlights the key aspects of the bystander's updating.[21]

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, holding the good ruler's strategy fixed, when the bad ruler represses the good activist more often, 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, the bystander's incentives to protest increase. However, when the good ruler represses the bad activist more, the bystander's incentives to protest *fall.* Further, as the bystander protests more, the incentives of both the good and the bad ruler fall. These underlying forces drive equilibrium behavior.

[21] The proof follows from (3) by simple algebraic manipulations and differentiations, and hence is omitted.

Proposition 2 *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, $\rho_b^G = 1$, 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:*

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

Figure 1 illustrates. 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.[22] This structure of best responses allows for multiple

[22]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.

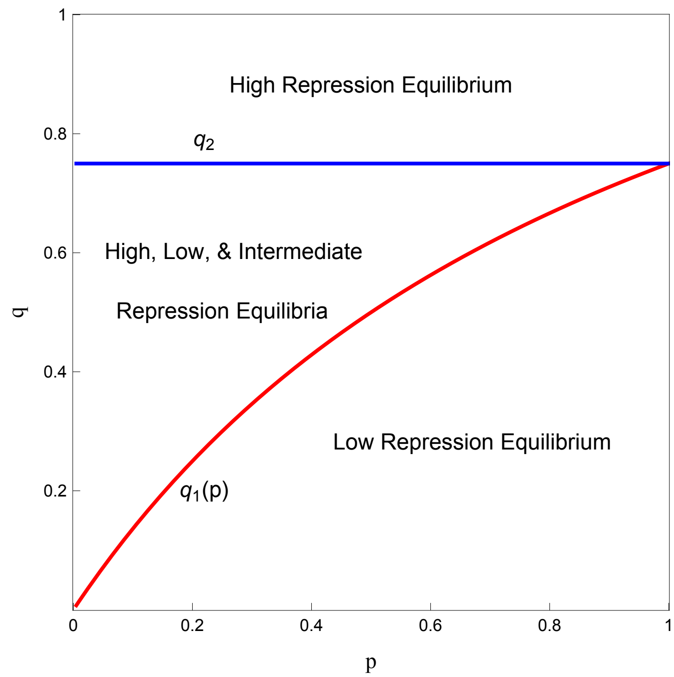

Figure 1: Equilibrium without restrictions.

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

Strategic forces 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 each 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, multiple equilibria arise in a larger area of the parameter space:* $\frac{\partial[q_2 - q_1(p)]}{\partial\beta} > 0$*.*

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. To offset 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 and repression. In particular, let G be 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, increases in G render the status quo under a bad ruler 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 rises, 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 vary across countries.

These results 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.[23] 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 the 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.*

[23] Examples include Miller et al. (1977), Weede (1981, 1986, 1987), Muller (1985, 1986), Midlarsky (1988), Muller and Seligson (1987), Lichbach (1989), Muller et al. and Midlarsky (1989), Muller and Weede (1990, 1994), Muller et al. (1991), Wang and Dixon et al. (1993). In the context of ethnic conflict, Esteban and Ray (2008, 2011) show the importance of intra-group inequality, revealing synergies between the rich and the poor within one ethnic group when fighting against another.

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, he 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.

5 EXOGENOUS LIMITS ON REPRESSION: INTERNATIONAL PRESSURE

Our analysis so far presumes that the only force that prevents bad governments from repressing legitimate activists is the endogenous threat of protest by the public. However, international pressure sometimes makes the repression of legitimate dissidents prohibitively costly. Foreign governments and international institutions can pressure a regime into reducing repression by suspending foreign aid, terminating trade benefits, or making membership contingent on human rights reform. U.S. pressure on Guatemala in the 1990s and European Union pressure on Turkey are two classic examples (Hafner-Burton 2013, Ch. 8). Moreover, regimes may inherit human rights treaties that sometimes make it too costly to repress legitimate activists, or they may join such treaties to receive foreign aid (Magesan 2013). Whether inherited or the result of a choice (to receive benefits that we do not explicitly model), international pressure can prevent regimes from repressing legitimate activists (Simmons 2009; Simmons and Danner 2010). However, mounting sufficient pressure requires significant economic and political resources which may be difficult to muster, especially when international actors have incentives to free-ride.[24] Consequently, the international community may fail to prevent the repression of legitimate dissidents.

To explore the effects of international pressure, we consider the following modification of our base model. Suppose that when a ruler attempts to repress the good activist, the attempt fails with (an exogenous) probability $C \in (0, 1)$ due to international pressure, in which case the ruler is forced to concede to the good activist's demands. With the remaining probability, $1 - C$, the ruler's attempt to repress the good activist succeeds. The bystander knows the likelihood C of a successful intervention, but he does not observe the actions of the ruler or international community.[25]

[24]For example, regarding the enforcement of human rights treaties, Simmons (2009) argues that the members of international community "have incentive to ignore violations, either because they are essentially unaffected by practice elsewhere, or because other foreign policy objectives swamp the concerns they have in a particular case, or because they hope someone else will pay the costs of enforcement" (p. 126).

[25]This "repression obstruction" does not automatically generate commitment power for the bad ruler: his strategy

The possibility that the bad ruler's repression of the good activist can fail changes how the bystander updates his beliefs from (3) to:

$$(4) \quad p' = \frac{p[q + (1-q)(1-C)\rho_g^B]}{p[q + (1-q)(1-C)\rho_g^B] + (1-p)q\rho_b^G} \qquad q' = \frac{q[(1-p)\rho_b^G + p]}{q[(1-p)\rho_b^G + p] + p(1-q)(1-C)\rho_g^B}.$$

When the bystander observes repression, he only assigns a probability $(1-C) \times [p(1-q)\rho_g^B]$ that a bad ruler has repressed a good activist. That is, holding the ruler's strategy fixed, increases in the strength of international pressure (i.e. increases in C) mean that repression is less likely to be carried out by a bad ruler against a good activist. Therefore, the bystander updates less negatively about the ruler and more negatively about the activist, shifting the bystander's beliefs in the ruler's favor. This effect has important implications for the equilibria, which are characterized in Proposition 3.

Proposition 3 *Suppose that the ruler's attempt to repress a good activist is blocked with probability $C \in (0,1)$. In equilibrium, $\rho_g^G = 0$ and $\rho_b^B = 1$. For a given C, there exists an increasing function $g(p) \in (0,1)$ such that:*

- ***Low Repression Equilibrium with International Pressure:*** *If $q < g(p)q_1(p)$, then a unique equilibrium exists in which $\rho_b^G = 0$, $\rho_g^B = \frac{\beta_b - \beta}{\beta_g + \beta}\frac{q}{1-q}\frac{1}{1-C}$, and $\pi = \alpha_g$.*
- ***High Repression Equilibrium with International Pressure:*** *If $q > g(1)q_2$, then a unique equilibrium exists in which $\rho_b^G = 1$, $\rho_g^B = 1$, and $\pi = 0$.*
- ***Intermediate Repression Equilibrium with International Pressure:*** *If $g(p)\ q_1(p) \leq q \leq g(1)\ q_2$, then, in addition to the above two equilibria, there exists an equilibrium in which:*

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

The structure of equilibria is similar to our base model. There can be three equilibria which are ranked according to their repression levels. Because international pressure shifts the bystander's beliefs in the ruler's favor, international pressure reduces the bystander's incentive to join the protest, leading the bad ruler to attempt repression of the good activist more often. As Figure 2 illustrates, the set of parameters in which the high repression equilibrium arises expands from $q > q_1(p)$ in our base model to $q > g(p, C)\ q_1(p)$, where we have made the dependence of g on C explicit. Without international pressure, in regions I and II the low repression equilibrium is unique, and the bad ruler represses the good activist with probability $\rho_g^B = \frac{\beta_b - \beta}{\beta_g + \beta}\frac{q}{1-q}$. With international pressure, the high

must still satisfy equation (2). We analyze the bad ruler's commitment power in the next section.

repression equilibrium is unique in region I, and all three equilibria arise in region II. In the high and intermediate repression equilibria with international pressure, the bad ruler always *attempts* to repress the good activist, succeeding with probability $1-C$. In the low repression equilibrium with international pressure the likelihood that the good activist is repressed is identical to the unique equilibrium without international pressure.

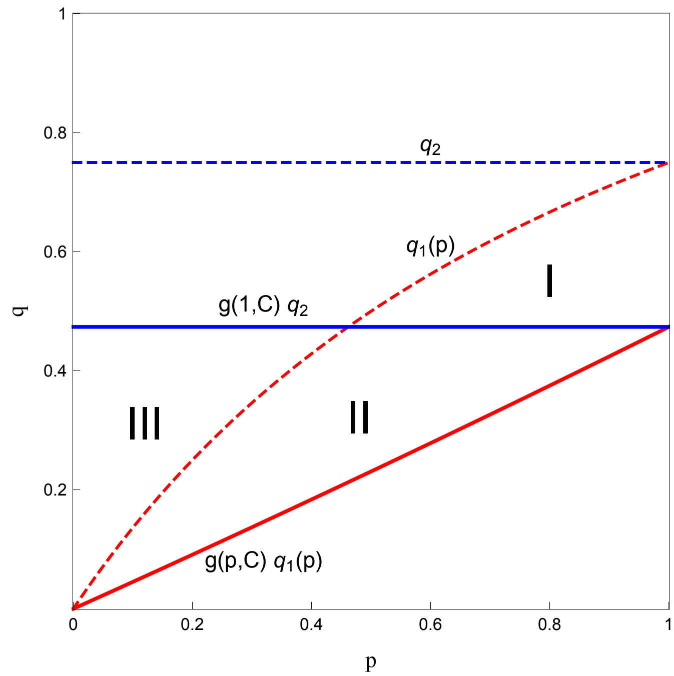

Figure 2: Equilibrium with exogenous restrictions. Region I: $g(1,C)q_2 < q < q_1(p)$. Region II: $g(p,C)q_1(p) < q < \min\{q_1(p), g(1,C)q_2\}$. Region III: $q_1(p) < q < g(1,C)q_2$.

Protection of Good Activists and Implementation of Good Reforms. Critically, even though under some conditions international pressure protects the good activist and raises the likelihood that good reforms are implemented, it can also generate the opposite effects. This is most clear in region II of Figure 2, where the low repression equilibrium is unique without international pressure, but the high and intermediate repression equilibria arise with international pressure.

Corollary 4 *In region II, the likelihood that the good activist is repressed and good reforms are blocked is higher in the intermediate and high repression equilibria with international pressure than in the unique equilibrium without international pressure.*

In Corollary 4, we highlight region II because the uniqueness of the low repression equilibrium without international pressure allows for a sharp comparison. In essence, in region II, international pressure makes it possible to sustain equilibria in which the good activist is more likely to be repressed successfully, even though the international community blocks the ruler's repression attempt some of the time.[26] In particular, the likelihood that the good activist is repressed in the intermediate or high repression equilibrium with international pressure is strictly larger than that likelihood in the unique equilibrium without international pressure, $1 - C > \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}$.[27] This result resonates with a "troubling and recurrent finding" of the empirical literature, "that participation in some international treaties correlates with worse human rights behavior" (Hafner-Burton 2012, p. 280).

Of course, international pressure can also be unambiguously helpful. In particular, international pressure reduces the likelihood that the good activist is repressed above the solid horizontal line in Figure 2 where $q > g(1, C)q_2$. In particular, in region I, $1-C < \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}$. In contrast, it does not affect this likelihood under the solid curve, where $q < g(p, C)q_1(p)$. However, even when international pressure reduces the overall likelihood that the good activist is repressed, it comes with a cost.

Corollary 5 *There exists $\hat{\alpha} \in (0, C)$ such that the likelihood that good reforms are blocked is higher with international pressure in region I if and only if $\alpha_g > \hat{\alpha}_g$.*

Corollary 5 implies that when good activists demand fundamental reforms (so that α_g is high) in countries with relatively unpopular rulers (relatively high p) and moderately unpopular activists ($q \in (g(1, C)q_2, q_1(p))$), international pressure necessarily hinders good reforms by reducing the public's incentives to protest—even as it reduces the overall likelihood that good activists are repressed.[28] To see the intuition, note that in region I, international pressure changes the unique equilibrium in three ways: the likelihood that the public protests following repression drops from α_g to 0, the likelihood that the bad ruler attempts to repress the good activist increases from $\frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}$ to 1, and the likelihood that the international community intervenes to prevent repression of the good activist increases from 0 to C. Because international pressure shifts the public's beliefs in favor of the ruler, it completely displaces domestic checks on repression generated by the potential for public protest, which in turn induces the bad ruler to try to repress the good activist with probability one. Therefore, with international pressure, a good reform can only be implemented due to international pressure that forces the bad ruler to concede. In contrast, in the unique (low

[26] In region III multiple equilibria arise, whether or not there is international pressure; the overall likelihood that the good activist is repressed is also higher in the intermediate and high repression equilibria with international pressure than in the low repression equilibrium without pressure. Therefore, even in region III it is also possible that international pressure results in a higher likelihood that a good activist is repressed.

[27] These are the likelihoods conditional on the bad ruler being in power and the activist being the good type; the corresponding ex ante likelihoods are obtained by multiplying both sides by $p(1-q)$, which then cancel.

[28] It is worth mentioning that public opinion parameters p and q can be estimated; for example, surveys are used to gauge the public's views of the government and opposition groups even in many authoritarian regimes (Cammett 2011; Jamal 2012; Beissinger et al. 2015).

repression) equilibrium without international pressure, a beneficial reform can be implemented *either* through a direct concession by the ruler or by a public protest. When α_g is large enough, to counter the bad ruler's large incentives to repress, the bystander protests so often that the overall likelihood of good reforms is higher without international pressure.

Strength of International Pressure (Magnitude of C). Our preceding results investigate the effects of a particular level of international pressure on different societies (with different primitive parameters). But international institutions and foreign governments can exert different *degrees* of pressure on a particular government. Thus, we study how different degrees of international pressure affect the same society. We focus on cases in which the good activists are more likely to arise ($q < q_1(p)$), where the uniqueness of equilibrium without international pressure permits sharper comparisons. Let $R_g(C)$ be the equilibrium likelihood that the good activist is successfully repressed by the bad ruler, given international pressure of degree C.[29]

Proposition 4 *Consider a society with parameters (p, q) such that $q < q_1(p)$. International pressure decreases the likelihood that the good activist is repressed if and only if it is sufficiently strong. When it is weaker, it can <u>increase</u> the likelihood that the good activist is repressed. Formally, there exist $0 < \underline{C} < \overline{C} < 1$ such that: (i) $R_g(C) < R_g(0)$ if and only if $C > \overline{C}$, and (ii) if $C \in (\underline{C}, \overline{C})$ and either the intermediate or high repression equilibrium is selected, then $R_g(C) > R_g(0)$.*

Figure 3 illustrates the results. The logic builds on Corollaries 4 and 5. As international pressure becomes stronger (i.e., as C increases) the lower boundaries of regions I and II, $g(1, C)q_2$ and $g(p, C)q_1(p)$, smoothly shift downward. Therefore, when C is small ($C < \underline{C}$), a society with $q < q_1(p)$ remains below region II and in the low repression equilibrium with international pressure, in which the overall likelihood that the good activist is repressed is unaffected by international pressure. As international pressure increases moderately to $C \in (\underline{C}, \overline{C})$, the society becomes part of region II where all three equilibria are possible. If the low repression equilibrium with international pressure is selected, again international pressure has no effect. However, if the intermediate or high repression equilibrium is selected, then the likelihood that the good activist is repressed is strictly higher than the likelihood without international pressure, $1 - C > \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}$. In this case, the strategic effect of international pressure—the reduction in the public's incentives to protest—dominates its direct effect. Finally, as international pressure rises even further ($C > \overline{C}$), the society becomes part of region I, in which $1 - C < \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}$. In this case, the direct effect of international pressure

[29] That is, if ρ_g^B comes from the bad ruler's equilibrium strategy, then $R_g(C) \equiv p(1-q)\rho_g^B(1-C)$.

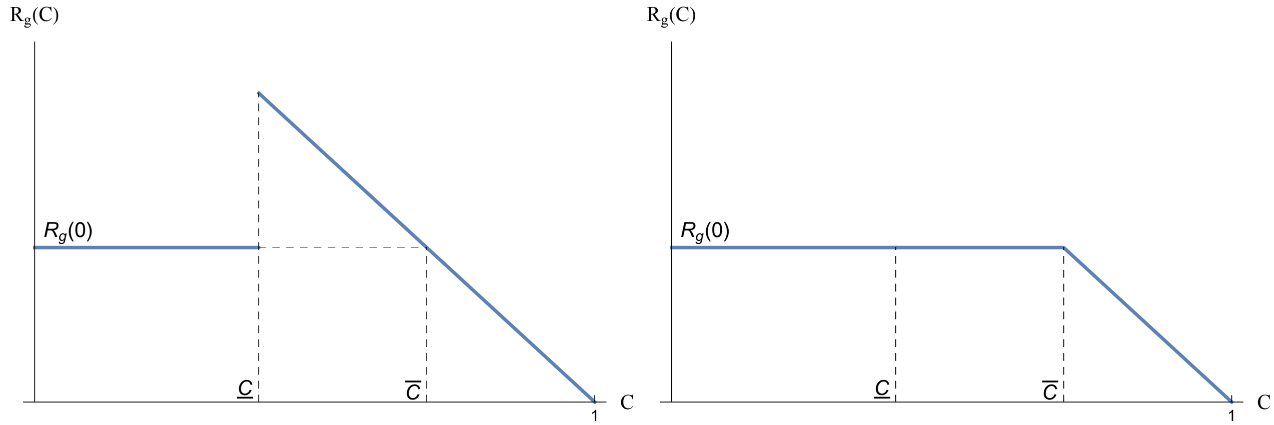

Figure 3: The likelihood that the good activist is repressed by the bad ruler $R_g(C)$ as a function of the strength of international pressure C. Right panel depicts the case in which the low repression equilibrium with international pressure is selected in region II. Left panel depicts the case in which the intermediate or high repression equilibrium with international pressure is selected in region II.

dominates, and the overall likelihood that the good activist is repressed falls.[30]

These results have an immediate policy consequence. Small levels of international pressure either have no effect or an adverse effect. Thus, international pressure should be used only if it is sufficiently forceful $C > \overline{C} = 1 - \frac{\beta_b - \beta}{\beta_g + \beta} \frac{q}{1-q}$. Moreover, because $\frac{\partial \overline{C}}{\partial q} < 0 < \frac{\partial \overline{C}}{\partial \beta}$, the minimum degree of international pressure that renders it beneficial is higher when (i) the good activists are likely to arise so that the public is less suspicious of the activists (q is lower), or (ii) when the status quo under the bad ruler is worse (β is higher, corresponding to lower income in our interpretation in Corollary 2). This suggests that situations in which international pressure appears most necessary are also more prone to generate ineffective or adverse results.

6 ENDOGENOUS LIMITS ON REPRESSION: DOMESTIC INSTITUTIONS

In the previous section, we studied the effect of exogenous restrictions on the repression of good activists, showing that such restrictions can lead to *more* repression of good activists. This finding suggests that bad governments may have incentives to set up institutions ex-ante that restrict their repression of good activists. In particular, 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. In this section, we consider the implications of the bad government's commitment power.

[30]The likelihood that good reforms are blocked depends on C in a similar way, except that when $C > \overline{C}$, international pressure increases the likelihood that good reforms are blocked if and only if $\alpha_g > \hat{\alpha}_g$.

To illustrate the settings that we model, consider a regime consisting of a good faction and a bad faction. Both factions participate in the government, but only one of the factions has real authority at a time, and the real authority can switch from one faction to another behind the scenes. When the bad faction is in power, it delegates repression decisions to an institution such as a security agency. This enables the bad faction to commit to a repression strategy: by selecting attributes of this agency (its composition, size, and funding), the bad faction can determine the degree to which this agency represses activists of different types. For example, by recruiting a larger fraction of officers who would be sympathetic to good activists, the bad faction can reduce the probability that the security agency represses good activists.[31] The good and bad factions have conflicting interests with respect to repression. Therefore, when the de facto power within the regime shifts from the bad to the good faction, the good faction also takes over the authority to make repression decisions from the bad faction's security agency.

Critically, the allocation of real authority inside an authoritarian regime is typically not transparent. 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 is leading the country, power may be shifting between different factions of the leadership behind the scenes. Because behind-the-scenes transitions of de facto power between factions remain unnoticed by the public, when repression is observed, the public cannot distinguish whether it was ordered by the bad faction's security agency or by the good faction.[32]

To investigate the effect of 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 whether the bad ruler has been replaced or remained in power. Next, the activist protests. As in the original

[31] Once the security agency's officers are selected, adjusting its composition can be very costly to the bad faction. The security service fulfills important tasks besides repressing activists (e.g. identifying and defusing foreign threats, disciplining the bureaucracy, etc.), and tampering with the security agency's attributes would interfere with these critical functions. Moreover, attempts to shuffle high-ranking officers could raise the potential for coup attempts or conflict among the elite that could seriously threaten the regime's survival.

[32] At the end of the section, we also consider a weaker form of commitment power whereby the bad faction ex-ante commits to a degree of judicial independence, allowing the judiciary to block its attempt to repress good activists some of the time. Judicial independence is a weaker form of commitment because it does not allow the bad faction to choose its repression strategy exactly, only to limit the probability with which good activists are repressed. In the Online Appendix II, we consider the case in which both factions simultaneously commit to their repression strategies ex-ante.

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.[33]

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:*

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

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

Lemma 4 in the Online 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

[33] As before, $\rho_g^G = 0$ in equilibrium because the good ruler prefers good reforms to the status quo.

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 (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. In the text, we focus on the interesting case in which the protest probability is δ_b whenever $F(r_b, r_g) \in (0, K)$.[34] Thus, 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 he sometimes protests, to $F > K$, where he 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 4 in the Online Appendix I. 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 Online Appendix I). Given these equilibrium selections, Proposition 5 formally describes the equilibrium, and Figure 4 illustrates the four equilibrium regions that arise.

Proposition 5 *Suppose that the bad ruler can commit to his strategy (r_b, r_g). 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).*

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 $p^* = \frac{\beta_b(\alpha_g - \delta_b)}{\beta_b\alpha_g + \delta_b\beta_g}$.

[34] Propositions 8 and 9 in the Online Appendix II 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.

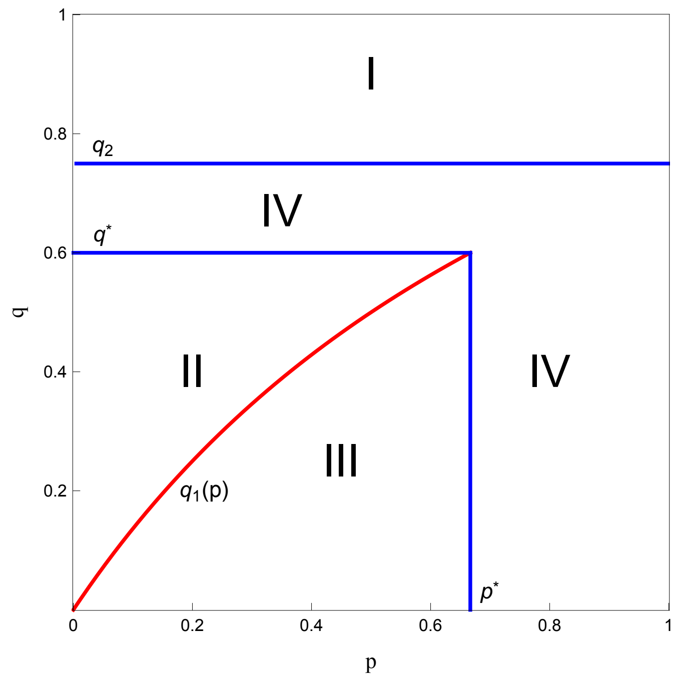

Figure 4: Equilibrium with endogenous restrictions. Proposition 5 describes the equilibrium.

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: as Lemma 2 shows, 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 choice, *in equilibrium*, 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.[35] 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 potentially optimal choices: (1) he can repress both types of activist with probability one ($r_2 = 1$) and face

[35]The thresholds $q = q_2$ and $q = q_1(p)$ determine the value of $F(1, 1, q)$, where we have made the dependency of F on q explicit. When $q > q_2$, $F(1, 1, q) < 0$, and hence the bad ruler who chooses $r_g = 1$ faces $\pi = 0$. When $q_1(p) < q < q_2$, $F(1, 1, q) \in (0, K)$, and hence the bad ruler who chooses $r_g = 1$ faces $\pi = \delta_b$. When $q < q_1(p)$, $F(1, 1, q) > K$, and hence the bad ruler who chooses $r_g = 1$ faces $\pi = 1$.

the bystander's protest with probability δ_b, or (2) he can limit the repression of the good activist to $r_1 = \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta} < 1$, 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:[36] (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. As p increases, the bystander believes that the ruler is more likely to be bad. As a result, he 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. Therefore, when p is large (region IV), the ruler chooses option (1) in which the bystander does not protest. When p is smaller (region III), $r_2(p)$ is sufficiently large that the gain from raising the repression probability from r_1 to $r_2(p)$ offsets the corresponding increase in the protest probability from 0 to δ_b, and hence the rulers chooses option (2).

Corollary 6 *In equilibrium, the bystander does not protest upon repression if and only if the probability that the ruler is bad <u>or</u> the probability that the activist is bad is sufficiently <u>large</u>. There exists $(p^*, q^*) \in (0,1)^2$ 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 his repression of good activists in regions III and IV. To understand the effects of this commitment power, we analyze how the ruler's equilibrium behavior varies with q, focusing on the more interesting case when $p < p^*$. Figure 5 illustrates. Increases in q have 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.[37]

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

[36]These choices correspond to $F(1, r_1) = 0$ and $F(1, r_2) = K$.

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

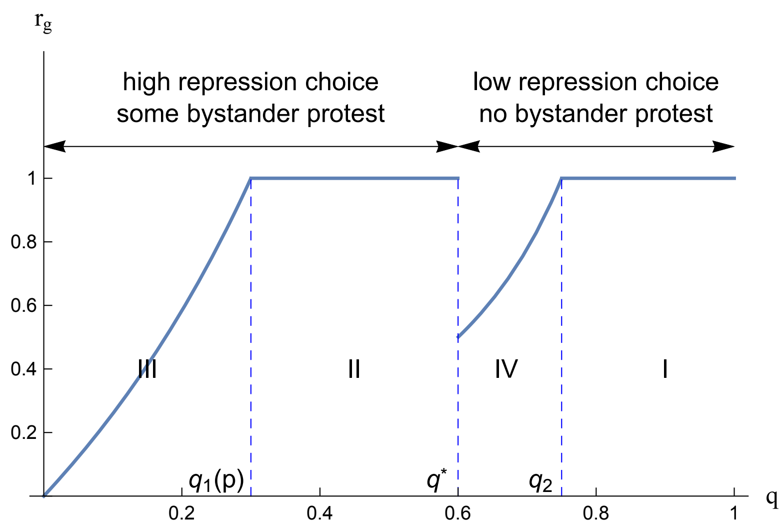

Figure 5: The bad ruler's equilibrium likelihood of repressing the good activist as a function of the prior likelihood q that the activist is bad—for a given $p < p^*$.

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 his protest can lead to bad reforms, effect (ii), is irrelevant. In contrast, when $r_g = r_2$, the bystander protests with probability $\pi = \delta_b$, and hence both (i) and (ii) contribute to reducing his 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 5 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 7 *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. Knowing that the bad ruler cannot raise his repression of the good activist beyond the committed level, the bystander updates less negatively about the ruler and more negatively about the activist following repression, and hence is less inclined to protest. This raises the good ruler's incentive to repress the bad activist, which, in turn, further reduces the bystander's incentive to protest, and so on. 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.[38]

Corollary 8 *When $q < q_1(p)$ and $p < p^*$, with commitment, the bad ruler represses the good activist more often than in the 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.

Independent Judiciary and Commitment Power. Our preceding analysis of the bad ruler's commitment is based on his ability to delegate repression to an agency that implements his desired repression strategy, repressing the good activist neither more, nor less often than the ruler desires ex-ante. This is a strong form of commitment power, and it leads us to investigate a setting in which the ruler's commitment power is less demanding.

Judicial institutions in authoritarian regimes provide a basis for a weaker form of commitment power. Until recently, the conventional wisdom held that judiciaries in authoritarian regimes are mere window dressing, at least with regard to the protection of activists. However, recent litera-

[38] From Propositions 2, when $q < q_1(p)$, including 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. Notably, absent commitment, the good ruler does not repress in equilibrium; in contrast, with commitment, the good ruler always represses the bad activist.

ture suggests that although judiciaries rarely oppose the executive in some authoritarian regimes, in others the judiciary *sometimes* protects dissidents, for example, by "refusing to prosecute people for exercising their human rights, and seeking to prosecute the police for abuses of power" Hilbink (2012, p. 598-99). "Reform-minded judges may work to push the envelope through their judgments, capitalizing on the important state functions that they perform to leverage pressure in the direction of political reform" (Moustafa 2014, p. 288; see also Rosberg (1995) ch. four and Ip (2012)). More broadly, this literature highlights that the degree of judicial independence varies across authoritarian regimes. Thus, we model the degree of judicial independence by the likelihood with which the judiciary strikes down the ruler's attempt to repress the good activist.[39] Our goal is to investigate the bad ruler's decision to commit ex-ante to a degree of judicial independence and to determine how this weaker form of commitment influences the outcome.[40]

In Proposition 6, we show that if the bad ruler commits to a level of judicial independence, then equilibrium outcomes are equivalent to the ones that arise if he commits to his repression strategy. To see the key intuition, note that in regions III and IV of Figure 4 (the two cases where the bad ruler's commitment power is relevant), the bad ruler supports an equilibrium with a low likelihood of protest by committing to repress the good activists with probability $r_g < 1$. However, given the low likelihood of protest ($\pi < \alpha_g$), if the bad ruler's commitment power were removed, then the bad ruler would like to deviate by repressing the good activists with probability one; absent other restrictions on the ruler's strategy, this would upset the equilibrium. Now, suppose that when the bad ruler's commitment power is removed, an independent judiciary is simultaneously introduced, which blocks the ruler's attempt to repress a good activist with probability $C = 1 - r_g$. The bad ruler still *tries* to repress the good activist with probability one, but he only succeeds with probability r_g. Therefore, the likelihood that the good activist is repressed is identical to the equilibrium with precise commitment, the bystander's beliefs are unchanged, and the strategies of the bystander and good ruler continue to be best responses. Thus, the equilibrium that arises when the bad ruler commits to his strategy can be recreated by selecting a particular level of judicial independence.

To make this argument precise, one must also select the equilibrium in a consistent manner in both settings—recall that both settings feature multiple equilibria in the subgame that follows the bad ruler's choice. The subgame following the bad ruler's choice of judicial independence is identical to the game in which repression of good activists is blocked with a particular probability, and the

[39]Analogous results obtain if the bad ruler commits to an upper bound on the likelihood of repressing good activists.

[40]Commitment to a degree of judicial independence is a weaker form of commitment because it does not allow the bad ruler to choose his repression strategy exactly, only to limit the probability with which good activists are repressed.

equilibria of this subgame are characterized in Proposition 3. For $q \in [g(p)q_1(p), g(1)q_2]$, we consider the following equilibrium selection: if $q = g(1)q_2$, then select the high repression equilibrium, and if $q \in [g(p)q_1(p), g(1)q_2)$, then select the intermediate repression equilibrium.[41] With this equilibrium selection, the equilibrium of the game in which the bad ruler commits to a repression strategy and the equilibrium of the game in which the bad ruler chooses the level of judicial independence are outcome-equivalent. That is, the probability that each type of ruler *successfully* represses each type of activist and the probability that the bystander joins the protest are identical in both.

Proposition 6 *Suppose that the bad ruler commits to a level of judicial independence ex-ante. The equilibrium of this game is outcome-equivalent to the game in which the bad ruler commits to his repression strategy.*

Proposition 6 implies that the ability to establish judicial independence allows the bad ruler to exploit the strategic complementarity between the bystander and the good government, which is the essence of his commitment power. By creating an independent judiciary that limits repression of the good activist, the bad ruler shifts the formation of public opinion in his favor, reducing the public's incentive to protest. Consequently, the good ruler is more inclined to repress the bad activist, further reinforcing the favorable shift in the public's beliefs. These effects reduce the endogenous cost of repression for the state, resulting in a lower likelihood of protest and a higher likelihood of repression in equilibrium. In particular in region III of Figure 2, the bad ruler introduces a judiciary with a relatively low degree of independence, reducing the equilibrium level of protest from $\pi = \alpha_g$ (in the absence of any restrictions on the ruler's strategy) to $\pi = \delta_b$ and increasing the likelihood with which he *successfully* represses the good activist (see Corollary 8). In region IV, the bad ruler introduces a judiciary with a higher degree of independence, supporting an equilibrium in which the good ruler always represses the good activist and the bystander never protests.[42]

Proposition 6 suggests a novel explanation for the emergence of independent judiciary in autocracy. Three explanations have been offered in the literature. First, independent judiciaries protect

[41] This equilibrium selection is consistent in the following sense with the one we introduced when analyzing the bad ruler's commitment to his strategy. Recall that in that setting equilibrium multiplicity arises when the bad ruler's choice of (r_b, r_g) satisfies $F(r_b, r_g) \in [0, K]$, and we select the equilibrium with $\pi = \delta_b$ when $F(r_b, r_g) \in (0, K]$, and the equilibrium with $\pi = 0$ when $F(r_b, r_g) = 0$. When multiple equilibria arise in the subgame following the ruler's choice of judicial independence, one of those equilibria features $\pi = \alpha_g$. This equilibrium is inconsistent with our preceding selection in which $\pi = \delta_b$ or $\pi = 0$, and hence we do not select it. In the other equilibria, the bad activist is repressed with probability 1, and the good activist is repressed successfully with probability $1 - C$. Thus, we select the equilibrium with $\pi = 0$ whenever $F(1, 1 - C) = 0$ and the equilibrium with $\pi = \delta_b$ whenever $F(1, 1 - C) \in (0, K]$.

[42] In some cases, the bad ruler prefers not to establish an independent judiciary that constrains repression; this happens when the activist is very unpopular (region I) or when both activist and ruler are popular but the ruler is relatively more popular than the activist (region II).

property rights and enforce contracts, promoting economic development. Second, they impose restrictions on a regime's successors, ensuring that outgoing elites will not face unfettered opposition rule. Third, they monitor the activities of low level officials, mitigating principal-agent problems between the elite and the bureaucracy. None of these explanations directly addresses the role of judicial independence in promoting human rights or constraining repression. "Even if dictators are willing to support judicial independence in the area of property rights, there is no guarantee that such independence will be allowed to spill over into human rights or political and civil liberties" (Helmke and Rosenbluth 2009 p. 358). According to the prevailing explanations, if such a spillover occurs, then it must be an *unintended byproduct* of judicial independence. In contrast, our analysis demonstrates that an independent judiciary whose *primary purpose* is to protect human rights and prevent repression can benefit the regime. An independent judiciary can play two seemingly contradictory roles, stabilizing the dictatorship by promoting human rights.

7 CONCLUSION

"The seed of revolution is repression," once said Woodrow Wilson.[43] While the state's repression of activists may terminate their movement, it may also spur the general public to join the activists, leading to revolution. The literature has focused on the deterrence aspect of repression, ignoring the empirical evidence that protest can spread following repression, generating an endogenous cost of repression for the state. We develop a model to study the interactions between the state's incentive to repress activists and the public's incentive to join the protest, based on the key observation that the general public has limited information about the nature of the activists' demands and the intentions of the state that represses them. We present three sets of results. Our first set of results, presented in Section 4, highlights the critical role played by social norms in this interaction, shedding light on conflicting empirical results regarding the determinants of repression. Our second set of results, presented in Section 5, analyzes the effect of international pressure, which places exogenous limits on the state's ability to repress legitimate activists. Paradoxically, these restrictions may increase the repression of good activists, resonating with empirical studies showing that a state's ratification of human rights treaties may increase its use of repression. Our third set of results, presented in

[43] In his 7th annual message to the congress in 1919, he wrote: "I would call your attention to the widespread condition of political restlessness in our body politic.... Broadly, they arise...from the machinations of passionate and malevolent agitators.... The only way to keep men from agitating against grievances is to remove the grievances. An unwillingness even to discuss these matters produces only dissatisfaction and gives comfort to the extreme elements in our country which endeavor to stir up disturbances *in order to provoke governments to embark upon a course of retaliation and repression. The seed of revolution is repression* [our emphasis]."

Section 6, considers endogenous restrictions on repression: the state introduces an institution that restricts its ability to repress good activists. By doing so, the state manipulates the public's beliefs, which allows it to increase the probability of repression while reducing the probability of protest, suggesting a novel explanation for the existence of partially independent judiciaries in autocracies.

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. 2008. "Coalition Formation in Non-Democracies." *Review of Economic Studies* 75: 987-1009.

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.

Beissinger, Mark, Amaney A Jamal, and Kevin Mazur. 2015. "Explaining Divergent Revolutionary Coalitions: Regime Strategies and the Structuring of Participation in the Tunisian and Egyptian Revolutions." *Comparative Politics* 48: 1-24.

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.

Cammett, Melani C. 2011. "Partisan Activism and Access to Welfare in Lebanon." *Studies in Comparative International Development* 46: 70-97.

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.

Cheibub, Jose A., Jennifer Gandhi, and James R. Vreeland. 2010. "Democracy and Dictatorship Revisited." *Public Choice* 143: 67-101.

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." *The Economic Journal. Forthcoming.*

Chen, Heng, and Wing Suen. 2016. "Falling Dominoes: A Theory of Rare Events and Crisis Contagion." *American Economic Journal: Microeconomics* 2016 8: 1-29.

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. "A Model of Ethnic Conflict." *Journal of the European Economic Association* 9: 496-521.

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, NY: Cambridge University Press.

Geddes, Barbara, Erica Frantz, and Joseph Wright. 2014. "Military Rule." *Annual Review of Political Science* 17: 147-62.

Ginsburg, Tom. 2003. *Judicial Review in New Democracies: Constitutional Courts in Asian Cases.* New York, NY: Cambridge Univ. Press

Ginsburg, Tom, and Tamir Moustafa. eds. 2008. *Rule by Law: The Politics of Courts in Authoritarian Regimes.* New York, NY: Cambridge University Press.

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

Hafner-Burton, Emilie. 2012. "International Regimes for Human Rights." *Annual Review of Political Science* 15: 265-86.

Hafner-Burton, Emilie. 2013. *Making Human Rights a Reality.* Princeton, NJ: Princeton University Press.

Hafner-Burton, Emilie, and Kiyoteru Tsutsui. 2005. "Human Rights in a Globalizing World: The Paradox of Empty Promises." *American Journal of Sociology* 110: 1373-411.

Hathaway, Oona. 2002. "Do Human Rights Treaties Make a Difference?" *Yale Law Journal* 111: 1935-2042.

Helmke, Gretchen and Frances Rosenbluth. 2009. "Regimes and the Rule of Law: Judicial Independence in Comparative Perspective." *Annual Review of Political Science* 12: 345-66.

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

Hilbink, Lisa. 2012. "The Origins of Positive Judicial Independence." *World Politics* 64: 587-621.

Hirschl, Ran. 2004. *Towards Juristocracy: The Origins and Consequences of the New Constitutionalism.* Cambridge, MA: Harvard University Press.

Hollyer, James, and Peter Rosendorff. 2011. "Why Do Authoritarian Regimes Sign the Convention Against Torture? Signaling, Domestic Politics, and Non-Compliance." *Quarterly Journal of Political Science* 6: 275-327.

Ip, Eric C. 2012. "Judicial Review in China: A Positive Political Economy Analysis." *Review of Law & Economics* 8: 331-66.

Jamal, Amaney A. 2012. *Of Empires and Citizens: Pro-American Democracy or No Democracy at All?* Princeton, NJ: Princeton University Press.

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

KONDA. 2014. "GEZI REPORT: Public Perception of the Gezi Protests. Who Were the People at Gezi Park?" Available at: http://konda.com.tr/en/raporlar/KONDA_Gezi_Report.pdf

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

Magesan, Arvind. 2013. "Human Rights Treaty Ratification of Aid Receiving Countries." *World Development* 45: 175-88.

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: 964-82.

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.

Moustafa, Tamir. 2014. "Law and Courts in Authoritarian Regimes." *Annual Review of Law and Social Science* 10: 281-99.

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.

Peerenboom, Randall. 2002. *China's Long March Toward Rule of Law.* Cambridge, UK: Cambridge University Press.

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

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 J. Benhabib, A. Bisin, and M. Jackson. Amsterdam: North-Holland.

Prato, Carlo and Bruno Strulovici. 2013. "Direct Democracy and Responsibility Substitution." Mimeo.

Rosberg, James H. 1995. "Roads to the Rule of Law: The Emergence of an Independent Judiciary in Contemporary Egypt." PhD Thesis, Department of Political Science, MIT.

Simmons, Beth. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics.* New York, NY: Cambridge University Press.

Simmons Beth, and Allison Danner. 2010. "Credible Commitments and the International Criminal Court." *International Organization* 64: 225-56.

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." *Games and Economic Behavior* 94: 97-121.

Shadmehr, Mehdi, and Dan Bernhardt. 2011. "Collective Action with Uncertain Payoffs: Coordination, 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.*

Shapiro, Martin. 1981. *Courts: A Comparative and Political Analysis.* Chicago, IL: University of Chicago Press.

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

Smith-Cannoy, Heather. 2012. *Insincere Commitments: Human Rights Treaties, Abusive States, and Citizen Activism.* Washington DC: Georgetown University Press

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

Svolik, Milan. 2012. *The Politics of Authoritarian Rule.* New York, NY: Cambridge Univ. Press.

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, NY: Cambridge University Press.

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

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.

Verner, Joel. 1984. "The Independence of Supreme Courts in Latin America: A Review of the Literature." *Journal of Latin American Studies* 16: 463-506.

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

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

For Online Publication

9 APPENDIX I: PROOFS

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. ■

9.1 PROOFS FOR SECTION 4: NO LIMITS ON REPRESSION

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:

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

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, $\rho_g^B \in [0,1] \Leftrightarrow \frac{(\beta+\beta_g)p}{\beta_b+\beta_g p} \le q \le \frac{\beta+\beta_g}{\beta_b+\beta_g}$, 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 $q_1(p) \le q \le q_2$.

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 $\beta_g(1-q') - \beta_b q' \le -\beta p'$. Substituting (p', q') gives:

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

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

9.2 PROOFS FOR SECTION 5: EXOGENOUS LIMITS ON REPRESSION

Proof of Proposition 3. Let $g(p) \equiv 1 - C\frac{\beta_b - p\beta}{\beta_b - p\beta + (1-C)p(\beta+\beta_g)}$. Note that $g(p) \in (0,1)$ for $C \in (0,1)$, and that $g'(p) = C(1-C)\frac{\beta_b(\beta+\beta_g)}{(\beta_b - p\beta + (1-C)p(\beta+\beta_g))^2} > 0$.

First, observe that the best responses of the bystander and the ruler are still described by equations (1) and (2). Hence, Lemma 1 applies. Hence, equation (2) implies that $\pi \le \alpha_g$. 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 $q' = q/(q+(1-q)(1-C))$. 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 (4) implies:

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

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

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

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

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, (4) implies:

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

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

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

which implies:

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

Hence,

$$\rho_g^B \in [0,1] \Leftrightarrow \frac{(\beta+\beta_g)p}{\beta_b + \beta_g p}\Big(1 - \frac{C(\beta_b - p\beta)}{\beta_b - p\beta + (1-C)p(\beta+\beta_g)}\Big) \leq q \leq \frac{\beta+\beta_g}{\beta_b+\beta_g}\Big(1 - \frac{C(\beta_b - \beta)}{\beta_b - \beta + (1-C)(\beta+\beta_g)}\Big),$$

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

Suppose that $\pi = 0$. Equation (2) implies that $\rho_b^G = 1$ and $\rho_g^B = 1$. Hence, (4) implies:

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

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

$$\beta_g(1 - \frac{q}{q+p(1-q)(1-C)}) - \beta_b \frac{q}{q+p(1-q)(1-C)} \leq -\beta \frac{p[q+(1-q)(1-C)]}{p[q+(1-q)(1-C)] + (1-p)q},$$

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

Proof of Corollary 4. In the intermediate and high repression equilibria with international pressure, from Proposition 3, the likelihood that the activist is repressed is $p(1-q)(1-C)$. Absent international pressure, from Proposition 2, in the low repression equilibrium, the corresponding likelihood is $p(1-q)\rho_g^B = p(1-q)\frac{\beta_b - \beta}{\beta_g + \beta}\frac{q}{1-q}$. Moreover,

$$\frac{\beta_b - \beta}{\beta_g + \beta}\frac{q}{1-q} < 1 - C \text{ if and only if } q < \frac{(1-C)(\beta_g+\beta)}{(1-C)\beta_g + \beta_b - C\beta} = g(1)q_2.$$

Further, the likelihood that the bystander protests is 0 in the high repression equilibrium with international pressure, δ_b in the intermediate repression equilibrium with international pressure, and $\alpha_g > \delta_b$ in the low repression equilibrium (without pressure). The result follows. ■

Proof of Corollary 5. Observe that, without international pressure, the low repression equilibrium is unique in region I. In this equilibrium, the good reform is blocked with probability $p(1-q)\rho_g^B(1-\alpha_g) = p(1-q)\frac{\beta_b-\beta}{\beta_g+\beta}\frac{q}{1-q}(1-\alpha_g)$. Moreover, the high repression equilibrium with international pressure is also unique in region I. In this equilibrium the good reform is blocked with probability $p(1-q)(1-C)$. Further, the boundaries of region I, $g(1)q_2$ and $q_1(p)$, do not depend on α_g. Thus, in region I, the likelihood that good reforms are blocked is higher with international pressure if and only if:

$$\frac{\beta_b - \beta}{\beta_g + \beta}\frac{q}{1-q}(1-\alpha_g) < 1 - C. \tag{5}$$

From the proof of Corollary 4, for $\alpha_g = 0$, $\frac{\beta_b-\beta}{\beta_g+\beta}\frac{q}{1-q} > 1-C$ if and only if $q > g(1)q_2$, which includes region I. Further, the left hand side of (5) is strictly decreasing in α_g, with $\rho_g^B(1-C) < 1-C$ at $\alpha_g = C$. Thus, there exists $\hat{\alpha}_g \in (0, C)$ such that $\frac{\beta_b-\beta}{\beta_g+\beta}\frac{q}{1-q}(1-\alpha_g) < 1-C \;\Leftrightarrow\; \alpha_g > \hat{\alpha}_g$. ■

Proof of Proposition 4: The results follow from the following Lemma together with our discussion in the text.

Lemma 3 *Consider a society with parameters (p,q) such that $q < q_1(p)$. There exists $0 < \underline{C}(p,q) < \overline{C}(p,q) < 1$ with $p(1-q)(1-\underline{C}) > R_g(0)$ and $p(1-q)(1-\overline{C}) = R_g(0)$ such that:*

- *If either intermediate or high repression equilibrium with international pressure is selected in region II, then*
$$R_g(C) = \begin{cases} R_g(0) & \text{if } C < \underline{C} \\ p(1-q)(1-C) & \text{if } C > \underline{C}. \end{cases}$$
- *If the low repression equilibrium with international pressure is selected in region II, then*
$$R_g(C) = \begin{cases} R_g(0) & \text{if } C \leq \overline{C} \\ p(1-q)(1-C) & \text{if } C \geq \overline{C}. \end{cases}$$

Proof of Lemma 3. From the proof of Proposition 3, $g(p,C) = 1 - C\frac{\beta_b - p\beta}{\beta_b - p\beta + (1-C)p(\beta+\beta_g)}$, and hence $\frac{\partial g(p,C)}{\partial C} = -\frac{(\beta_b - p\beta)(\beta_b + p\beta_g)}{(\beta_b + p\beta_g - Cp(\beta+\beta_g))^2} < 0$. That is, $g(p,C)$ is strictly decreasing in C on $C \in [0,1]$ and for all $p \in [0,1]$, with $g(p,0) = 1$ and $g(p,1) = 0$. Thus, for a given p, if $q < q_1(p)$, (i) there exists a unique $\underline{C}$ that satisfies $g(p,\underline{C})q_1(p) = q$, and (ii) there exists a unique $\overline{C}$ that satisfies $g(1,\overline{C})q_2 = g(1,\overline{C})q_1(1) = q$, and (iii) $0 < \underline{C} < \overline{C} < 1$. Thus, given a p we have: $q < g(p,C)q_1(p)$ for $C < \underline{C}$; $q = g(p,\underline{C})q_1(p)$; $g(p,C)q_1(p) < q < g(1,C)q_1(1)$ for $\underline{C} < C < \overline{C}$; $q = g(1,\overline{C})q_1(1)$; and $q > g(1,C)q_1(1)$ for $C > \overline{C}$. Thus, a (p,q) with $q < q_1(p)$ is below region II when $C \in (0,\underline{C})$, in region II when $C \in (\underline{C},\overline{C})$, and in region I when $C \in (\overline{C},1)$.

From Proposition 3, in the intermediate and high repression equilibrium with international pressure, the likelihood that the good activist is repressed $R_g(C)$ is $p(1-q) \times (1-C)$; and in the low repression equilibrium with international pressure, that likelihood is $p(1-q) \times \frac{\beta_b-\beta}{\beta_g+\beta}\frac{q}{1-q}\frac{1}{1-C} \times (1-C) = p(1-q) \times \frac{\beta_b-\beta}{\beta_g+\beta}\frac{q}{1-q}$, which is the same as the likelihood that the good activist is repressed in the low repression equilibrium without international pressure $R_g(0)$. Moreover, from the proof of Proposition 3, $\frac{\beta_b-\beta}{\beta_g+\beta}\frac{q}{1-q}\frac{1}{1-C}$ is strictly increasing in q for $q < g(1,C)q_2$, and becomes 1 at $q = g(1,C)q_2$. Hence, $R_g(0) < p(1-q)(1-C)$ for $q < g(1,C)q_2$, and $R_g(0) = p(1-q)(1-C)$ at $q = g(1,C)q_2$, i.e., as C increases $p(1-q)(1-C)$ crosses $R_g(0)$ from above at $\overline{C}$. ■

9.3 PROOFS FOR SECTION 6: ENDOGENOUS LIMITS ON REPRESSION

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:

$$(6) \qquad B(r_b, r_g, \pi) = q\;[r_b(1-\pi) + (1-r_b)(1-\alpha_b)] + (1-q)\;[r_g(1-\pi) + (1-r_g)(1-\alpha_g)],$$

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

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:*

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

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

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

Lemma 4 *Fix the bad ruler's strategy (r_b, r_g). The following characterizes the equilibria 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 5 *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 4. 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 $(r_b, r_g) \neq (1, 1)$. ■

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 11.

Lemma 6 *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 (6), 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 (6), 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:

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

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 7 *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 5 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 6 implies that R_0 dominates any strategy for which $F(r_b, r_g) \leq K$, and Lemma 7 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 6 implies that R_1 dominates any strategy for which $F(r_b, r_g) \leq K$. Because $B_1 > B_0$, Lemma 7 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 6,

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

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 6 implies that R_0 dominates any strategy for which $F(r_b, r_g) \leq K$, and Lemma 7 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 6 implies that R_2 dominates any strategy for which $F(r_b, r_g) \leq K$. Because $B_2 > B_0$, Lemma 7 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 6,

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

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 $\pi = 0$. ■

9.3.1 Proofs Related to Proposition 6

Terminology. To ease exposition of the results, we introduce the following terminology. We refer to the game in which the bad ruler can commit to a repression strategy (r_b, r_g) as *the game with precise commitment*. The game in which the ruler's attempt to repress the good activist is blocked with an exogenous probability is called *the game with exogenous* C. We refer to the game in which the bad ruler can ex ante choose this probability as *the game with endogenous* C. With this terminology, Proposition 6 concerns the equilibria of the game with endogenous C. The equilibria of the game with precise commitment are presented in Proposition 5 and the equilibria of the game with exogenous C are presented in Proposition 3. Moreover, we refer to the *high, intermediate, and low repression equilibrium with international pressure* as *high, intermediate, and low repression equilibrium with exogenous blocking*, accordingly. An equilibrium of the game with precise commitment is *outcome-equivalent* to an equilibrium of the game with exogenous C or an equilibrium of the game with endogenous C if the probability that each type of ruler successfully represses each type of activist, and the probability that the bystander protests following repression are identical.

Equilibrium Selection with Endogenous C. The subgame following the bad ruler's choice of C is identical to the game with exogenous C, and the equilibria of this subgame are characterized in Proposition 3. For $q \in [q_1(p)g(p), q_2 g(1)]$, multiple equilibria exist in the subgame. We consider the following selection in these cases: if $q = q_2 g(1)$, then we select the *high repression equilibrium with exogenous blocking*, and if $q \in [q_1(p)g(p), q_2 g(1))$, then we select the *intermediate repression equilibrium with exogenous blocking.*

Lemma 8 *Fix a choice of C in the game with endogenous C. There exists an (r_b, r_g) such that the bad ruler's payoff of selecting (r_b, r_g) in the game with precise commitment is weakly greater than the bad ruler's payoff of selecting C in the game with endogenous C.*

Proof. We divide the proof into three cases: $q \in (0, q_1(p)g(p))$, $q \in [q_1(p)g(p), q_2 g(1))$, and $q \in [q_2 g(1), 1)$, where we recognize that $g(p)$ depends on C.

Case I. Suppose C is such that $q < q_1(p)g(p)$, that is,

$$q < q_1(p) \left(1 - C\frac{\beta_b - p\beta}{\beta_b - p\beta + (1-C)p(\beta + \beta_g)}\right) \Leftrightarrow C < 1 - \frac{q}{1-q}\frac{\beta_b - p\beta}{p(\beta + \beta_g)}.$$

Following this choice of C, the *low repression equilibrium with exogenous blocking* is the unique equilibrium of the subgame, and it generates an expected payoff of $\hat{B}(C) = 1 - \alpha_g$. Suppose that in the game with precise commitment, the bad ruler chooses $(r_b, r_g) = (1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta})$, so that $F(r_b, r_g) = 0$. Then, in the subgame following $(r_b, r_g) = (1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta})$, we have $\pi = 0$. Hence, from equation (6), the bad ruler's expected payoff is $B(1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}, 0) = q + (1-q)\left[\frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta} + (1 - \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta})(1 - \alpha_g)\right] = 1 - \alpha_g + \frac{q}{q_2}\alpha_g > \hat{B}(C)$.

Case II. Suppose C is such that $q_1(p)g(p) \le q < q_2 g(1)$, that is,

$$q_1(p) \left(1 - C\frac{\beta_b - p\beta}{\beta_b - p\beta + (1-C)p(\beta + \beta_g)}\right) \le q < q_2 \left(1 - C\frac{\beta_b - \beta}{\beta_b - \beta + (1-C)(\beta + \beta_g)}\right) \Leftrightarrow$$
$$1 - \frac{q}{1-q}\frac{\beta_b - p\beta}{p(\beta + \beta_g)} \le C < 1 - \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}.$$

Following this choice of C, the *intermediate repression equilibrium with exogenous blocking* is selected in the subgame, and it generates an expected payoff of $\hat{B}(C) = q(1 - \delta_b) + (1-q)\left[(1 - C)(1 - \delta_b) + C(1 - \alpha_g)\right]$. Suppose that in the game with precise commitment, the bad ruler chooses $(r_b, r_g) = (1, 1 - C)$. To determine which equilibrium arises in the subgame following this choice, we calculate $F(1,1-C)=(1-q)(\beta_g+\beta)(1-C)-q(\beta_b-\beta)$. Note that

$$\begin{aligned}F(1,1-C) &> (1-q)(\beta_g+\beta)\frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}-q(\beta_b-\beta)=0,\\ F(1,1-C) &\leq (1-q)(\beta_g+\beta)\frac{q}{1-q}\frac{\beta_b-p\beta}{p(\beta_g+\beta)}-q(\beta_b-\beta)=\frac{1-p}{p}q\beta_b=K,\end{aligned}$$

where the last equality follows from the definition of K in Lemma 2. Hence, $\pi=\delta_b$ in the subgame following the choice of $(r_b,r_g)=(1,1-C)$, and hence, from equation (6), the bad ruler's payoff is $B(1,1-C,\delta_b)=q(1-\delta_b)+(1-q)\ [(1-C)(1-\delta_b)+C(1-\alpha_g)]=\hat{B}(C)$.

Case III. Suppose C is such that $q\geq q_2g(1)$, that is,

$$q\geq q_2\ \left(1-C\frac{\beta_b-\beta}{\beta_b-\beta+(1-C)(\beta+\beta_g)}\right)\ \Leftrightarrow\ C\geq 1-\frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}.$$

Following this choice of C, the *high repression equilibrium with exogenous blocking* is the unique equilibrium of the subgame,[44] and generates an expected payoff of $\hat{B}(C)=q+(1-q)(1-C)$. Suppose that in the game with precise commitment, the bad ruler chooses $(r_b,r_g)=(1,1-C)$. To determine which equilibrium arises in the subgame following this choice, we calculate $F(1,1-C)=(1-q)(\beta_g+\beta)(1-C)-q(\beta_b-\beta)$. Note that

$$F(1,1-C)\leq(1-q)(\beta_g+\beta)\frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}-q(\beta_b-\beta)=0,$$

and hence, $\pi=0$ in the subgame following $(r_b,r_g)=(1,1-C)$. Thus, from equation (6), the bad ruler's payoff is $B(1,1-C,0)=q+(1-q)(1-C)=\hat{B}(C)$. ■

Lemma 9 *Fix a pair of priors (p,q). Let (r_b,r_g) be the <u>equilibrium</u> strategy of the bad ruler in the game with precise commitment. In the game with endogenous C, suppose the bad ruler chooses $C=1-r_g$. Then, the equilibrium of the subgame following this choice of C is outcome-equivalent to the equilibrium of game with precise commitment.*

Proof. Proposition 5 characterizes the equilibrium choices of (r_b,r_g). First, consider case 1 of Proposition 5, where $(r_b,r_g)=(1,1)$ in equilibrium. Let $C=0$. Then, both in the equilibrium of the game with endogenous C and in the equilibrium of the game with precise commitment, the good activist is always successfully repressed by the bad ruler, the bad activist is always repressed by both types of the ruler, and the bystander never protests.

Second, consider case 2 in Proposition 5, which corresponds to $q_1(p)<q<q^*<q_2$. Let

[44] If $q=q_2g(1)$, this equilibrium is selected by our equilibrium selection described above.

$C = 1 - r_g = 0$, and observe that, $C = 0$ implies $g(p) = 1$, so that $[q_1(p)g(p), q_2 g(1)) = [q_1(p), q_2]$. Given our equilibrium selection, when $q \in [q_1(p)g(p), q_2 g(1)) = [q_1(p), q_2)$, the *intermediate repression equilibrium with exogenous blocking* (from Proposition 3) is selected. For $C = 0$, this equilibrium is identical to the *intermediate repression equilibrium* (from Proposition 2) that obtains in the game with precise commitment.

Third, consider case 3 in Proposition 5, where $r_b = 1$ and $r_g = \frac{q}{1-q}\frac{\beta_b - p\beta}{p(\beta_g+\beta)}$ in equilibrium. Let

$$C = 1 - \frac{q}{1-q}\frac{\beta_b - p\beta}{p(\beta_g + \beta)}. \tag{7}$$

Observe that this choice of C implies $q = q_1(p)g(p)$. Given our equilibrium selection, with a C such that $q = q_1(p)g(p)$, the equilibrium of the subgame is *the intermediate repression equilibrium with exogenous blocking* in which $\pi = \delta_b$, $\rho_g^B = 1$, $\rho_b^B = 1$, $\rho_g^G = 0$, and

$$\rho_b^G = \frac{p}{1-p}\Big(\frac{(\beta + \beta_g) - (\beta_b + \beta_g)q}{\beta_b q} - \frac{1-q}{q}\frac{\beta + \beta_g}{\beta_b}\ C\Big). \tag{8}$$

Substituting from equation (7) to (8) yields $\rho_b^G = 1$. Moreover, in the game with endogenous C, the probability that repression against the good activist succeeds is $\rho_g^B \times (1 - C) = \frac{q}{1-q}\frac{\beta_b - p\beta}{p(\beta_g+\beta)}$, which is the same as the equilibrium level of r_g in the game with precise commitment. Thus, these equilibria are outcome-equivalent.

Fourth, consider case 4 in Proposition 5, where $r_b = 1$ and $r_g = \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}$ in equilibrium. Let $C = 1 - \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}$. Observe that this choice of C implies $q = q_2 g(1)$. Given our equilibrium selection, with a C such that $q = q_1(p)g(1)$, the equilibrium of the subgame is the *high repression equilibrium with exogenous blocking* in which $\rho_g^G = 0$, $\rho_b^G = \rho_g^B = \rho_b^B = 1$ and $\pi = 0$. In this equilibrium, $\rho_g^B = 1$, and hence the probability that repression against the good activist succeeds is $\rho_g^B \times (1 - C) = r_g$. Hence, this equilibrium is outcome-equivalent to equilibrium of the game with precise commitment. ■

Lemma 10 *The equilibrium of the game with endogenous C is generically unique.*

Proof. In the game with endogenous C, the bad ruler's payoff of selecting C is

$$\hat{B}(C) = q(1 - \pi(C)) + (1 - q)\ [(1 - C)\rho_g^B(C)(1 - \pi) + (1 - (1 - C)\rho_g^B(C))(1 - \alpha_g)],$$

where $\pi(C)$ and $\rho_g^B(C)$ depend on the equilibrium of the subgame following the choice of C.

Consider $q \in (0, q_1(p))$. First, note that

$$q \geq q_2 g(1) \Leftrightarrow C \geq C_H \equiv 1 - \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}, \quad \text{and} \quad C_H \in (0,1).$$

Second, note that

$$q \geq q_1(p)g(p) \Leftrightarrow C \geq C_L \equiv 1 - \frac{q}{1-q}\frac{\beta_b - p\beta}{p(\beta_g + \beta)}, \quad \text{and} \quad C_L \in (0, C_H).$$

It follows that (1) for $C < C_L$ the *low repression equilibrium with exogenous blocking* is unique in the subgame following the bad ruler's initial choice, generating $\pi(C) = \alpha_g$ and $\rho_g^B(C) = \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}\frac{1}{1-C}$, (2) for $C_L \leq C < C_H$ the *intermediate repression equilibrium with exogenous blocking* is selected in the subgame following the bad ruler's initial choice, generating $\pi(C) = \delta_b$ and $\rho_g^B(C) = 1$, and (3) for $C \geq C_H$ *high repression equilibrium with exogenous blocking* is selected, generating $\pi(C) = 0$ and $\rho_g^B(C) = 1$. Hence, for $C < C_L$, the bad ruler's payoff is $\hat{B}(C) = 1 - \alpha_g$, which does not depend on C. At $C = C_L$, the protest probability $\pi(C)$ jumps down from α_g to δ_b, and it is δ_b for all $C \in [C_L, C_H)$; similarly, $\rho_g^B(C)$ jumps up from $\frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}\frac{1}{1-C}$ to 1, and it is 1 for $C \geq C_L$. Hence, $\hat{B}(C)$ has an upward jump discontinuity at $C = C_L$ at which it is right continuous, and for $C \in [C_L, C_H)$, $\hat{B}(C)$ is a decreasing linear function of C. At $C = C_H$, the protest probability $\pi(C)$ jumps down from δ_b to 0, and it is 0 for all $C \geq C_H$. Hence, $\hat{B}(C)$ has an upward jump discontinuity at $C = C_H$ at which it is right continuous, and for $C \geq C_H$, $\hat{B}(C)$ is a decreasing linear function of C. Hence, only $C = 0$, $C = C_L$, or $C = C_H$ could be optimal. Next, note that $\hat{B}(0)$ does not depend on q, and $\hat{B}(C_H)$ are both linear functions of q with different slopes, and hence each of the following equations holds for no more than a single value of q: $\hat{B}(0) = \hat{B}(C_H)$, $\hat{B}(0) = \hat{B}(C_L)$, or $\hat{B}(C_L) = \hat{B}(C_H)$. Hence, outside of knife-edge cases, the equilibrium value of C is unique. The remaining cases, $q > q_2$ and $q \in (q_1(p), q_2)$, are analogous and simpler. ■

Proof of Proposition 6. Lemma 8 implies that the bad ruler's equilibrium payoff in the game with precise commitment is an upper bound for his equilibrium payoff in the game with endogenous C. Lemma 9 shows that for each (p, q), the bad ruler can always select a C to achieve this upper bound. Thus, such a C must be an optimal choice. Moreover, Lemma 9 also shows that when the bad ruler selects this optimal C, the equilibrium of the game is outcome-equivalent to the equilibrium of the game with precise commitment. Finally, Lemma 10 show that the optimal C is unique. ■

9.4 EQUILIBRIUM SELECTION IN THE GAME WITH COMMITMENT

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

ES1. From Lemma 4 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 4, 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 4 (the others are considered in Propositions 8, 9). 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 maximum value. 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 $(r_b^*, r_g^*) \equiv (1, \frac{q}{1-q} \frac{\beta_b - p\beta}{p(\beta + \beta_g)})$, yielding the bad ruler's payoff $B^* \equiv B(r_b^*, r_g^*, \delta_b)$. This is the only instance in which the bad ruler's equilibrium choice is such that $F(r_b, r_g) = K$.

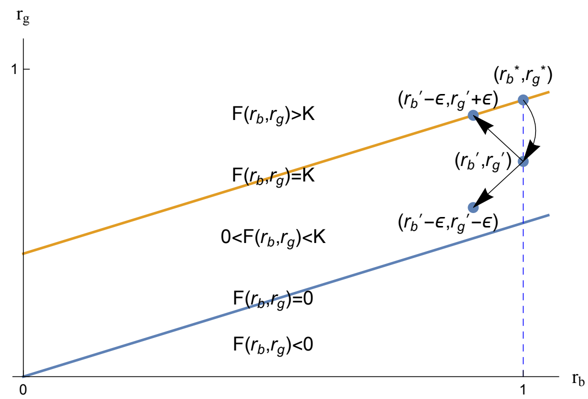

Figure 6: The nature of the deviation from (r_b^*, r_g^*) to (r_b', r_g').

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:* $\lim_{\epsilon\to 0} B^*(\epsilon) = B^*$
2. *The protest probability converges to δ_b:* $\lim_{\epsilon\to 0} \Pr\{\pi^*(R(r_b^*(\epsilon)), R(r_g^*(\epsilon))) = \delta_b\} = 1$.
3. *The bad ruler's strategy converges to (r_b^*, r_g^*):* $\lim_{\epsilon\to 0} r_i^*(\epsilon) = r_i^*$ *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

$$B^* > B^*(\epsilon). \tag{9}$$

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 6). For 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 $k \equiv 2 + \frac{q}{1-q}\frac{\beta_b - \beta}{\beta+\beta_g}$, so that $B(r_b^* - \epsilon, r_g^* - k\epsilon, \delta_b) = B(r_b' - \epsilon, r_g' - \epsilon, \delta_b)$. Then,

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

The first inequality follows from monotonicity properties of B, the second inequality follows from optimality of $B^*(\epsilon)$, and the third is (9). 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:

$$\begin{aligned} B^*(\epsilon) = {} & \Pr\{\pi^*(R_b^*, R_g^*) = 1\} E[B(R_b^*, R_g^*, 1) | \pi^*(R_b^*, R_g^*) = 1] \\ & + \Pr\{\pi^*(R_b^*, R_g^*) = \delta_b\} E[B(R_b^*, R_g^*, \delta_b) | \pi^*(R_b^*, R_g^*) = \delta_b] \\ & + \Pr\{\pi^*(R_b^*, R_g^*) = 0\} E[B(R_b^*, R_g^*, 0) | \pi^*(R_b^*, R_g^*) = 0]. \end{aligned} \tag{10}$$

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 $B(\hat{r}_b, \hat{r}_g, \delta_b) < B(r_b^*, r_b^*, \delta_b) = B^*$; 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 (10) 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:

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

From part (1), $\lim_{\epsilon\to 0} B^*(\epsilon) = B^*$. From part (2), $\lim_{\epsilon\to 0} \Pr\{\pi^*(R_b^*, R_g^*) = \delta_b\} = 1$. Thus, (10) implies $\lim_{\epsilon\to 0} E[B(R_b^*, R_g^*, \delta_b) | \pi^*(R_b^*, R_g^*) = \delta_b] = B^*$. By continuity, $\lim_{\epsilon\to 0} r_i^*(\epsilon) = r_i^*$. ■

ES2 is justified by the following lemma.

Lemma 11 *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 $F(r_b, r_g) \leq 0$:

$$(11) \quad \max_{(r_b, r_g) \in [0,1]^2} B(r_b, r_g, \pi(r_b, r_g)) \quad \text{subject to} \quad F(r_b, r_g) \leq 0, \quad \text{and } \pi(r_b, r_g) = \pi \text{ if } F(r_b, r_g) = 0.$$

Lemma 6 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 7, 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, optimization (11) has no solution, and no equilibrium exists. ■

Appendix II: For Online Publication

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 8 *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 5 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 9 *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 5 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. ■

11 COMMITMENT BY BOTH THE GOOD AND THE BAD RULER

Suppose that before nature selects the faction that controls the state and before factions observe the activist's type, both the good and the bad faction simultaneously commit to their repression strategies (r_b^B, r_g^B) and (r_b^G, r_g^G), where r_i^J is the probability with which the type J ruler commits to repress the type i activist. The bystander observes both factions' repression strategies. Then, a faction gains control of the state, but the bystander does not observe which faction has gained control. As in the previous sections, the bad faction gains control with probability p and the good faction gains control with probability $1 - p$. The faction in power observes the activist's type, and represses or concedes according to its strategy. If the faction in power concedes, then the game ends. If the faction in power represses, then the bystander observes repression and decides whether to protest. Consider equilibria in which factions do not randomize their choice of repression probability (I think it can be proved that this is without loss of generality).

Let $U_B(r_b^B, r_g^B, \pi)$ be the bad faction's payoff from (r_b^B, r_g^B), anticipating that the bystander protests with probability π, and let $U_G(r_b^G, r_g^G, \pi)$ be the good faction's payoff:

$$U_B(r_b^B, r_g^B, \pi) = p\Big(q\ [r_b^B(1-\pi) + (1-r_b^B)(1-\alpha_b)] + (1-q)\ [r_g^B(1-\pi) + (1-r_g^B)(1-\alpha_g)]\Big),$$
$$U_G(r_b^G, r_g^G, \pi) = (1-p)\Big(q\ [r_b^G(1-\pi) + (1-r_b^G)(1-\delta_b)] + (1-q)[r_g^G(1-\pi) + (1-r_g^G)(1+\delta_g)]\Big).$$

Lemma 12 *Given the strategy of the bad ruler (r_b^B, r_g^B) and the strategy of the good ruler (r_b^G, r_g^G), the bystander's best response is:*

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

where $F_B(r_b^B, r_g^B) \equiv (1-q)(\beta_g + \beta)r_g^B - q(\beta_b - \beta)r_b^B$ and $F_G(r_b^G, r_g^G) \equiv K(r_b^G - \frac{(1-q)\beta_g}{q\beta_b} r_g^G)$, with $K \equiv \frac{1-p}{p} q\beta_b$.

Proof. Given the strategies of the good and bad ruler, the bystander's posterior belief is:

$$p' = \frac{p[qr_b^B + (1-q)r_g^B]}{p[qr_b^B + (1-q)\rho_g^B] + (1-p)(qr_b^G + (1-q)r_g^G)},$$
$$q' = \frac{q[(1-p)r_b^G + pr_b^B]}{q[(1-p)\rho_b^G + pr_b^B] + (1-q)[pr_g^G + (1-p)r_g^B]}.$$

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

Lemma 13 *Any strategy for the good ruler (r_b^G, r_g^G) such that $r_g^G > 0$ is strictly dominated by $(r_b^G, 0)$.*

Proof. Let the protest probability following (r_b^G, r_g^G) be π, and the protest probability following $(r_b^G, 0)$ be π'. Because $F_G(r_b^G, r_g^G)$ is decreasing in r_g^G, it follows that $\pi' \leq \pi$. Hence, $U_G(r_b^G, r_g^G, \pi) - U_G(r_b^G, 0, \pi') = (1-p)[q(\pi - \pi') + (1-q)(\pi + \delta_b)] > 0$. ■

Because all strategies for the good ruler with $r_g^G > 0$ are strictly dominated, we delete them from the game. The bystander's best response can be represented in the following simpler form.

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

where we have omitted the subscript B on function $F_B(r_b^B, r_g^B)$ to simplify notation.

Lemma 14 *(1) A repression strategy (r_b^B, r_g^B) satisfying $F(r_b^B, r_g^B) > K r_b^G$ gives the bad ruler a strictly smaller payoff than $(r_b^B, r_g^B) = (0, 0)$. (2) For the bad ruler a repression strategy (r_b^B, r_g^B) satisfying $F(r_b^B, r_g^B) > K$ is strictly dominated by $(r_b^B, r_g^B) = (0, 0)$.*

Proof.

(1) If $F(r_b^B, r_g^B) > K r_b^G$, then $\pi = 1$. Furthermore, $F(r_b^B, r_g^B) > K r_b^G$ implies that $F(r_b^B, r_g^B) \geq 0$, and hence $r_g^B > 0$. Hence, the bad faction's payoff of selecting such (r_b^B, r_g^B) is $U_B(r_b^B, r_g^B, 1) = p\Big(q(1 - r_b^B)(1 - \alpha_b) + (1 - q)(1 - r_g^B)(1 - \alpha_g)\Big)$. Hence, $r_g^B > 0$ implies that $U_B(r_b^B, r_g^B, 1) < p\Big(q(1 - \alpha_b) + (1 - q)(1 - \alpha_g)\Big)$. Observing that $p\Big(q(1 - \alpha_b) + (1 - q)(1 - \alpha_g)\Big)$ is the payoff of selecting $(r_b^B, r_g^B) = (0, 0)$ completes the proof.

(2) $F(r_b^B, r_g^B) > K \Rightarrow F(r_b^B, r_g^B) > K r_b^G$ for any $r_b^G \in [0, 1]$. From (1), any such strategy is strictly dominated by $(r_b^B, r_g^B) = (0, 0)$. ■

Because repression strategies (r_b^B, r_g^B) satisfying $F(r_b^B, r_g^B) > K$ are strictly dominated, we delete them from the game.

Let $\theta(r_b^B, r_g^B, r_b^G)$ represent the bystander's protest probability for ruler strategies (r_b^B, r_g^B, r_b^G) satisfying $F(r_b^B, r_g^B) = K r_b^G$.

Lemma 15 *(Good ruler's best responses)*

1. *The good ruler's best response to (r_b^B, r_g^B) such that $F(r_b^B, r_g^B) < K$ is $r_b^G = 1$.*

2. *The good ruler's best response to (r_b^B, r_g^B) such that $F(r_b^B, r_g^B) = K$ is*

$$r_b^G = \begin{cases} 1 & \textit{if} \quad \theta(r_b^B, r_g^B, 1) < \delta_b \\ 0 \textit{ or } 1 & \textit{if} \quad \theta(r_b^B, r_g^B, 1) = \delta_b \\ 0 & \textit{if} \quad \theta(r_b^B, r_g^B, 1) > \delta_b \end{cases}$$

Proof. (1) Suppose $F(r_b^B, r_g^B) < K$. If $r_b^G > \frac{F(r_b^B, r_g^B)}{K}$, then $\pi = 0$. Among $r_b^G > \frac{F(r_b^B, r_g^B)}{K}$ the best choice is $r_b^G = 1$, yielding payoff $U_G(1,0)$. Because $U_G(1,0) > U_G(r_b^G, \pi)$ for all possible $(r_b^G, \pi) \neq (1,0)$, the good ruler's best response is $r_b^G = 1$.

(2) Suppose $F(r_b^B, r_g^B) = K$. If $r_b^G < 1$, then $\pi = 1$. Among $r_b^G < 1$, the best choice is $r_b^G = 0$, yielding payoff $U_G(0,1)$. If $r_b^G = 1$, then $\pi = \theta'' \equiv \theta(r_b^B, r_g^B, 1)$, yielding payoff $U_G(1, \theta'')$. Because $U_G(1,\theta'') - U_G(0,1) = (1-p)q(\delta_b - \theta'')$, the result follows. ■

Lemma 16 *An equilibrium in which $r_b^G = 0$ does not exist.*

Proof. Suppose that such an equilibrium exists. Lemma 14 part (2) implies that any (r_b^B, r_g^B) such that $F(r_b^B, r_g^B) > 0$ is worse for the bad ruler than $(r_b^B, r_g^B) = (0,0)$, and it therefore cannot be the bad ruler's best response. Hence, it must be that the bad ruler's equilibrium strategy satisfies $F(r_b^B, r_g^B) \leq 0$, but Lemma 15 implies that the best response to such (r_b^B, r_g^B) is $r_b^G = 1$. ■

Lemma 17 *(Bad ruler's best response). For $q < q_1(p)$, let $\hat{r}_1$ be such that $F(1, \hat{r}_1) = K$.*

1. *If $q > q_1(p)$, then the bad ruler's best response to $r_b^G = 1$ is $(r_b^B, r_g^B) = (1,1)$.*
2. *If $q < q_1(p)$ and $\theta(1, \hat{r}_g, 1) = 0$, then the bad ruler's best response to $r_b^G = 1$ is $(r_b^B, r_g^B) = (1, \hat{r}_g)$.*
3. *If $q < q_1(p)$ and $\theta(1, \hat{r}_g, 1) > 0$, then the bad ruler's best response to $r_b^G = 1$ does not exist.*

Proof. (1) If $q > q_1(p)$, then $F(1,1) < K$. Hence, if the bad ruler selects $(r_b^B, r_g^B) = (1,1)$, then the bystander's best response is $\pi = 0$. The bad ruler's payoff $U_B(1,1,0) > U_B(r_b^B, r_g^B, \pi)$ for all $(r_b^B, r_g^B, \pi) \neq (1,1,0)$, and hence it is the bad ruler's best response.

(2) By definition, $F(1, \hat{r}_g) = K$. Because $\theta(1, \hat{r}_g, 1) = 0$ the bad ruler's payoff of choosing $(r_b^B, r_g^B) = (1, \hat{r}_g)$ is $U_B(1, \hat{r}_g, 0)$. Suppose that the bad ruler chooses (r_b^B, r_g^B) such that $F(r_b^B, r_g^B) < K$. Following such a choice $\pi = 0$, yielding payoff $U_B(r_b^B, r_g^B, 0)$. Furthermore, $F(1, \hat{r}_g) - F(r_b^B, r_g^B) = (1-q)(\beta_g + \beta)(\hat{r}_g - r_g^B) - (1 - r_b^B)q(\beta_b - \beta) > 0$. Because $r_b^B \leq 1$, it follows that $r_g^B < \hat{r}_g$. Hence, $U_B(1, \hat{r}_g, 0) - U_B(r_b^B, r_g^B, 0) = p(q\alpha_b(1 - r_b^B) + \alpha_g(\hat{r}_g - r_g^B)(1-q) > 0$, and hence, $(1, \hat{r}_g)$ dominates any such choice. Next, suppose that the bad ruler chooses $(r_b^B, r_g^B) \neq (1, \hat{r}_g)$ such that $F(r_b^B, r_g^B) = K$. Following such a choice $\pi = \theta(r_b^B, r_g^B, 1)$, yielding payoff $U_B(r_b^B, r_g^B, \theta)$. Furthermore, $F(1, \hat{r}_g) - F(r_b^B, r_g^B) = (1-q)(\beta_g + \beta)(\hat{r}_g - r_g^B) - (1 - r_b^B)q(\beta_b - \beta) = 0$. Because $r_b^B < 1$, it follows that $r_g^B < \hat{r}_g$. Hence, $U_B(1, \hat{r}_g, 0) - U_B(r_b^B, r_g^B, \theta) = p(q(\alpha_b(1 - r_b^B) + \theta r_b^B) + \alpha_g((\hat{r}_g - r_g^B) + r_g^B\theta)(1-q) > 0$, and hence, $(1, \hat{r}_g)$ dominates any such choice.

(3) Consider a strategy (r_b^B, r_g^B) satisfying $F(r_b^B, r_g^B) < K$. Following any such strategy the bystander does not protest, and hence, any such strategy delivers payoff $U_B(r_b^B, r_g^B, 0)$. If $r_b^B < 1$, then increasing r_b^B by ϵ, allows the bad ruler to maintain $\pi = 0$ and increases his payoff to $U_B(r_b^B + \epsilon, r_g^B, 0)$. If $r_b^B = 1$, then $r_g^B < \hat{r}_g$, and hence, increasing r_g^B by ϵ allows the bad ruler to maintain $\pi = 0$ and increases his payoff to $U_B(1, r_g^B + \epsilon, 0)$. Consider a strategy (r_b^B, r_g^B) satisfying $F(r_b^B, r_g^B) = K$ with $r_b^B < 1$. Following such a strategy the bystander protests with probability $\theta(r_b^B, r_g^B, 1)$, yielding payoff $U_B(r_b^B, r_g^B, \theta)$. Increasing r_b^B by sufficiently small ϵ ensures that $\pi = 0$, yielding a payoff of $U_B(1, r_g^B + \epsilon, 0) > U_B(r_b^B, r_g^B, \theta)$. If $(r_b^B, r_g^B) = (1, \hat{r}_g)$, then the bystander protests with probability $\theta(1, \hat{r}_g, 1)$, yielding payoff $U_B(1, \hat{r}_g, \theta)$. Suppose that bad ruler selects $(r_b^B, r_g^B) = (1, \hat{r}_g - \epsilon)$. Because $F(1, \hat{r}_g - \epsilon) < K$, following this choice $\pi = 0$, yielding payoff $U_B(1, \hat{r}_g - \epsilon, 0)$. For ϵ sufficiently small $U_B(1, \hat{r}_g - \epsilon, 0) > U_B(1, \hat{r}_g, \theta)$. ■

Proposition 10 *Suppose that the good and bad faction simultaneously commit to their repression strategies.*

1. *If $q > q_1(p)$, then in the unique equilibrium $(r_b^G, r_g^G) = (1, 0)$, $(r_b^B, r_g^B) = (1, 1)$, and $\pi = 0$*

2. *If $q < q_1(p)$, then in the unique equilibrium $(r_b^G, r_g^G) = (1, 0)$, $(r_b^B, r_g^B) = (1, \hat{r}_g)$, and $\pi = 0$.*

Proof. From Lemma 13, $r_g^G = 0$. Lemma 15 implies that the good ruler's best response can only be either $r_b^G = 0$ or $r_b^G = 1$, and Lemma 16 rules out the possibility that the good ruler selects $r_b^G = 0$ in equilibrium. Hence, $r_b^G = 1$.

(1) From Lemma 17, the bad ruler's unique best response to $r_b^G = 1$ is $(r_b^B, r_g^B) = (1, 1)$. If $q > q_1(p)$, then $F(1, 1) < K$, and hence Lemma 15 and Lemma 13 imply that the good ruler's unique best response to $(r_b^B, r_g^B) = (1, 1)$ is $r_b^G = 1$.

(2) From Lemma 17, if $\theta(1, \hat{r}_g, 1) > 0$, then the bad ruler's best response does not exist. Hence, an equilibrium must have $\theta(1, \hat{r}_g, 1) = 0$. Hence, from Lemma 16 the bad ruler's unique best response to $r_b^G = 1$ is $(r_b^B, r_g^B) = (1, \hat{r}_g)$. Because $F(1, \hat{r}_g) = K$ and $\theta(1, \hat{r}_g, 1) = 0$, the good ruler's best response to $(r_b^B, r_g^B) = (1, \hat{r}_g)$ is $r_b^G = 1$. ■