

# Information Manipulation and Reform in Authoritarian Regimes\*

JIDONG CHEN AND YIQING XU

*We develop a theory of how an authoritarian regime interactively uses information manipulation, such as propaganda or censorship, and policy improvement to maintain social stability. The government can depict the status quo policy more popularly supported than it actually is, while at the same time please citizens directly by enacting a costly reform. We show that the government's ability of making policy concessions reduces its incentive to manipulate information and improves its credibility. Anticipating a higher chance of policy concessions and less information manipulation, citizens are more likely to believe the government-provided information and support the regime. Our model provides an explanation for the puzzling fact that reform coexists with selective information disclosure in authoritarian countries like China.*

How does an authoritarian government like China's govern its people and manage to stay in power? Providing better policies (through reform) and manipulating information citizens receive (through propaganda and censorship) are two obvious answers and have been separately studied by many scholars.<sup>1</sup> An authoritarian regime can provide citizens with policy improvement including economic benefits, reducing their dissatisfaction against the regime and preventing them from coordinating with each other in the political domain (see e.g., Oi 2003; Bueno de Mesquita and Downs 2005; Cox 2009; Svobik 2012; Dimitrov 2013; Miller 2015). Alternatively, it can also use propaganda or censorship to distort citizens' incentives of joining the protest (see, e.g., Shirk 2010; Shadmehr and Bernhardt 2011; Stockmann and Gallagher 2011; King, Pan and Roberts 2013; Stockmann 2013; Dimitrov 2014; King, Pan and Roberts 2014; Lorentzen, Landry and Yasuda 2014; Wallace 2014; Huang 2015; Crabtree, Fariss and Kern 2015; Wallace 2015). The Chinese government seems to be using both strategies. It sustains unprecedented economic growth in the past three decades, which has greatly lifted the living standard of ordinary Chinese citizens. It also selectively censors the information citizens receive and manipulates it in a way that benefits the regime.<sup>2</sup>

\* Jidong Chen is an Assistant Professor in the Beijing Normal University Business School, Rear Main Building, Beijing Normal University, 19 Xijiekouwai St., Beijing 100875, China; Visiting fellow, Wallis Institute of Political Economy, University of Rochester, 107 Harkness Hall, Box 027-0158, University of Rochester, Rochester, NY 14627, USA (jidongc@princeton.edu). Yiqing Xu is a PhD Candidate in the Department of Political Science, Massachusetts Institute of Technology, 77 Mass. Ave, E53-470, Cambridge, MA 02142, USA (xyq@mit.edu). The authors are indebted to Jonathan Hanson, Mehdi Shadmehr and two anonymous referees for their helpful comments and suggestions.

<sup>1</sup> Propaganda is commonly used to describe the regime's purposeful communication toward its citizens in order to change their attitudes, whereas censorship is commonly understood as a selective information disclosure. We focus on the common feature of the two types of information manipulation in this paper. Specifically we see both of them as ways regime uses to change citizens' belief of a societal fundamental.

<sup>2</sup> For instance, the Chinese government not only deletes negative online news that can potentially spur collective action (King, Pan and Roberts 2013), but also hires internet commentators to post favorable comments toward government policies as a way to sway public opinion (Kalathil and Boas 2003). In addition, the local

However, how an authoritarian government can use reform and information manipulation interactively is less studied in the literature.

In this paper, we develop a model to understand how an authoritarian government can simultaneously use both instruments, enacting reform and manipulating information, to prevent citizens' collective action against the regime and stay in power. The simple game-theoretic framework has one government actor and two citizens. The government receives a signal of the social fundamental, which measures the (un)popularity of the status quo policy. It then chooses to (1) release a message as an attempt to manipulate citizens' beliefs about the fundamental and (2) make policy concessions to please the citizens directly. Specifically, the government chooses the extent to which it depicts the status quo policy more popularly supported than it actually is. Though the citizens do not directly observe the level of information distortion, they rationally internalize the government's incentives when forming their beliefs. Enacting reform, on the one hand, directly appeases citizens' anger and prevents them from joining collective action; on the other, partially reveals the government's private information of the social fundamental. After receiving the government's message and observing the policy outcome, both citizens simultaneously decide whether to participate in collective action demanding their desired policy.<sup>3</sup>

The main contribution of this paper is to illustrate a channel through which the government's decisions of making policy concessions and manipulating information are linked together. The government's higher ability of policy adjustment reduces its incentive to manipulate information, thus making its words about the popularity of the status quo more believable. As a result, citizens are more likely to abstain from joining collective action when they are told that the status quo is supported by many. The risks of collective action hence decrease.

Consider the following two scenarios. When the cost of reform is very high, as the citizens know that revealing an unpopular status quo is a dominated strategy for a government that never reforms, the government finds itself in a *babbling/pooling equilibrium* where it cannot persuade citizens not to join collective action even when the reform policy is unpopular.

When the cost of reform is relative low, however, the government may consider changing the status quo policy so as to avoid collective action. The government adopts the following strategy: when the reform is unpopular, the government will announce that the reform is unpopular (or equivalently, the status quo is popular); when the reform is popular, the government will sometimes implement the reform and admit that it is popular and sometimes maintain the status quo policy and lie to its citizens. In such a *partially informative equilibrium*, the government makes a tradeoff between cost of implementing the reform and maintaining the status quo while lying to the citizens to reduce the chance of collective action.

After we characterize the government's endogenous policy choice, a less surprising, though meaningful, result emerges: the probability of enacting the reform policy is increasing in the credibility of the threat of citizens' collective action, which in our model is captured by the probability that the citizens prefer the reform policy. In this highly simplified framework, if the government knows that the citizens are unable to collectively protest against a disliked policy, it will not make any effort to reform even though it clearly knows that a majority of the citizens favor it. However, if the threat of collective action is highly credible, it will push the

---

(Footnote continued)

governments in China also actively use their official websites to signal their attentiveness and concern toward the masses, as well as their economic and fiscal achievements (Pan 2013).

<sup>3</sup> Collective action can take form of either a small-scale protest or a revolutionary social movement that leads to a regime change. They differ in terms of the cost the government suffers from the collective action. In the rest of this paper, we use collective action and protest interchangeably.

government to change the policy. In addition, we show that the government's policy responsiveness also depends on the its policy-adjustment cost.

In an extension of the benchmark model, we consider a situation in which the government has commitment power when choosing strategies of information manipulation and enacting policies. The optimal mechanism from the government's perspective involves a moderate degree of policy improvement and a moderate degree of information manipulation. Commitment is required to sustain such a mechanism because the government has an *ex post* incentive to deviate. In the real world, a potential solution to the commitment problem is to institutionalize certain bureaucratic and legislative agencies (e.g., Nathan 2003; Lee and Zhang 2013; Truex 2015). The result of this paper implies that even if an authoritarian regime is able to control its agents, it lacks the incentive to hold them fully accountable. This finding complements the literature on the principal–agent problem of central–local relations in authoritarian countries (Egorov, Guriev and Sonin 2009; Lorentzen 2014). It is also consistent with the empirical findings that the official petition system helps mediate social conflicts only to a certain extent: in most cases, the authorities only address claims that may lead to social instability, which is often referred to as the strategy of “maintaining social stability” (*weiiwen*) by China scholars (Cai 2004; Chen 2012).

Our paper is closely related to Besley and Prat (2006) and Shadmehr and Bernhardt (2015), in which the authors study how the government manipulates citizens' political actions through the media. Our model differs from their models in two aspects. First, in our model, the government has an option to adjust the policy depending on cost that directly affects citizens' incentives in political actions, but there is no direct cost of lying.<sup>4</sup> Second, Besley and Prat (2006) and Shadmehr and Bernhardt (2015) focus on “hard information,” information that is verifiable after being disclosed, and the government's decision is whether to truthfully disclose this piece of information. Nevertheless, in our model, as in Edmond (2013) and Huang (2014), we allow “soft information,” information that is not verifiable by citizens, and consider the government's strategic choice of not only whether to disclose information but also how to disclose it.

This paper also contributes to a growing literature on the relationship between government manipulation of information and citizens' collective action (e.g., Bueno de Mesquita 2010; Hollyer, Rosendorff and Vreeland 2011; Shadmehr and Bernhardt 2011; Egorov and Sonin 2012; Little 2012; Edmond 2013; Lorentzen 2013; Casper and Tyson 2014; Dimitrov 2014; Gehlbach and Sonin 2014; Little 2014a; Little 2014b; Lorentzen 2014; Shadmehr 2014a; Shadmehr 2014b; Smith and Tyson 2014; Hollyer, Rosendorff and Vreeland 2015a; Hollyer, Rosendorff and Vreeland 2015b; Little 2015; Lorentzen 2015; Rundlett and Svolik 2015; Shadmehr and Bernhardt 2015). Different from collective-action models based on (*partially*) *common-value global games*, our model is built upon the technics in Palfrey and Rosenthal (1985), and assumes private and independent values among citizens.<sup>5</sup> The slightly variant version of the public good provision game gives us a feature of uniqueness in the collective-action stage and provides an alternative, tractable framework to analyze the choice of policy platform in the context of information manipulation and collective action.

The arrangement of this paper is as follows. The next section introduces the basic framework and characterizes the equilibria. The penultimate section studies the optimal mechanism design when the government has the commitment power. The final section concludes.

<sup>4</sup> The reform option in our framework is an opportunity cost for lying when the status quo is unpopular and therefore plays a similar role as the exogenous cost of censorship in Besley and Prat (2006) and Shadmehr and Bernhardt (2015) to sustain an informative equilibrium.

<sup>5</sup> Morris and Shin (2006) may call our game a *private-value interaction/global game*.

## A BENCHMARK MODEL

In this section, we introduce the setup of the benchmark model and characterize its equilibria.<sup>6</sup>

*Setup*

*Players and policy preferences.* There are three players: a government, Citizen 1 and Citizen 2. The government can choose a policy  $x \in \{Q, R\}$ . We call  $Q$  the status quo policy and  $R$  the reform policy. It costs the government  $\mu > 0$  to implement the reform policy relative to the status quo (hereafter  $\mu$  is called the policy-adjustment cost), as if the government draws a positive payoff gain  $\mu$  from  $Q$  relative to  $R$ . For notational convenience, we assign values for  $Q$  and  $R$  such that  $Q = 0$  and  $R = 1$ .

Citizen  $i$  only knows her own type  $t_i \in \{\underline{t}, \bar{t}\}$ , with  $\underline{t}$  and  $\bar{t}$  representing a *non-activist*, who is indifferent between the two policies, and an *activist*, who strictly prefers the reform policy to the status quo policy, respectively. We normalize a citizen's policy gain from the status quo policy to 0, no matter what type she is, i.e.:

$$u_i(Q; t_i) = 0, \quad i = 1, 2.$$

A non-activist is indifferent between the reform policy and the status quo policy<sup>7</sup>:

$$u_i(R; t_i = \underline{t}) = \underline{U} = 0, \quad i = 1, 2.$$

An activist gets a strictly positive payoff from the reform policy:

$$u_i(R; t_i = \bar{t}) = \bar{U} > 0, \quad i = 1, 2.$$

Both  $\underline{U}$  and  $\bar{U}$  are common knowledge.

*Information structure.* The government receives a private signal  $\theta \in \{L, H\}$ , which represents the societal demand for the reform policy (or the social fundamental). The government infers the citizens' preferences based on this signal. Specifically, we assume that the information structure satisfies the following form:

$$(t_1, t_2) |_{\theta=H} = (\bar{t}, \bar{t}) \text{ and } (t_1, t_2) |_{\theta=L} \neq (\bar{t}, \bar{t}).$$

Citizen  $i$  knows her own type but does not know the other citizen's type  $t_{-i}$ . However, both citizens share a common prior that the other citizen is an activist with probability  $p \in (0, 1)$ :

$$\Pr(t_i = \bar{t}) = p, \quad i = 1, 2.$$

The two citizens' types are independent and this fact is also common knowledge.

*Collective action and payoffs.* Successful collective action requires both citizens' participation. When it succeeds, the government is forced to implement the reform policy and suffers

<sup>6</sup> In the Appendix, we provide the proofs for most of the results based on a more general model of collective action, of which the benchmark model is a special case. A key difference is that in the benchmark model, citizens' collective-action gain comes from policy outcomes only, whereas in the generalized model, citizens' collective-action gain takes a flexible form. Such flexibility allows richer interpretation of citizens' motives of participating in collective action, including psychological factors such as grievances (e.g., Passarelli and Tabellini 2013).

<sup>7</sup> When  $\underline{U} < 0$ , the main results of this paper remain qualitatively the same.

$\rho_2 > 0$ . If only one of the citizens participates, with probability  $\lambda > 0$ , the individual protest is successful and the reform policy is implemented; with probability  $(1 - \lambda)$ , it is not successful and the original policy  $x$  remains unchanged. Thus, on average the government suffers a cost  $\rho_1 > 0$ , which may depend on  $\lambda$  and  $\mu$ .<sup>8</sup> If neither citizen participates, no policy change happens and the government suffers no cost.

Citizen  $i$ 's collective-action payoff is represented by

	participate ( $j$ )	abstain ( $j$ )
participate ( $i$ )	$u_i(R; t_i) - k_i$	$\lambda u_i(R; t_i) + (1 - \lambda)u_i(x; t_i) - k_i$
abstain ( $i$ )	$\lambda u_i(R; t_i) + (1 - \lambda)u_i(x; t_i)$	$u_i(x; t_i)$

where  $k_i$  is citizen  $i$ 's cost of participating in collective action;  $k_i$  the  $i$ 's private information that is not known by the government and the other citizen.  $k_i$  is assumed to be independent and identically distributed between 0 and 1 with cumulative distribution function  $F(\cdot)$ .

The government's total payoff consists of two parts: a policy implementation cost and a cost from collective action. We summarize its total payoff as follows (recall  $Q = 0$  and  $R = 1$ ):

	participate ( $j$ )	abstain ( $j$ )
participate ( $i$ )	$-x\mu - (1 - x)\rho_2$	$-x\mu - (1 - x)\rho_1$
abstain ( $i$ )	$-x\mu - (1 - x)\rho_1$	$-x\mu$

Throughout this paper, we make the following parametric assumptions:

1. the cost that the government suffers when only one citizen protests  $\rho_1$  is smaller than the cost of the reform policy  $\mu$ , i.e.,  $\rho_1 < \mu$ ;
2. the average cost of the government is increasing in the number of participants, i.e.,  $\rho_2 > \rho_1 > 0$ ;
3.  $F(\cdot)$  is weakly concave; and  $f(k) = F'(k) > 0, \forall k \in (0, 1)$ <sup>9</sup>;
4. an activist's gain from the reform policy is smaller than the upper bound of the collective-action cost, i.e.,  $\overline{U} \leq 1$ ; and
5. the probability that an individual challenge succeeds  $\lambda$  is sufficiently small but always positive, i.e.,  $\lambda \in (0, \frac{1}{2})$ .<sup>10</sup>

Denote  $A \equiv (1 - \lambda)\overline{U}$ , which is the payoff gain of joining a protest (excluding the protest cost) provided that the other citizen also participates. Similarly,  $B \equiv \lambda\overline{U}$  is the payoff gain when the other citizen does not participate. Hence, we have  $0 < B < A < 1$ .

Suppose  $S = \{L, H\}$  is the set of messages that the government can announce. The government can choose a propaganda strategy  $f(s|\theta)$ , where  $f(s|\theta)$  is the probability density

<sup>8</sup>  $\rho_1$  could be a function of  $\lambda$ , e.g.,  $\rho_1 = \lambda\mu + \delta$ , where  $\delta$  can be interpreted as an additional cost of dealing with a protester. We do not make any specific assumptions about how  $\rho_2$  and  $\rho_1$  depend on  $\lambda$  and  $\mu$ .

<sup>9</sup> It can be verified that the uniform distribution, or any distribution with a cumulative distribution function  $F(k) = k^\delta$  ( $0 < \delta < 1$ ), satisfies this property. The concavity of the distribution is used merely to guarantee a unique prediction in the collective-action stage. Without this assumption, we may need to deal with the problem of multiple equilibria, although the properties in the equilibrium we focus on are still valid.

<sup>10</sup> The condition that  $\lambda > 0$  is crucial for a unique prediction in the collective-action stage. When  $\lambda = 0$ , there always exists an equilibrium in that citizens never protest no matter what the government does.

function of the signal  $s$  conditional on the fundamental  $\theta$ . For simplicity, we consider the following particular form of information manipulation:<sup>11</sup>

$$\Pr(s = L | \theta = L) = 1, \Pr(s = H | \theta = H) = \varepsilon, \Pr(s = L | \theta = H) = 1 - \varepsilon.$$

Thus the government's choice of  $f(s|\theta)$  is simplified to the choice of  $\varepsilon \in [0, 1]$ . Note that  $(1 - \varepsilon)$  represents the extent to which the government lies about the (un)popularity of the status quo policy.

*Timing and actions.* The timing of actions is as follows.

Period (1) *Information manipulation and policy choice.* The government chooses a strategy of information manipulation  $\varepsilon \in [0, 1]$  and policy adjustment  $x \in \{Q, R\}$ .

Period (2) *Collective action.* Both citizens observe the government message  $s \in \{H, L\}$  and policy  $x \in \{Q, R\}$  and simultaneously decide whether to participate in a popular protest ( $a_i = 1$ ) or not ( $a_i = 0$ ).

The equilibrium notion is *Perfect Bayesian Equilibrium*. The government's equilibrium strategies consist of manipulating released information  $\varepsilon^* \in [0, 1]$ , and choosing a probability of reform  $\sigma^*(\theta, s): \{L, H\}^2 \rightarrow [0, 1]$  as a function of the fundamental  $\theta$  and the released message  $s$ .<sup>12</sup> Citizen  $i$ 's equilibrium strategy of whether to protest  $a_i^*(x, s, k_i, t_i): \{Q, R\} \times \{H, L\} \times [0, 1] \times \{t, \bar{t}\} \rightarrow \{0, 1\}$  is a function of the realized policy  $x$ , the realized signal  $s$ , the private cost of protesting  $k_i$  and the reference type  $t_i$ .  $a_i^*(x, s, k_i, t_i)$  is derived based on the citizen's equilibrium belief about the other citizen's reference  $t_{-i}$ , which is formed by Bayes' rule and the citizen's rational expectation of the government's equilibrium strategies  $\varepsilon^*$  and  $\sigma^*(\theta, s)$ .

Whenever the societal demand for the reform policy is low, i.e.,  $\theta = L$ , the government will never choose the reform policy. This is because without reform, the highest possible cost it suffers is  $\rho_1$ , which is smaller than the cost of reform  $\mu$ . Therefore,  $\sigma^*(\theta = L, \cdot) = 0$ .

Upon a high societal demand for reform, i.e.,  $\theta = H$ , we denote the government's probability of reform as  $\sigma_H = \sigma(\theta = H, s = H)$  or  $\sigma_L = \sigma(\theta = H, s = L)$  when "H" or "L" is reported.

When an activist observes  $s = H$ , she knows for sure that the other citizen is also an activist, so her equilibrium belief that the other citizen is an activist is 1, i.e.:

$$\gamma_H^* = \Pr(t_j = \bar{t} | s = H, x = Q) = 1. \quad (1)$$

When an activist observes  $s = L$  and  $x = Q$ , her equilibrium belief that the other citizen is an activist is

$$\gamma_L^* \equiv q(z^*) = \Pr(t_j = \bar{t} | s = L, x = Q) = \frac{p(1 - z^*)}{p(1 - z^*) + (1 - p)}, \quad (2)$$

<sup>11</sup> Models that study information manipulation have to make (parametric) assumptions about the feasible information structure  $\{f(s|\theta), S\}$  (as well as the associated cost) of one kind or another. For example, Besley and Prat (2006) and Shadmehr and Bernhardt (2015) assume that  $\{f(s|\theta), S\}$  can be either  $\Pr(s = \theta|\theta) = 1$  for all  $\theta$ , or  $\Pr(s = 0|\theta) = 1$  for all  $\theta$ . Our model allows an infinite number of choices, of which the case of binary choice ( $\varepsilon = 1$  and  $\varepsilon = 0$ ) is a special case.

<sup>12</sup> Slightly different from Austen-Smith and Banks (2000) and Kartik (2007), we allow information manipulation and policy choice to be *conditionally correlated*. Specifically, the government's policy choice is not only contingent on its private signal  $\theta$  indicating the social fundamental  $(t_1, t_2)$ , but also depends on its released message: "H" and "L." In addition, we allow the government to choose a mixed strategy over the two policy options  $Q$  and  $R$ .



where

$$z^* \equiv 1 - (1 - \sigma_L^*)(1 - \varepsilon^*). \quad (3)$$

It can be verified that  $q(z^*)$  is strictly decreasing in  $z^*$ .  $q(0) = p$ ,  $q(1) = 0$ ,  $q'(z) = -\frac{p(1-p)}{(1-pz)^2} < 0$ .

### Equilibrium Characterization

The following lemma shows that in any equilibrium, the citizens' strategy in the collective-action stage is uniquely determined.

LEMMA 1: (*Characterizing the collective-action stage*) In equilibrium, a non-activist never protests, and an activist never protests upon observing the reform policy. Suppose no reform is implemented (i.e.,  $x = Q$  is the realized policy), and let  $\gamma$  be an activist's (endogenous) belief that the other citizen is also an activist. For a given  $\gamma$ , an activist protests if and only if her cost  $k_i$  is weakly lower than a threshold  $k^*(\gamma)$ .  $k^*(\gamma)$  is unique and well defined by  $k^*(\gamma) = \gamma(A-B)F(k^*(\gamma)) + B$ , where  $F(\cdot)$  is the distribution function of the collective action cost  $k_i$ .  $k^*(\gamma)$  is strictly increasing in  $\gamma$ . Define

$$p_0(\gamma) = F(k^*(\gamma)), \quad (4)$$

which is the probability with which an activist protests.  $p_0(\gamma)$  is also strictly increasing in  $\gamma$  (see Appendix for the proof).

As neither the government nor the other citizen observes a citizen's cost, citizens' equilibrium choices of whether to join a protest appear random both to the government and to each other.  $p_0(\gamma)$ , the probability of an activist participating in collective action, is increasing in  $\gamma$ , which is an activist's endogenous belief that the other citizen is also an activist. This is because an activist understands that with bigger  $\gamma$ , the likelihood that her fellow citizen also demands reform is higher, even though the two cannot communicate with each other directly. In the extreme case when  $\gamma = 0$ , the probability of an activist joining a protest reaches the minimum. As  $\gamma$  goes up, it increases until reaching the highest value  $p_0(1)$ .

We first claim that, if the government does not make any policy changes under any circumstances, the citizens will never believe that it truthfully discloses any information (whenever  $\theta = H$ ). Therefore, information manipulation of the government will not effectively change citizens' prior belief. To show this, we need to demonstrate that, if the government does not reform at all, the citizens will think that it has an incentive to fully distort the information, i.e.,  $\varepsilon^* = 0$ . Specifically, if an equilibrium with  $\varepsilon^* > 0$  exists, by announcing "H," both citizens know that the other citizen is also an activist, as suggested by Equation 1, thus the government gets  $-W(p_0(1))$ , where

$$W(x) = \rho_2 x^2 + 2\rho_1 x(1-x). \quad (5)$$

As long as  $\rho_2 > \rho_1$ , we can verify that  $W(x) = (2\rho_2 - \rho_1)x^2 + 2\rho_1 x$  is always strictly increasing when  $x \in [0, 1]$ .

When  $\theta = H$  and the government never adjusts its policy, by announcing "L" it receives  $-W(p_0(\gamma_L^*))$ , where  $\gamma_L^*$  is the probability that each activist believes that the other citizen is also an activist, which is determined by Equation 2.

Because  $\gamma_L^* < 1$ ,  $W(p_0(\gamma_L^*)) < W(p_0(1))$ . As a result, the citizens are fully aware that the government always gets a strictly higher payoff by reporting "L" than by reporting "H." Hence, we have the following lemma.

LEMMA 2: *If the government does not choose the reform policy under any circumstances, the citizens will never believe that it reveals the truth when  $\theta = H$ , hence, the government's information manipulation cannot change citizens' prior beliefs.*

Lemma 2 means that the government becomes more credible when making policy concessions, hence, its revealed message is more likely to change citizens' beliefs in the fundamental. In other words, information manipulation is only effective when the government is willing to make concessions on the policy front. Next we characterize the rest of the equilibrium.

When  $W(p_0(1)) \leq \mu$ , the government has no incentive to adjust the policy. According to Lemma 2, it will never disclose any true information, i.e.,  $\varepsilon^* = 0$ .

When  $W(p_0(p)) \leq \mu < W(p_0(1))$ , we can show that the government has no incentive to disclose true information either. Suppose  $\varepsilon^* > 0$  and the government faces two activists. After receiving the message “H,” citizens know that both of them are activists, thus will join the protest with probability  $p_0(1)$  if the government keeps the status quo (Lemma 1). As a result, the government will be forced to reform as  $\mu < W(p_0(1))$  and receive a payoff  $-\mu$ . However, if the government reports “L” and does not reform, it will get  $-W(p_0(\gamma_L^*))$ , where  $\gamma_L^*$  is the probability with which an activist believes that the other citizen is also an activist, which is determined by Equation 2. Notice that  $\gamma_L^*$  is smaller than  $p$ . Because  $-W(p_0(\gamma_L^*)) > -W(p_0(p)) \geq -\mu$ , truthfully reporting “H” is a strictly dominated option. Therefore, we always have  $\varepsilon^* = 0$  when  $W(p_0(p)) \leq \mu < W(p_0(1))$ . Suppose  $\sigma_L^* > 0$ . By keeping the status quo, the government gets  $-W(p_0(\frac{p(1-\sigma_L^*)}{p(1-\sigma_L^*)+(1-p)}))$ , which is strictly higher than  $-\mu$ . Therefore, we know that the government never reforms, i.e.,  $\sigma_L^* = 0$ .

Without loss of generality, we now focus on the situation in which  $W(p_0(p)) > \mu$ . Suppose that the government faces two activists. In any equilibrium with  $\varepsilon^* > 0$ , the government sends “H” with a positive probability. After receiving “H,” citizens know that both of them are activists, thus joining a protest with probability  $p_0(1)$  if the government keeps the status quo (Lemma 1). As a result, it is always optimal for the government to reform when it announces: “H,” i.e.,  $\sigma_H^* = 1$  as  $\mu < W(p_0(p)) < W(p_0(1))$ .

By Lemma 1, an activist protests under the status quo with probability  $p_0(\gamma_L^*)$ . When both citizens are activists, if the government announces “H,” it is forced to reform and gets  $-\mu$ . However, if the government announces “L,” it gets a payoff of  $\max_{\sigma_L \in [0,1]} \{ -[(1-\sigma_L)W(p_0(\gamma_L^*)) + \sigma_L\mu] \}$ .<sup>13</sup> The following proposition characterizes all possible equilibrium.

PROPOSITION 1: (Characterizing the equilibria) *Provided that  $\lambda$  is sufficiently small:*

(1) *when  $\mu < W(p_0(p))$ , the unique class of equilibria  $(\varepsilon^*, \sigma^*)$  is determined by*

$$\mu = W(p_0(q(1-(1-\varepsilon^*)(1-\sigma_L^*)))),$$

*and  $\sigma_H^* = 1$ ; in this case, when both citizens strictly prefer the reform policy to the status quo, the degree of information manipulation without reform  $\Pr(s = L, x = Q | \theta = H) = 1 - q^{-1}[p_0^{-1}(W^{-1}(\mu))] < 1$  and it is strictly increasing in  $\mu$ ;*

(2) *when  $\mu \geq W(p_0(p))$ , there exists a unique equilibrium in which  $\varepsilon^* = 0$  and the government never chooses the reform policy; in this case, when both citizens strictly prefer the reform policy to the status quo, the degree of information manipulation without reform  $\Pr(s = L, x = Q | \theta = H) = 1$  (see Appendix for the proof).*

<sup>13</sup>  $(1-\sigma_L)W(p_0(\gamma_L^*)) + \sigma_L\mu$  is a linear function in  $\sigma_L$ . When the government calculates the equilibrium  $\sigma_L^*$ , it takes the citizens' equilibrium belief  $\gamma_L^*$  and behavior  $p_0(\gamma_L^*)$  as given.



Proposition 1 characterizes two different types of equilibrium, resulting from different parameters. If, e.g., the policy-adjustment cost is very high, i.e.,  $\mu \geq W(p_0(p))$ , the government never reforms and always lies to its citizen about the (un)popularity of the status quo policy. Being aware of this, the citizens never believe the messages sent by the government. Thus, we have a pooling equilibrium, in which the government is unable to persuade an activist not to participate in a protest when the other citizen is actually a non-activist.

However, if the policy-adjustment cost is low enough, i.e.,  $\mu < W(p_0(p))$ , the government exhibits a certain level of responsiveness to the citizens' demand and occasionally reveals the truths of the (un)popularity of the status quo policy to its citizens. In this more interesting case, when the reform policy is unpopular (i.e., not supported by both citizens), the government announces that it is unpopular and does not reform; when the reform policy is popular, it sometimes admits its popularity and enact the reform policy and sometimes lies and keeps the status quo policy. The government's announcement is more credible to its citizens in this equilibrium than in the previous one. Hence, now it is possible for the government to change the citizens' belief and reduce the risks of collective action when it maintains the status quo.

In the case when  $\mu$  is prohibitively high, the game is equivalent to the case when the government does not have an option to adjust policies at all. Thus, an important driving force of the different properties of the two equilibrium is the option of making policy changes. As a general result, Proposition 1 also shows that the probability with which the government chooses to lie  $\Pr(s = L, x = Q|\theta = H)$  decreases as  $\mu$  decreases. In other words, the more freedom the government has to adjust policies, the lower its cost of allowing the citizens to express discontent publicly, and the higher the opportunity cost it faces when censoring the bad news.

Based on Proposition 1, we can fully characterize the equilibrium probability of enacting a reform when there are two activists:

$$z^* = \begin{cases} 1 - \frac{1-p}{p} \frac{p_0^{-1}(W^{-1}(\mu))}{1-p_0^{-1}(W^{-1}(\mu))} & \text{if } \mu < W(p_0(p)) \\ 0 & \text{if } \mu \geq W(p_0(p)) \end{cases}.$$

Because  $W(\cdot)$  and  $p_0(\cdot)$  are strictly increasing, we know that  $z^*$  is increasing in  $p$  and decreasing in  $\mu$ . Thus, we have:

**COROLLARY 1:** *Provided that  $\lambda$  is sufficiently small, the probability of enacting a reform is increasing in the probability that citizens strictly prefer the reform policy  $p$  and decreasing in the policy-adjustment cost  $\mu$ .*

Note that  $p$  measures the credibility of the citizens' collective-action threat against the government. Corollary 1 is a straightforward result. It suggests that, in this highly simplified framework, the source of policy responsiveness of an authoritarian government ultimately comes from the threat of citizens' collective action. If the government knows that the citizens are unable to solve the collective action problem, it is not willing to make any effort to reform. In addition, policy responsiveness in authoritarianism also depends on the government's ability to adjust policies.

## INSTITUTIONAL DESIGN

In this section, we assume that the government is able to commit to a mechanism of information manipulation and reform (i.e., a set of rules that specify strategies *ex ante*) and study the optimal design of such a mechanism. The commitment may originate from a set of institutional

arrangements set by the authoritarian ruler. For simplicity, we also assume that  $W(p_0(1)) > \mu$  such that whenever “H” is announced, the government always finds it optimal to reform. For any mechanism  $(\varepsilon, \sigma_L)$  with  $\varepsilon > 0$ , it induces the same outcome (including the citizens’ induced beliefs) as the mechanism  $(\hat{\varepsilon}, \hat{\sigma})$  with  $\hat{\varepsilon} = 0$  and  $\hat{\sigma} = 1 - (1 - \sigma_L)(1 - \varepsilon)$ . Therefore, without loss of generality, we can focus on the mechanism in which the government never reports popular anger, i.e.,  $\varepsilon = 0$ .

The probability to adjust the policy is denoted as  $\sigma$ , which we call a *pure responsiveness mechanism*. We first write down the government’s expected total payoff:

$$E(\tilde{G}) = -p^2[(1 - \sigma)W(p_0(q(\sigma))) + \sigma\mu] - 2p(1 - p)p_0(q(\sigma))\rho_1. \quad (6)$$

When  $\sigma = 1$ , each citizen perfectly observes the other citizen’s type. When  $\sigma = 0$ , horizontal communication between citizens is shut down and no policy change occurs.

**PROPOSITION 2:** *(The optimal mechanism) Provided that  $W(p_0(1)) > \mu$  and the probability of a successful individual protest  $\lambda$  is sufficiently small:*

- (1) *in the best mechanism, even though both citizens strictly prefer the reform policy to the status quo, the government never fully responds to their demand, namely  $\sigma^* < 1$ ;*
- (2) *when  $\mu < W(p_0(p))$ , we have  $\sigma^* \in (0, 1)$  and this mechanism requires a commitment: because the policy-adjustment cost is strictly higher than the cost of facing collective action, i.e.,  $W(p_0(q(\sigma^*))) < \mu$ , the government has an ex post incentive to deviate (see Appendix for the proof).*

Proposition 2 suggests that in the best mechanism, the government uses a mixture of information manipulation and reform to reshape the citizens’ belief in its favored direction. When an activist sees that the government keeps the status quo policy, she is likely to believe that the other citizen is not an activist, hence, her incentive of joining a protest is compromised. Under such a mechanism, although the government has to make more efforts to implement the reform policy than in the case without commitment, it gains because of having the opportunity of discouraging the citizens from joining a protest when they are of different types.

## CONCLUSIONS

We develop a model in which an authoritarian government interactively uses information manipulation and reform to prevent citizens’ collective action against the regime. The two citizens’ incentives to join a protest depend on their beliefs about a societal fundamental and the government’s policy. We show that the government’s exerting efforts to reform serves two purposes: (1) it pleases the citizens directly and (2) it reveals the government’s private information on the (un)popularity of the status quo policy. Information manipulation, on the other hand, may reshape the citizens’ beliefs of the social fundamental and discourage them from joining collective action, granted that reform sometimes actually takes place.

We characterize the equilibrium level of information manipulation and policy changes. We show that, in equilibrium, the degree of information manipulation depends on the government’s ability to make policy concessions. This is because a high ability to adjust policies reduces the cost of revealing the truth. Expecting the information to be more credible, the citizens are more likely to abstain from participating in collective action when they are told that the reform policy is unpopular. In other words, occasional policy concessions lends credibility to government-provided information, which in turn reduces the risks of collective action.

We also consider the optimal institutional design with which the government can maximize its expected payoff by committing to a set of rules that *ex ante* specify its strategies. We show that, when commitment is possible (e.g., by institutionalizing certain government functions), the government always uses a mixture of information manipulation and reform.

Our model sheds lights on the puzzling fact that authoritarian regimes, such as China, control the information their citizens receive through propaganda and censorship, whereas at the same time make policy concessions to them.

## REFERENCES

- Austen-Smith, David, and Jeffrey S. Banks. 2000. 'Cheap Talk and Burned Money'. *Journal of Economic Theory* 91(1):1–16.
- Besley, Timothy, and Andrea Prat. 2006. 'Handcuffs for the Grabbing Hand? Media Capture and Government Accountability'. *The American Economic Review* 96(3):720–736.
- Bueno de Mesquita, Bruce, and George W. Downs. 2005. 'Development and Democracy'. *Foreign Affairs* 84:77–86.
- Bueno de Mesquita, Ethan. 2010. 'Regime Change and Revolutionary Entrepreneurs'. *The American Political Science Review* 104(3):446–66.
- Cai, Yongshun. 2004. 'Managed Participation in China'. *Political Science Quarterly* 119(3):425–51.
- Casper, Brett Allen, and Scott A. Tyson. 2014. 'Popular Protest and Elite Coordination in a Coup d'état'. *The Journal of Politics* 76(2):548–64.
- Chen, Xi. 2012. *Social Protest and Contentious Authoritarianism in China*. Cambridge: Cambridge University Press.
- Cox, Gary. 2009. 'Authoritarian Elections and Leadership Succession, 1975–2000'. Mimeo, Stanford University, Stanford, CA, USA.
- Crabtree, Charles, Christopher Fariss, and Holger Kern. 2015. 'Truth Replaced by Silence: Private Censorship in Russia'. Presented at NYU CESS 8th Annual Experimental Political Science Conference, 6 March 2015, NYU, New York, NY, USA.
- Dimitrov, Martin. 2013. 'Vertical Accountability in Communist Regimes: The Role of Citizen Complaints in Bulgaria and China'. In Martin Dimitrov (ed.), *Why Communism Did Not Collapse: Understanding Authoritarian Regime Resilience in Asia and Europe*. 276–302. Cambridge, UK: Cambridge University Press.
- . 2014. 'Dictatorship and Information: Autocratic Resilience in Communist Europe and China'. Book Manuscript, Tulane University, New Orleans, LA.
- Edmond, Chris. 2013. 'Information Manipulation, Coordination, and Regime Change'. *The Review of Economic Studies* 80(4):1422–58.
- Egorov, Georgy, and Konstantin Sonin. 2012. 'Incumbency Advantage in Non-Democracies'. Mimeo, Northwestern University, Evanston, IL.
- Egorov, Georgy, Sergei Guriev, and Konstantin Sonin. 2009. 'Why Resource-Poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data'. *The American Political Science Review* 103(4):645–69.
- Gehlbach, Scott, and Konstantin Sonin. 2014. 'Government Control of the Media'. *Journal of Public Economics* 118:163–71.
- Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland. 2011. 'Democracy and Transparency'. *The Journal of Politics* 73(4):1191–1205.
- . 2015a. 'Transparency, Protest and Autocratic Instability'. *American Political Science Review*: (forthcoming).
- . 2015b. 'Why Do Autocrats Disclose?' Prepared for the Annual Meeting of the Midwest Political Science Association, 16–19 April 2015, Chicago, IL.
- Huang, Haifeng. 2014. 'Propaganda as Signaling'. *Comparative Politics* (forthcoming).
- . 2015. 'A War of (Mis)Information: The Political Effects of Rumors and Rumor Rebuttals in an Authoritarian Country'. *British Journal of Political Science* (forthcoming).

- Kalathil, Shanthi, and Taylor C. Boas. 2003. *Open Networks, Closed Regimes: The Impact of the Internet on Authoritarian Rule*. Washington, DC: The Brookings Institution Press.
- Kartik, Navin. 2007. 'A Note on Cheap Talk and Burned Money'. *Journal of Economic Theory* 136(1): 749–58.
- King, Gary, Jennifer Pan, and Margaret Roberts. 2013. 'How Censorship in China Allows Government Criticism But Silences Collective Expression'. *The American Political Science Review* 107(2):1–18.
- . 2014. 'Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation'. *Science* 345(6199):1251722.
- Lee, Ching Kwan, and Yonghong Zhang. 2013. 'The Power of Instability: Unraveling the Microfoundations of Bargained Authoritarianism in China'. *American Journal of Sociology* 118(6):1475–1508.
- Little, Andrew. 2014a. 'Communication Technology and Protest'. Prepared for the Annual Meeting of the Midwest Political Science Association, 3–6 April 2014, Chicago, IL.
- Little, Andrew T. 2012. 'Elections, Fraud, and Election Monitoring in the Shadow of Revolution'. *Quarterly Journal of Political Science* 7(3):249–83.
- . 2014b. 'Are Non-Competitive Elections Good for Citizens?' Mimeo, Cornell, Ithaca, NY.
- . 2015. 'Fraud and Monitoring in Non-Competitive Elections'. *Political Science Research and Methods* 3(1):21–41.
- Lorentzen, Peter. 2013. 'Regularizing Rioting: Permitting Public Protest in An Authoritarian Regime'. *Quarterly Journal of Political Science* 8(2):127–58.
- . 2014. 'China's Strategic Censorship'. *American Journal of Political Science* 58(2):402–14.
- . 2015. 'China's Controlled Burn: Information Management and State-Society Relations Under Authoritarianism'. Book Manuscript, University of California Berkeley, Berkeley, CA.
- Lorentzen, Peter, Pierre Landry, and John Yasuda. 2014. 'Undermining Authoritarian Innovation: The Power of China's Industrial Giants'. *The Journal of Politics* 76(1):182–94.
- Miller, Michael. 2015. 'Elections, Information, and Policy Responsiveness in Autocratic Regime'. *Comparative Political Studies* 48(6):691–727.
- Morris, Stephen, and Hyun Song Shin. 2006. 'Heterogeneity and Uniqueness in Interaction Games'. In Lawrence E. Blume and Steven N. Durlauf (eds), *The Economy as an Evolving Complex System III: Current Perspectives and Future Directions*, 207–42. New York, NY: Oxford University Press.
- Nathan, Andrew. 2003. 'Authoritarian Resilience'. *Journal of Democracy* 14(1):6–17.
- Oi, Jean. 2003. 'Bending Without Breaking, the Adaptability of Chinese Political Institutions'. In Nicholas Hope, Dennis T. Yang and Mu Y. Li (eds), *How Far Across the River? Chinese Policy Reform at the Millennium*, 450–68. Stanford University Press: Stanford, CA.
- Palfrey, Thomas, and Howard Rosenthal. 1985. 'Voter Participation and Strategic Uncertainty'. *The American Political Science Review* 79(1):62–78.
- Pan, Jennifer. 2013. 'Measuring the Goals and Incentives of Local Chinese Officials'. Mimeo, Harvard University, Cambridge, MA.
- Passarelli, Francesco, and Guido Tabellini. 2013. 'Emotions and Political Unrest'. Working Paper, IGIER, Milano, Italy.
- Rundlett, Ashlea, and Milan Svolik. 2015. 'Deliver the Vote! Micromotives and Macrobehavior in Electoral Fraud'. Mimeo, UIUC, Champaign, IL.
- Shadmehr, Mehdi. 2014a. Mobilization, Repression, and Revolution: Grievances and Opportunities in Contentious Politics'. *The Journal of Politics* 76(3):621–35.
- . 2014b. 'Repression Structure and Extremism in Revolutionary Movements'. Mimeo, University of Miami, Miami, FL.
- Shadmehr, Mehdi, and Dan Bernhardt. 2011. 'Collective Action with Uncertain Payoffs: Coordination, Public Signals and Punishment Dilemmas'. *American Political Science Review* 105(4):829–51.
- Shadmehr, Mehdi, and Dan. Bernhardt. 2015. 'State Censorship'. *American Economic Journal: Microeconomics* (forthcoming).
- Shirk, Susan. 2010. *Changing Media, Changing China*. Oxford, UK: Oxford University Press.
- Smith, Alastair, and Scott Tyson. 2014. 'Two-Sided Coordination: Competition in Collective Action'. Mimeo, NYU, New York, NY.

- Stockmann, Daniela. 2013. *Media Commercialization and Authoritarian Rule in China*. New York, NY: Cambridge University Press.
- Stockmann, Daniela, and Mary E. Gallagher. 2011. 'Remote Control: How the Media Sustain Authoritarian Rule in China'. *Comparative Political Studies* 44(4):436–67.
- Svolik, Milan. 2012. *The Politics of Authoritarian Rule*. Cambridge: Cambridge University Press.
- Truex, Rory. 2015. 'Consultative Authoritarianism and Its Limits'. *Comparative Political Studies*.
- Wallace, Jeremy L. 2014. 'Juking the Stats? Authoritarian Information Problems in China'. *British Journal of Political Science* 1–19.
- . 2015. 'Information Politics in Dictatorships'. In Robert Scott and Stephen Kosslyn (eds), *Emerging Trends in the Social and Behavioral Sciences*. Hoboken, NJ: John Wiley and Sons.

# APPENDIX

In the Appendix, we prove Lemma 1 with a model that is more general and more abstract than the one in this paper. The latter can be seen as a special case of former. First, we present the model.

The two citizens' *policy preferences* in the policy-adjustment stage are  $u_i(x, t_i, t_j)$ ,  $i = 1, 2$ . Collective-action payoff is characterized by

	participate	abstain
participate	$V_i^{11}(x, t_i, t_j) - k_i$	$V_i^{10}(x, t_i) - k_i$
abstain	$V_i^{01}(x, t_i, t_j)$	$V_i^{00}(x, t_i)$

where  $k_i$  is citizen  $i$ 's cost of participating in collective action, which is private information.  $k_i$  is independently and identically distributed with a cumulative distribution function  $F(\cdot)$  and support  $[\underline{k}, \bar{k}]$ . It is only observed by citizen  $i$  after she observes government policy  $x$ . Moreover, we define

$$A_i(x, t_i, t_j) = V_i^{11}(x, t_i, t_j) - V_i^{01}(x, t_i, t_j), \quad (\text{A1})$$

$$B_i(x, t_i) = V_i^{10}(x, t_i) - V_i^{00}(x, t_i). \quad (\text{A2})$$

We can verify that assumptions in the benchmark model of the Model section are a special case of the following assumptions.

ASSUMPTION 1:  $F(\cdot)$  is weakly concave;  $f(k) = F'(k) > 0$ ,  $\forall k \in [\underline{k}, \bar{k}]$ .

ASSUMPTION 2:  $\exists i \in I = \{1, 2\}$ ,  $A_i(R, t_i, t_j) \leq \underline{k}$ ,  $B_i(R, t_i) \leq \underline{k}$ ,  $B_{-i}(R, t_{-i}) \leq \underline{k}$ ,  $\forall t_i, t_j, t_{-i} \in \{\underline{t}, \bar{t}\}$ ;  $\forall i \in I = \{1, 2\}$ ,  $A_i(Q, t_i = \underline{t}, t_j) \leq \underline{k}$ ,  $B_i(Q, t_i = \underline{t}) \leq \underline{k}$ ,  $\forall t_j \in \{\underline{t}, \bar{t}\}$ .

ASSUMPTION 3:  $A = A_i(R, t_i = \bar{t}, t_j = \bar{t})$ ,  $B = B_i(R, t_i = \bar{t})$  do not depend on  $i$ .  $\bar{k} \geq A > B > \underline{k}$ .<sup>14</sup>

## Proof of Lemma 1

(a) Conditions  $A_i(R, t_i, t_j) \leq \underline{k}$ ,  $B_i(R, t_i) \leq \underline{k}$  in Assumption 2 suggest that it is a dominant strategy for citizen  $i$  not to protest under the reform policy  $R$ . Expecting  $i$ 's behavior, citizen  $j$  also finds it profitable to abstain because  $B_{-i}(R, t_{-i}) \leq \underline{k}$ . As a result, under the reform policy, no one protests.

<sup>14</sup> A weaker version is:  $A > B$  and  $\bar{k} > B > \underline{k}$ . The condition that  $A > \bar{k}$  is merely to simplify the analysis and does not qualitatively change the results.

(b) Under the status quo policy, conditions in Assumption 2,  $A_i(Q, t_i = \underline{t}, t_j) \leq \underline{k}$ ,  $B_i(Q, t_i = \underline{t}) \leq \underline{k}$ , imply that the non-activist type  $\underline{t}$  will never protest. Thus, the only uncertainty is to what extent an activist will join the protest.

(b.1) An activist's payoff gain in protest is

$$\gamma \Pr(j \text{ protests} | t_i = \bar{t})A + (1 - \gamma \Pr(j \text{ protests} | t_i = \bar{t}))B - k_i.$$

Hence in equilibrium an activist will use a cut-point strategy. In other words, she will protest if and only if her cost of protest  $k_i$  is smaller than a threshold  $k^*$ .

(b.2) Now suppose  $i$ 's cut-point is  $k_i^*$ ,  $i = 1, 2$ . According to (b.1), the payoff gain of player  $i$  is

$$\gamma F(k_i^*)(A - B) + B - k_i.$$

It then can be verified that the equilibrium condition is equivalent to

$$k_1^* = \gamma F(k_2^*)(A - B) + B, \quad (\text{A3})$$

$$k_2^* = \gamma F(k_1^*)(A - B) + B. \quad (\text{A4})$$

(b.3) Without loss of generality, let us assume  $k_1^* \leq k_2^*$ , so we get

$$\gamma F(k_2^*)(A - B) + B \leq \gamma F(k_1^*)(A - B) + B,$$

therefore  $k_1^* \geq k_2^*$  so that we have  $k_1^* = k_2^* \in [B, \bar{k}]$ . Let us denote them as  $k^*$ , we then have

$$k^* = \gamma F(k^*)(A - B) + B. \quad (\text{A5})$$

Because of Assumption 1,  $\psi(x) \triangleq \gamma F(x)(A - B) + B - x$  is also weakly concave. In addition, we have  $\psi(\underline{k}) = B - \underline{k} > 0$ ,  $\psi(\bar{k}) = \gamma(A - B) - (\bar{k} - B) < 0$ .

Because of continuity of  $\psi(x)$ , by the *Intermediate Value Theorem*,  $\exists$  a solution  $k^* \in (\underline{k}, \bar{k})$  such that  $k^* = \gamma F(k^*)(A - B) + B$ .

Because of concavity of  $\psi(x)$ ,  $\forall k \in (\underline{k}, k^*)$ ,  $\psi(k) > 0$  and  $\forall k \in (k^*, \bar{k})$ ,  $\psi(k) < 0$ . As a result,  $k^*$  is the unique cut-point equilibrium. As  $k^*$  depends on  $\gamma$ , we also write it as  $k^*(\gamma)$ .

(b.4)  $k^*$  is uniquely and well defined by

$$k^* = \gamma F(k^*)(A - B) + B. \quad (\text{A6})$$

By Equation A6, we know that  $k^*(\gamma) \geq B > 0$ . Therefore, we can rewrite the equation as

$$\gamma = \frac{k^* - B}{(A - B)F(k^*)}.$$

Therefore the function  $k^*(\gamma)$  is invertible. In the following we will show that  $\gamma$  is strictly increasing in  $k^*$  when  $k^* \geq B$ . Because an inverse of a strictly increasing function is also a strictly increasing function, we will know that  $k^*(\gamma)$  is also strictly increasing.

(b.5) It is obvious that  $\frac{k^* - B}{(A - B)F(k^*)}$  is differentiable in  $k^*$ , thus we have

$$\frac{d \frac{k^* - B}{(A - B)F(k^*)}}{dk^*} = \frac{F(k^*) - (k^* - B)f(k^*)}{(A - B)(F(k^*))^2}.$$

We merely need to show  $F(k^*) - (k^* - B)f(k^*) > 0$ . Because  $F(k^*)$  is differentiable,  $\exists \xi \in [\underline{k}, k^*]$  s.t.  $F(k^*) = F(\underline{k}) + f(\xi)(k^* - \underline{k}) > f(\xi)(k^* - B) \geq (k^* - B)f(k^*)$ . The last inequality comes from concavity. ■

*Proof of Proposition 1*



(a) When  $\lambda$  is small,  $p_0(0) = F(B) = F(\lambda \bar{U})$  and  $W(p_0(0))$  are small, so that  $W(p_0(0)) < \mu$ . This condition guarantees that the government never reforms with probability 1 in the case of two activists.

(b) We first focus on the equilibrium with  $\varepsilon^* > 0$ .

(b.1) If  $W(p_0(q(z^*))) > \mu$ , we must have  $\sigma_L^* = 1$ . Thus,  $W(p_0(q(z^*))) = W(p_0(q(1))) = W(p_0(0)) < \mu$ , contradiction.

(b.2) If  $W(p_0(q(z^*))) < \mu$ , we must have  $\sigma_L^* = 0$ . In this case, if there are two activists, the government would always report “L” instead of “H” to get a strictly profitable deviation,  $-W(p_0(q(z^*))) > -\mu$ .

(b.3) As a result, when  $\varepsilon^* > 0$ , we must have  $W(p_0(q(z^*))) = \mu$ .  $W(p_0(q(z^*)))$  is strictly decreasing as  $z^*$  increases. The equation has a solution if and only if  $W(p_0(q(0))) > \mu$ , i.e.,  $W(p_0(p)) > \mu$ , provided  $W(p_0(q(1))) = W(p_0(0)) < \mu$ . We can also check that when  $W(p_0(p)) > \mu$ , all the other incentives of the government are satisfied. As a result, the equilibrium is determined by  $W(p_0(q(z^*))) = \mu$ , which induces a unique solution for  $z^*$ . The uniqueness comes from the fact that  $W(\cdot)$ ,  $p_0(\cdot)$  are strictly increasing and  $q(\cdot)$  is strictly decreasing.

(c) Let us now investigate the equilibrium with  $\varepsilon^* = 0$  and  $\sigma_L^* = 0$ . When  $\varepsilon^* = 0$ , in order to make  $\sigma_L^* = 0$  part of the equilibrium strategies, we must have  $\mu \geq W(p_0(p))$ . In this case, the government does not have an incentive to deviate from the strategy  $\sigma_L^* = 0$  when there are two activists and it reports the message “L.”

(d) Following the similar logic in (b), we can show that when  $W(p_0(p)) > \mu$ ,  $(\varepsilon^* = 0, \sigma_L^* > 0)$  is an equilibrium and is characterized by  $W(p_0(q(z^*))) = \mu$ . ■

### Proof of Proposition 2

As  $W(x) = (2\rho_2 - \rho_1)x^2 + 2\rho_1x$ , we have

$$W(p_0(0)) = \rho_2 p_0(0)^2 + 2\rho_1 p_0(0)(1 - p_0(0)).$$

(a) As long as  $\rho_2 > \rho_1$ ,  $W(x) = (2\rho_2 - \rho_1)x^2 + 2\rho_1x$  is always strictly increasing when  $x \in [0, 1]$ .

If  $\rho_2 \geq 2\rho_1$ ,  $W(x)$  is strictly increasing when  $x > 0$ ;

If  $\rho_2 < 2\rho_1$ ,  $W(x)$  is strictly increasing when  $x \in [0, \frac{\rho_1}{\rho_1 - (\rho_2 - \rho_1)}]$ .

(b) When  $\lambda$  is sufficiently small, we know that  $B = \lambda L$  and  $p_0(0) = F(B)$  are both sufficiently small. Therefore,  $p_0'(0) = f(p_0(0))p_0(0)(A - B)$  is also very small. Thus,  $\mu > W(p_0(0)) + 2p_0'(0)\rho_1$ , provided  $\lambda$  is sufficiently small. As a result, the assumptions imply that

$$\max\{\rho_1, W(p_0(0)) + 2p_0'(0)\rho_1\} < \mu. \quad (A7)$$

(c) From Equation 6 we get

$$\frac{dE(\tilde{G})}{d\sigma} = p^2[-(\mu - W(p_0(q))) - (1 - \sigma)\frac{dW}{dp_0}\frac{dp_0}{dq}\frac{dq}{d\sigma}] - 2p(1 - p)\frac{dp_0}{dq}\frac{dq}{d\sigma}\rho_1, \quad (A8)$$

where  $\frac{dW}{dp_0} > 0$ ,  $\frac{dp_0}{dq} \geq 0$ ,  $\frac{dq}{d\sigma} < 0$ . Because  $q(1) = 0$ ,  $q'(1) = \frac{p}{1-p}$ , we have

$$\frac{dE(\tilde{G})}{d\sigma} \Big|_{\sigma \rightarrow 1} = p^2[-(\mu - W(p_0(0))) + 2p_0'(0)\rho_1].$$

By inequality (A7),  $\frac{dE(\tilde{G})}{d\sigma} \Big|_{\sigma \rightarrow 1} < 0$ . Because  $E(\tilde{G})$  is continuous in  $\sigma$ ,  $\sigma^*$  always exists in  $[0, 1]$ . As  $\frac{dE(\tilde{G})}{d\sigma} \Big|_{\sigma \rightarrow 1} < 0$ , we must have  $\sigma^* \in [0, 1)$ .

(d) When  $\mu < W(p_0(p))$ ,  $\frac{d\tilde{E}(\tilde{G})}{d\sigma}\big|_{\sigma \rightarrow 0} \geq -p^2[\mu - W(p_0(p))] > 0$ , therefore  $\sigma^* \in (0, 1)$ . The first-order condition implies

$$\mu - W(p_0(q^*)) = -(1 - \sigma^*) \frac{dW}{dp_0} \frac{dp_0}{dq} \frac{dq}{d\sigma} - 2 \frac{1 - \gamma}{\gamma} \frac{dp_0}{dq} \frac{dq}{d\sigma} \rho_1,$$

therefore,  $\mu - W(p_0(q^*)) > 0$ . ■