Radicalism in Mass Movements: Asymmetric Information and Endogenous Leadership*

HENG CHEN

University of Hong Kong

WING SUEN

University of Hong Kong

August 26, 2020

Abstract. Asymmetric information and diverse preferences for reform create an agency problem between opposition leaders and citizens. Dissatisfied citizens are unsure of how bad the current situation is but infer this information from the scale of the leader's reform proposal. Because radical leaders have an incentive to exaggerate and mislead, to command credibility, they must paradoxically radicalize the proposal further as a way of signaling the necessity of change. Radicalism motivated by signaling is costly, as it reduces a movement's chances of success. This mechanism also contributes to leadership radicalization when the leaders of movements arise as a compromise among diverse interests.

Keywords. political agency problem; signaling; endogenous leadership; regime change

JEL Classification. D74; D82

*Heng Chen: hengchen@hku.hk, tel 852-28578506. Wing Suen: wsuen@econ.hku.hk, tel 852-28578505.

We thank Navin Kartik, Balazs Szentes, Xianwen Shi, and seminar participants at the Asian Meeting of the Econometric Society (2019), Political Selection Workshop (2019) and Allied Social Sciences Association annual meeting (2020) for their comments. This research project is partly supported by the General Research Fund of the Research Grants Council of Hong Kong (Project No. 17504718) and by the Seed Fund for Basic Research of Hong Kong University (Project No. 201811159002).

1. Introduction

Leaders of mass movements provide direction that gives shape to popular discontent. In addition to organizing the masses, their key task is to formulate an alternative policy proposal or reform agenda to rally support among dissatisfied citizens to replace the status quo. However, the interests of leaders and their followers are seldom the same. First, the preferences of different individuals are naturally different: some prefer radical solutions, while others prefer more moderate ones. In addition, the stakes are much higher for leaders, and the possible sanctions they face are also higher. Furthermore, leaders are well informed about the political situation because they are specialized political actors, while the masses are generally less informed because the extent of their involvement in the movement is much smaller. These systematic differences present an agency problem in the relationship between leaders and their followers. In this paper, we analyze how this agency problem may distort the reform agenda of opposition leaders.

The often-observed radical agendas that accompany mass movements may be a dramatic manifestation of such an agency problem. There is no shortage of examples in which radical leaders propose unrealistic demands or extremist agendas that sow the seeds of failure. The fight for a democratically elected chief executive in Hong Kong offers a good case in point. Around 2013, electoral reforms became a focus in the political arena in Hong Kong. While there were more moderate proposals that would attempt to squeeze the greatest degree of democracy within the strictures of the Basic Law, Hong Kong's mini-constitution, leaders of prodemocratic political parties advocated for more radical reforms that would completely sidestep the role of the "nomination committee" specified in the Basic Law. Meanwhile, a group of individuals outside established political parties started the Occupy Central movement in 2014 and proposed using civil disobedience to signal their resolve to achieve a full-fledged "genuine democracy." The majority of Hong Kong citizens endorse the values of modern democracy, but "many people consider Occupy Central too radical a movement to strive for true democracy," as acknowledged by one of the leaders, Benny Tai.[1] In the

[1] See "Central Issues of the Occupy Central Movement," 23 May, 2013, *South China Morning Post*.

end, the movement eventually ran out of steam without any political achievement: the unrealistic political demands could not maintain popular support, and the undemocratic electoral system remains largely intact.[2]

Radicalism is a common feature of many mass movements; similar examples from the Tiananmen democracy movement and the student movement in Tokyo also suggest that radical and unrealistic reform agendas indeed suppressed popular support.[3,4] This phenomenon seems puzzling, given that one cannot be fully convinced by simply arguing that leaders of mass social movements have ideologically extreme preferences. Even if they are indeed radical, they still face a trade-off between proposing a radical reform agenda that suits their personal ideology and proposing a moderate agenda that appeals to a broader spectrum of citizens. Why do radical leaders often refuse to settle for less radical but more realistic agendas to boost support? Why is it often the case that radicals rather than moderates occupy leadership positions to propose agendas at the far end of the spectrum?

In this paper, we develop a theory to analyze radicalism in mass movements. The theoretical backdrop is a simple regime change model. Citizens dislike the status quo policy because it does not accord with the current situation (the "state"). They agree that a change is needed but disagree over the alternative policy to be implemented: radicals prefer larger changes, and moderates like smaller ones. The opposition leader proposes a reform agenda, and citizens join in protests if they are sufficiently attractive. The chances of success increase in the mass of protesters. If the agenda proposed is very close to the status quo, it may not be sufficiently attractive to draw followers, given that protest actions are costly. However, if the agenda is very radical, it may discourage relatively moderate citizens from participat-

[2] See details of the Umbrella Movement in Hong Kong in Dapiran (2017).

[3] In May 1989, during the Tiananmen democracy movement, the leader of the radicals announced the movement's key demands, including an end to martial law, withdrawal of the army, amnesty for participants in the movement, and a complete end to press censorship (Ogden et al. 1992, p. 247). None of the demands seemed realistic in the context of the political situation in China, especially given that press censorship was one of the key instruments that the regime relied on to retain its grip on power. The movement's radical leadership could not maintain its appeal to the popular masses: "From May 28 to June 3, the student presence in Tiananmen Square subsided considerably" (Ogden et al. 1992, p. 238).

[4] Leaders in the 1967–1969 student movement in Tokyo advanced the goal of "debunk[ing] fake democracy" to justify their resort to violence, an escalation over the largely peaceful demonstrations in the early 1960s aimed at the "defense of parliamentary democracy" (Tsurumi 1970). Sunada (1969, p. 471) commented that the radicals had a vision "so idealistic, indeed Utopian ..., that it is impossible for the general public to understand."

ing. Not surprisingly, the leader strategically chooses an agenda by trading off the chances of success against her own policy preference.

The key element in our theory is that we introduce asymmetric information about the political situation or the "state": the opposition leader knows the state but citizens do not. A high state would warrant a large reform, while a low state would warrant an incremental reform.[5] Citizens can only make inferences about the state based on the scale of reform proposed by the opposition leader, and a larger fraction of citizens will be mobilized in the high state (in which conditions are favorable for the opposition's success). Therefore, the leader may have an incentive to exaggerate the level of the state by proposing a large reform even in the low state. Because citizens are aware of such a motive, the leader cannot merely choose a strategic agenda that optimally balances the trade-off between the chances of success against her own policy preference. Instead, she may have to resort to "irrational radicalism" by choosing a scale of reform that is even larger than what is warranted by the high state. Although she knows that radicalism would not be popular with citizens, she cannot soften her position; otherwise, citizens would interpret that she is choosing an agenda to shore up her support.

If the leader is known to have a moderate preference, she has no incentive to pretend that a large reform is necessary when it is not. Only leaders with radical preferences are plagued by this problem of asymmetric information. In other words, a radical leader has to radicalize the agenda even more to convince citizens that the current situation is indeed bad. This type of signaling is costly for the leader: popular support is suppressed, and the probability of success is reduced. If the leader's preference happens to be very radical, her agenda would be distorted so much that the winning probability in the high state is even lower than that in the low state.

[5] In connection to the motivating example, the "state" can be interpreted as the extent to which Beijing interferes or meddles in the internal affairs of Hong Kong or the degree of Hong Kong's autonomy. If Beijing is largely self-restrained and the interference is limited ("low state"), then a moderate change to the electoral system can be satisfactory. If, instead, the government of Hong Kong had loses a large degree of autonomy ("high state"), then a free election with "one person, one vote" is really necessary to select a chief executive accountable to local citizens and to shield the city from outside manipulation. Further, there is still a spectrum of political options, from moderate to very radical, should Hong Kong's autonomy be seriously compromised: accepting elections with a larger nomination committee, Western-style full-blown democracy, or self-determination with a referendum.

Given that only the radical leader suffers from citizens' suspicion, which compels her to radicalize his proposal to signal, we investigate further under what circumstances it is radicals (instead of moderates) who lead the movements. To this end, we introduce a mechanism of endogenous leadership determination in Section 4: citizens can pay an extra cost to join the leadership and enjoy additional benefits from the likely reform. A "representative leader" of the group proposes a reform agenda, which is a compromise among the diverse interests of leaders. This mechanism is intuitive, requiring a minimum set of reasonable assumptions. We show that an equilibrium with leadership determination mechanisms of this type can exist and under some circumstances, it is unique.

In Section 5.1, we show that in a setting with both signaling mechanisms and endogenous leadership, stronger suppression by the regime may radicalize the leadership and therefore its reform agenda. On the one hand, it is intuitive that the leadership will be more radical when the cost of taking up the leadership role is higher. On the other hand, it is novel that a higher participation cost of followers can also radicalize the leadership. Further, we illustrate in Section 5.2 that the interaction of these two mechanisms can shed light on an empirical puzzle: in societies with structural roots of political instability, political upheavals with mass support are not observed as often (Geddes 1990; Goldstone 2001). According to our model, this puzzle is less surprising than it seems because a society ripe for revolt is also a breeding ground for radical leaders, whose radical agendas likely undermine their movements' prospects of success.

To be sure, our model can account for certain types of radicalism but not all. Our theory is useful for understanding mass movements where there exists a large degree of uncertainty in political situations and the leaders lack other means to credibly transmit the information they possess. We model uncertainty over payoffs in this framework, but this modelling choice is not crucial. An alternative model with uncertainty about the technology of regime change would produce similar results.[6] Another contribution made in this paper is that we

[6] Opposition leaders are often political insiders; they may possess private information about when the regime is vulnerable and when it is not (i.e., uncertainty over the technology of regime change). Because citizens are more likely to participate if the regime is vulnerable, the leader has an incentive to mislead them into believing that this is the case. With suitable assumptions and modifications, our main results still hold under this specification of uncertainty.

analyze the agency problem in political leadership. In parallel to a standard principal-agent problem, we can think of citizens as "delegating" the task of collecting information and formulating reform proposals to leaders. However, the choice of taking up the leadership role and becoming "agents" is endogenous in our model.

2. Literature Review

In the literature that analyzes how political radicalism arises in protests, the following two studies are the most relevant. Shadmehr (2015) examines how the proposal offered by the leader influences the outcome of a protest game, which is closest to the issues that we investigate. In that paper, agents can choose their participation effort, and the leader proposes a revolutionary agenda to attract followers and to induce their effort. Therefore, the leader strategically chooses the degree of radicalism in his agenda by trading off between the extensive and intensive margins. Michaeli and Spiro (2018) analyze under what conditions radicals rather than moderates initiate dissent and the strength of the dissent that they choose to express. Each agent trades off the cost of being sanctioned and the cost of deviating from his preferred position; the rise of extremism largely depends on the cost structures involved in this trade-off. Our work differs from those papers in three aspects. First, we only allow for a binary choice among the followers (to participate or not), and there is no intensive margin. Second, leaders endogenously emerge in our model, and such a mechanism reinforces radicalism in equilibrium. Finally, given the presence of asymmetric information in our model, a radical agenda is used as a costly signal and may not be the optimal choice for the leader, which differs from the results in the aforementioned studies.

A few recent papers explicitly consider the informational role of leadership in regime-change games. The common feature of this literature is that leaders' action is informative about either payoffs (e.g., Shadmehr and Bernhardt 2019) or the aggregate state (e.g., Bueno de Mesquita 2010). In our signaling game, radicalism refers to an unrealistic agenda that lacks popular support, and it does not refer to the violent tactics that are sometimes deployed by vanguards.[7]

[7] Other studies focus on the motivating role of leadership in regime change games (e.g., Morris and Shad-

In an electoral competition setting, Acemoglu, Egorov, and Sonin (2013) show that honest politicians may resort to populist policies (i.e., those to the left of the median voter's preference) as a way to signal that they are not captured by right-wing special interests. Bils (2019) shows that an informed officeholder can use extreme but unnecessary policies to signal his expertise to voters, while the uninformed officeholder favors moderate and measured policies.[8] In our collective action setting, signaling is about a common factor (the "state") that affects the payoffs of all citizens, rather than about the personal qualities of the leader. Banks (1990) studies the impact of an agenda-setter's superior information on her proposal to voters in a monopoly agenda control model. The key concern of that model is how voters extract information from the proposed agenda in a semipooling equilibrium, whereas we study a separating equilibrium in which citizens infer the state perfectly from the reform agenda and the leader's known propensity to reform causes herself to behave radically.

In the organization context, Hermalin (1998) considers a signaling game in which the leader's effort serves as a signal of the return to efforts, which in turn motivates the followers to give more effort. In this vein, Fu, Li, and Qiao (2020) is the most recent work on leadership in organizations with signaling and reputation concerns. In our model, the agenda proposed by the leader is a signal as well. The "cost" of such a proposal is endogenously determined in a simple protest game.

3. Asymmetric Information and Radicalism

The key to our theory is that the underlying political situation is uncertain and that the opposition leader possesses more information than citizens do. In this section, we embed this mechanism into a simple regime-change model, where citizens make collective participation decisions in response to the reform agenda proposed by the opposition leader. We first specify both the information structure and the payoff structures of leaders and citizens in

mehr 2020) and the coordinating role of leadership (e.g., Dewan and Myatt 2007). Landa and Tyson (2017) argue that leaders with strong coercive power can improve the transmission of information to followers.

[8] Boleslavsky and Cotton (2015) show that voters may tolerate extremist policies more when they are better informed about the quality of candidates during political campaigns.

Section 3.1. Then we analyze the equilibrium of the pretest game in Section 3.2, where the reform agenda is taken as given. The analysis of agenda-making of the opposition leader is conducted in both the full information case and the asymmetric information case, and the corresponding results are contrasted in Sections 3.3 and 3.4. In this section, the leadership preference is fixed and known to citizens; however, we extend our model and allow the leadership to form endogenously in Section 4.

3.1. The setup

Consider a society populated by a unit mass of citizens, indexed by i. They are not satisfied with the status quo policy y_0. In general, citizens prefer a policy, denoted by y, that is aligned with the current situation in society, which we refer to as the "state" and denote by θ. However, citizens have heterogeneous preferences regarding the appropriate policy for society. The preference of citizen i is parameterized by x_i, which is uniformly distributed on $[0, 1]$. The payoff to the citizen with preference x_i when the policy is y and the state is θ is:

$$u(y, \theta, x_i) = \overline{u} - |x_i + \theta - y|,$$

where $\overline{u}$ is a constant. The most preferred policy of citizen i is $x_i + \theta$, which we refer to as her *ideal policy*. Because preferences are different, the interests of citizens are not perfectly aligned. We say that citizen i is *more radical* if her preference x_i is higher.[9]

We assume that the political situation or the state θ is uncertain and unknown to citizens. Specifically, there are two states, θ_H and θ_L. The prior probabilities of the two states are π_H and π_L, respectively. The status quo policy and the two states are such that $y_0 < \theta_L < \theta_H$. As a result, regardless of the state of the world, there exists a mismatch between the status quo policy and the state, which leads to dissatisfaction and drives demand for reform. The difference is that in the high state θ_H, the discrepancy between the current state and the status quo policy is larger and the discontent is stronger.

[9] The assumption that all citizens have ideal policies to the right of y_0 is made only for simplicity of exposition. A variant of the model can accommodate the case where some citizens' ideal policies are to the left of the status quo.

In addition to the group of citizens, there exists an opposition leader, with preference x^m, who formulates and proposes an alternative policy y_1 to incite a revolt, intending to replace the status quo y_0. Such a policy can be interpreted as the reform agenda, political demands, or a blueprint for the new society. We say that a reform agenda is *more radical* if y_1 is higher. The key assumption is that the leader observes the state θ, but she cannot produce objective verifiable evidence that credibly transmits her private information to citizens. The leader chooses a reform agenda $y_1 = y_1^L$ if the state is low and $y_1 = y_1^H$ if the state is high. Once agenda y_1 is announced, citizens observe it and make an inference about the state.

Our assumption about information asymmetry across the leader and citizens is reasonable. Each individual citizen has little influence on the outcome; hence, the incentive for citizens to obtain accurate information about the state in society is small.[10] The leader, on the other hand, has a much greater stake in reform. She has to organize the masses, formulate a policy, and convince citizens that her alternative is superior to the status quo. Because her choice of agenda has a material effect on the outcome of the mass movement, the leader tends to spend more time and effort learning about the environment. In this static model, we abstract from the information acquisition decision and assume that the leader is endowed with private information about the state. The assumption that the leader of an organization possesses private information unavailable to followers is common in the literature on leadership (e.g., Hermalin 1998).

Once citizens observe the opposition leader's reform agenda y_1 and make inferences about the state θ, they need to decide whether to participate in a mass movement against the existing regime ($a_i = 1$) or not ($a_i = 0$). We label a citizen who chooses $a_i = 1$ a *follower* of the movement, and a citizen who chooses $a_i = 0$ a *bystander*. All citizens agree that some reform is desirable, since every citizen's ideal policy is larger than y_0, but they disagree on which reform is best. Indeed, if a reform agenda is farther from a citizen's ideal policy than the status quo is, this citizen will prefer to maintain the status quo.

[10] For a more concrete example, regarding the extent to which the political system in Hong Kong was crippled by the influence of Beijing, citizens in Hong Kong had no more information than some anecdotes that were discussed in media. Investing in knowledge about the constitutional issues concerning the degree of Hong Kong's autonomy may also be too costly to undertake.

The success probability of the movement depends on the total mass of citizens who choose to follow the opposition. Let A represent the mass of citizens who choose $a_i = 1$, and let $G(A)$ be the probability of success. If the movement succeeds, the reform agenda y_1 is implemented; otherwise, the status quo y_0 prevails.[11]

The utility from regime change and the cost incurred differ across three types of citizens: bystanders, followers, and leaders. Under the assumption that each citizen is atomistic, her participation decision has no influence on the total size of the attack in this model. The payoff of a bystander of the movement, for a given size of attack A and a given reform agenda y_1, is:

$$U_b(x_i) = u(y_0, \theta, x_i) + G(A)\left(u(y_1, \theta, x_i) - u(y_0, \theta, x_i)\right).$$

The utility difference to citizen i under the alternative policy y_1 and under the status quo y_0 is the *reward from success* (i.e., $u(y_1, \theta, x_i) - u(y_0, \theta, x_i)$). For any $y_1 > y_0$, the reward from success increases in θ and in x_i. Bystanders may gain or lose in the new society, depending on whether the reward from success is positive or negative. If the reform agenda y_1 is far to the right of the ideal policy of bystanders, they may be worse off when the movement succeeds.

Followers who attack the regime need to bear a cost of participating, $c_f > 0$. Followers join the movement because they benefit from the new policy, that is, the reward from success is positive. Further, we assume that they attach a higher weight, $k_f > 0$, to the reward from success than do bystanders. The payoff to a follower of the mass movement is:

$$U_f(x_i) = U_b(x_i) + k_f G(A)\left(u(y_1, \theta, x_i) - u(y_0, \theta, x_i)\right) - c_f.$$

The additional weight k_f represents the extra psychological reward received by the revolutionaries, and captures the notion of the "pleasure in agency," which is conceptualized

[11] The assumption that the opposition can commit to a policy proposal is common in models of electoral politics (e.g., Wittman 1983; Lindbeck and Weibull 1987). In revolutionary movements, political developments are often more chaotic, and the ability to carry out the announced policy after the rebels come to power may be curtailed. Nevertheless, it would be unrealistic to assume that revolution leaders can completely ignore their prerevolutionary promises with impunity. In this paper, we abstract from the issue of commitment.

in Wood (2003) and formalized subsequently in Morris and Shadmehr (2020). It refers to "the value they [revolutionaries] put on being part of the making of history" (Wood 2003, p. 38). Morris and Shadmehr (2020) stress that such a subjective value arises from the authorship of the changes in society, even though each participant cannot wield influence on the likelihood of success. Our formulation of the payoff structure corresponds to this notion.[12] We denote the *follower's premium* as $FP(x_i) := U_f(x_i) - U_b(x_i)$. Given θ, A, and y_1, citizen i chooses $a_i = 1$ if and only if $FP(x_i) \geq 0$.

The leader is involved in organizing the opposition movement and formulating a policy alternative y_1 and incurs a cost of $c_l > 0$ for these leadership activities. The leadership role is also more rewarding: heavier involvement in the movement entails that the pleasure in agency has a stronger intensity, represented by an extra weight $k_l > 0$ attached to the reward from success. The payoff to a leader is:

$$U_l(x_i) = \max\{U_b(x_i), U_f(x_i)\} + k_l G(A)\left(u(y_1, \theta, x_i) - u(y_0, \theta, x_i)\right) - c_l. \tag{1}$$

We denote the *leader's premium* as $LP(x_i) := U_l(x_i) - \max\{U_b(x_i), U_f(x_i)\}$. The leader's premium plays a role in determining whether a citizen chooses to be a leader or not, a decision that we analyze in Section 4. In this section, the leader chooses a reform agenda y_1 to maximize $U_l(x^m)$, while her preference x^m is known to citizens.

Throughout this paper, we maintain the assumption that the probability of success $G(A)$ is strictly increasing and weakly log-concave in A, with $G(0) > c_f/k_f$. Diminishing returns from having more attackers implies log-concavity of G. It is plausible, however, that successful revolts may require a critical mass of attackers, meaning that the success determination technology may exhibit increasing and then decreasing returns. Our assumption can accommodate this type of success determination technology, because log-concavity is

[12] If $k_f = 0$, the free-riding problem would be so severe that no one ever participates in a mass movement even when the cost of doing so is negligible. Our assumption that those who take costly political action can derive an extra portion of the reward from success is a common device used to rationalize the motive for taking part in collective action and to abstract from free-riding issues when modeling citizens as atomistic agents. DellaVigna, List, Malmendier and Rao (2017) examine the hypothesis that people vote because they derive pride from telling others that they voted. This behavioral motive in the context of voting is the counterpart to the pleasure in agency from participation in large political movements.

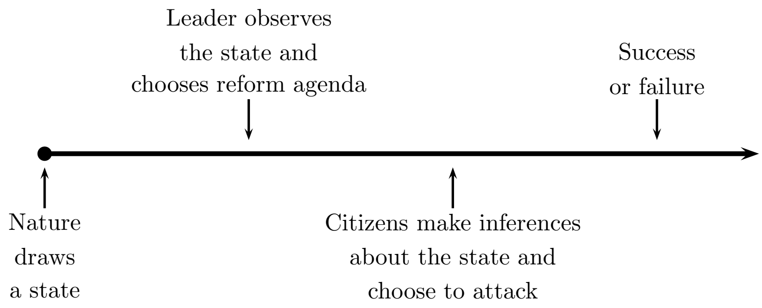


Figure 1. *Timeline of the game with asymmetric information.*

consistent with increasing returns as long as $G'(A)/G(A)$ is nonincreasing. The second part of our assumption (i.e., $G(0) > c_f/k_f$) is a sufficient condition that guarantees the existence of nontrivial equilibria with a positive mass of attackers. This condition requires that $G(0)$ be positive. A regime facing mass discontent may implode or collapse due to many forces (e.g., internal strife among elites, economic pressure, or foreign intervention) other than the actions of revolutionary leaders and followers. We can interpret $G(0)$ as the probability that these "background factors" bring down the regime.[13]

In this game with asymmetric information, nature picks the state first. Then the leader observes the state and proposes an agenda. Citizens observe the agenda, make inferences about the state, and make participation decisions. Finally, the revolt can be a success or failure, depending on the mass of attackers and some random factors. The timeline of the game is summarized in Figure 1.

3.2. Equilibrium of the protest game

In the following, we solve the model backwards. We first deal with the last stage, where citizens *correctly* infer about the state from the agenda proposed by the opposition leader and decide to attack the regime or not, taking the agenda as given. We then analyze the agenda decision of the leader under full information and asymmetric information in Sections 3.3 and 3.4.

Given the status quo policy y_0 and the reform agenda y_1, citizens decide whether to

[13] An equilibrium with a positive mass of attackers can exist even when $G(0) \leq c_f/k_f$. For example, for $G(A) = A$, we can show that such an equilibrium exists if $c_f/k_f < 1/4$.

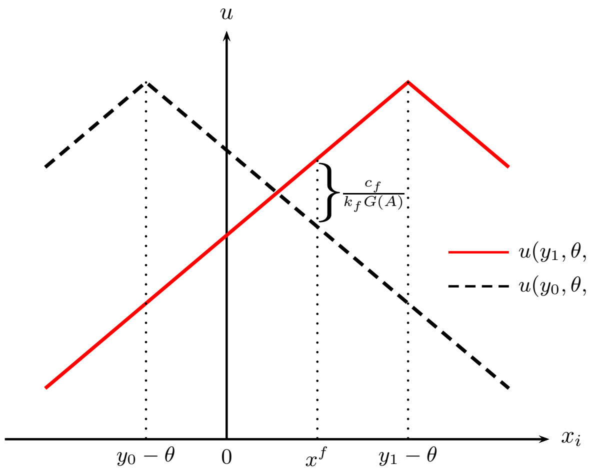


Figure 2. *The determination of the marginal attacker x^f. The utility difference between the solid line and the dashed line represents the reward from success for citizens with preference x_i. The follower's premium is nonnegative if and only if $x_i \geq x^f$.*

attack the regime or not in a simple protest game, when the state θ is learnt and becomes common knowledge. For any expected size of attack A, citizen i chooses $a_i = 1$, if and only if the follower's premium (defined in Section 3.1) $FP(x_i; A, y_1) \geq 0$. The function $FP(\cdot; A, y_1)$ is weakly increasing; therefore there exists a *marginal attacker*, denoted x^f and satisfying $FP(x^f; A, y_1) = 0$, such that citizen i attacks if and only if $x_i \geq x^f$. Figure 2 illustrates the determination of the marginal attacker x^f given a pair of policies y_0 and y_1.

Under this decision rule and the uniform distribution assumption, the total mass of attackers is $A = 1 - x^f$. The equilibrium of the protest game can be characterized by the indifference condition for the marginal attacker, which can be written as:

$$FP(x^f; 1 - x^f, y_1) = k_f G(1 - x^f)\big(u(y_1, \theta, x^f) - u(y_0, \theta, x^f)\big) - c_f = 0. \tag{2}$$

To emphasize its dependence on the reform agenda and the state, we use $x^f(y_1; \theta) \in (0, 1)$ to denote the equilibrium marginal attacker that satisfies condition (2). If $FP(0; 1, y_1) \geq 0$, then $x^f(y_1; \theta) = 0$, and all citizens attack. If $FP(1; 0, y_1) < 0$, then $x^f(y_1; \theta) = 1$, and no one attacks.

Lemma 1. *There exists $y_{\min}(\theta) \in (y_0, 1 + \theta)$ such that (a) if $y_1 < y_{\min}(\theta)$, then the only equilibrium is a no-attack equilibrium; and (b) if $y_1 \geq y_{\min}(\theta)$, then there exists only one* *equilibrium with a positive mass of attackers and with $x^f(y_1;\theta) \leq y_1 - \theta$. Further, in case (b), the mass of attackers increases when the agenda is less radical or when the state is higher.*

In case (a) of Lemma 1, the reward from success is small if the reform agenda y_1 is close to the status quo y_0. Given a positive cost of participation, no one chooses to participate in the movement. In case (b), the reform agenda y_1 is sufficiently far from the status quo, that guarantees at least a fraction of the relatively more radical citizens would find it worthwhile to attack; i.e., a nontrivial equilibrium exists. In this case, there may be multiple equilibria, one with $x^f \leq y_1 - \theta$ and another with $x^f > y_1 - \theta$.[14] Throughout the paper, we focus on the equilibrium with the largest equilibrium attack size (i.e., the one with $x^f \leq y_1 - \theta$). This is the only interior equilibrium in which the cutoff strategy (i.e., attack if and only if $x_i \geq x^f$) can be reasonably justified, and it is stable with meaningful comparative statics.

In a nontrivial equilibrium of the protest game, a more radical reform agenda y_1 suppresses participation. First, because $x^f + \theta \leq y_1$, the reform agenda is already to the right of the ideal policy of the marginal attacker. Raising y_1 further would make the agenda even less appealing to this citizen. Thus, holding A fixed, the marginal attacker shifts to the right. Second, as fewer citizens participate (i.e., A decreases), the follower premium falls for every citizen, which shifts the marginal attacker to the right even further. Consequentially, the marginal attacker becomes more radical and the probability of success falls correspondingly.

Equally important, a larger mismatch between the state and status quo policy causes more citizens to prefer the proposed policy y_1 to the status quo y_0, because the reward from success is increasing in the state θ for any $y_1 > y_0$. Lemma 1 shows that the size of the attack is larger when θ is higher, which implies that the probability of success, $G(1 - x^f(y_1;\theta))$, increases in θ. Such a comparative static is particularly useful when the agenda decision is analyzed.

[14] Multiple equilibria may exist because the protest game is a coordination game with strategic complementarity: the payoff from attacking rises as more citizens choose to attack.

3.3. Strategic reform agenda in the full information benchmark

Now we turn to the agenda-making stage. The crux of our theory is that the agenda proposed by the leader has the dual role of informing citizens about the state and laying down a policy alternative that matters for their welfare after the regime change. In this section, we study a benchmark case in which we disentangle the informational role from the material role of the agenda by assuming that the state is known even before the leader announces the agenda. The results obtained will be utilized and contrasted with those obtained in the subsequent section, where the impact of asymmetric information is studied.

When the state θ is common knowledge, the leader simply chooses y_1 to maximize her own payoff $U_l(x^m)$, taking into account its impact on the protest game, characterized in Lemma 1. The leader's maximization problem can be written as:

$$\max_{y_1} \quad u(y_0,\theta,x^m)+\kappa G(1-x^f(y_1;\theta))[u(y_1,\theta,x^m)-u(y_0,\theta,x^m)]-c_l, \tag{3}$$

where $\kappa = 1+k_f+k_l$ if she chooses to attack in the protest game and $\kappa = 1+k_l$ if he chooses not to attack.

We restrict our attention to the interesting case where the leader's ideal policy is large enough (i.e., $x^m+\theta > y_{\min}$).[15] The leader would never find it optimal to choose an agenda which is more radical than her ideal policy (i.e., $y_1 > x^m+\theta$), because she can always choose a less radical policy that is closer to her ideal policy and that gives rise to a higher probability of success. She would not choose a very moderate agenda, either (i.e., $y_1 < y_{\min}$), because such an agenda would induce no attack in the protest game. For $y_1 \in [y_{\min}, x^m+\theta]$, the objective function (3) is quasi-concave; therefore, the optimal agenda y_1^* is unique and is characterized by the first-order condition. We also show that it is not optimal to choose $y_1 = y_{\min}$. The optimal agenda y_1^* satisfies:

$$1-\frac{G'(1-x^f(y_1^*;\theta))}{G(1-x^f(y_1^*;\theta))}\frac{\partial x^f(y_1^*;\theta)}{\partial y_1}(y_1^*-y_0)\geq 0, \tag{4}$$

[15] If $x^m+\theta \leq y_{\min}$, then the optimal agenda is to choose $y_1 = y_{\min}$ as long as the gap between $G(1-x^f(y_{\min}))$ and $G(0)$ is not too small.

with $y_1^* = x^m + \theta$ if (4) holds as a strict inequality.

The leader faces a trade-off between the reward from success and the chances of success. The marginal utility from having a policy closer to the leader's ideal policy is 1. The marginal cost is that the chance of obtaining the reward from success is lowered as $\partial x^f / \partial y_1$ is positive by Lemma 1. The first-order condition (4) optimally balances this trade-off. Because this optimal agenda y_1^* reflects the leader's concern about the equilibrium outcome in the subsequent protest game, we sometimes refer to it as a *strategic agenda*, and we write $y_1^*(x^m; \theta)$ to emphasize its dependence on the leader's preference and on the state.

Proposition 1. *For any leader with $x^m \geq y_{\min}(\theta) - \theta$, there exists a unique cutoff $\hat{x}(\theta)$ such that the strategic agenda is:*

$$y_1^*(x^m; \theta) = \begin{cases} x^m + \theta & \text{if } x^m < \hat{x}(\theta); \\ \hat{x}(\theta) + \theta & \text{if } x^m \geq \hat{x}(\theta). \end{cases} \tag{5}$$

Further, a higher state makes the strategic agenda more radical, but increases the probability of success in the protest game.

When the leader's preference x^m is relatively moderate (i.e., $x^m < \hat{x}(\theta)$), she can afford to choose her own ideal policy (i.e., $x^m + \theta$) without suppressing citizens' support very much; i.e., the optimal solution $y_1^*(x^m; \theta)$ is a corner solution. When the leader's preference is relatively radical (i.e., $x^m \geq \hat{x}(\theta)$), choosing her ideal policy would discourage too many moderate citizens from joining the mass movement. Therefore, the leader must compromise by choosing the agenda $y_1^* = \hat{x}(\theta) + \theta$, which is less radical than her ideal policy. In general, the strategic agenda $y_1^*(x^m; \theta)$ is weakly increasing in x^m.[16]

The second part of the proposition says that the opposition leader chooses a more radical agenda in response to a higher state, but would never raise the equilibrium agenda y_1^* to the point where it hurts the chances of success. In other words, $G(1 - x^f(y_1^*(x^m; \theta); \theta))$

[16] The fact that the strategic policy y_1^* is constant for $x^m \geq \hat{x}(\theta)$ is a consequence of the piecewise linear utility function. With the quadratic utility function, the reward from success strictly increases in the leader's preference x^m, and so does her strategic policy. The two specifications deliver very similar results, but the linear specification is more tractable.

always increases in θ. In connection to the two states specified in our model, we let $y_1^{L*} := y_1^*(x^m; \theta_L)$ and $y_1^{H*} := y_1^*(x^m; \theta_H)$ represent the strategic agenda chosen by the leader with preference x^m in the low and high states, respectively, under full information. Further, let $G^{L*} := G(1 - x^f(y_1^{L*}; \theta_L))$ and $G^{H*} := G(1 - x^f(y_1^{H*}; \theta_H))$ represent the corresponding probabilities of success. Proposition 1 states that $y_1^{H*} > y_1^{L*}$ and $G^{H*} > G^{L*}$.

The implication of this result is important for our subsequent analysis. Once citizens believe that the environment is more favorable for a revolt (i.e., the state is high), the leader can choose a more radical agenda, but the chances of success are also greater. That provides the leader an incentive to mislead citizens when the leader knows the state but citizens do not know. In the next section, such an incentive will be the focus of our analysis. We will also show that, under some circumstances, asymmetric information may cause the leader to "over-react": she chooses a policy so radical that it reduces the probability of success, despite the state being more favorable to the mass movement.

3.4. Radical reform agenda in separating equilibrium

In this section, we analyze the case of asymmetric information, in which only the leader knows the state but citizens do not. In this setting, the leader lacks an effective means to convince citizens that it is her information rather than her own interest that drives her choice of agenda. She has to rely on the reform agenda to credibly signal her private information. Such a signaling mechanism is costly, because the agenda has to be distorted to command credibility.

In the following formal analysis, we restrict our attention to leaders with a preference higher than a threshold (specifically, $x^m \geq y_{\min}(\theta_L) - \theta_L$) to ensure that in each state, there is a positive mass of attackers in the equilibrium of the protest game.

Lemma 2. *There exists a unique threshold $x^\dagger \in (\hat{x}(\theta_H), \hat{x}(\theta_H) + \theta_H - \theta_L)$ such that the full-information outcome (y_1^{L*}, y_1^{H*}) is an equilibrium outcome under asymmetric information if and only if the leader is moderate, i.e., $x^m \leq x^\dagger$.*

The leader in the low state may have an incentive to mislead citizens into believing that the state is high, and the more radical her preference is, the greater the incentive to mislead. First, for the same reform agenda, the mass of attackers is larger if citizens believe that the state is high (Lemma 1) when it is actually low. Second, when the leader misleads citizens with y_1^{H*}, the reward from success could be either larger or smaller, depending on whether y_1^{H*} or y_1^{L*} is closer to her ideal policy. Specifically, by choosing y_1^{H*} instead of y_1^{L*} in the low state, the reward from success would change from $y_1^{L*} - y_0$ to $2x^m + 2\theta_L - y_1^{H*} - y_0$. For a leader with low x^m, the policy y_1^{H*} is too far from her ideal policy in the low state, and the net change is negative. For a leader with high x^m, the policy y_1^{H*} can be closer to her ideal policy in the low state, and the net change is positive. In general, the net change is larger when the leader is more radical. Therefore, in the low state θ_L, a leader with a moderate preference (small x^m) faces a trade-off between a higher chance of success and a lower reward from success when she lies. However, both are higher for a leader with a radical preference (large x^m). Lemma 2 shows that there exists a unique $x^\dagger$ such that the concern over a lower reward dominates the benefit of a higher chance of success if and only if $x^m \leq x^\dagger$. We say that the leader is moderate if $x^m \leq x^\dagger$, and radical otherwise.

Suppose the leader's preference is moderate ($x^m \leq x^\dagger$); she thus has no incentive to mislead, and citizens know this. In equilibrium, the leader would choose the strategic agenda y_1^{L*} in the low state and the strategic agenda y_1^{H*} in the high state, and citizens correctly infer about the state. In other words, the leader indeed behaves "optimally," this is, as she would behave under full information.

Suppose the leader's preference is radical ($x^m > x^\dagger$); she thus has an incentive to mislead by proposing y_1^{H*} in the low state. Of course, citizens would not be fooled into believing that the state is high simply because they observe the leader choosing the reform agenda y_1^{H*}. In this case, the full-information outcome cannot be supported as an equilibrium outcome under asymmetric information. In the following proposition, we characterize the least-cost separating equilibrium that satisfies the D1 criterion (Banks and Sobel 1987; Cho and Kreps 1987). We show that the agenda chosen must be even more radical than y_1^{H*} in the high state, so that the leader would not have an incentive to mimic the high state by choosing this very radical policy when the state is low. In other words, the radical leader cannot behave "optimally" in the high state; she has to resort to "irrational radicalism" to separate herself from leaders who may otherwise bluff in order to boost support.

To formalize, let $(\hat{y}_1^L, \hat{y}_1^H)$ represent the agenda choices in the separating equilibrium in the respective states, and let $\hat{G}^L := G(1-x^f(\hat{y}_1^L;\theta_L))$ and $\hat{G}^H := G(1-x^f(\hat{y}_1^H;\theta_H))$ represent the corresponding probabilities of success.

Proposition 2. *Suppose the leader is radical (i.e., $x^m > x^\dagger$). In the least-cost separating equilibrium that satisfies the D1 refinement, compared with the full information benchmark, we have the following: (i) in the low state, the equilibrium agenda remains the same (i.e., $\hat{y}_1^L = y_1^{L*}$ and $\hat{G}^L = G^{L*}$); and (ii) in the high state, the equilibrium agenda is more radical and chances of success are lower (i.e., $\hat{y}_1^H > y_1^{H*}$ and $\hat{G}^H < G^{H*}$). Further, (iii) as the leader's preference x^m is more radical, the agenda $\hat{y}_1^H$ is strictly more radical.*

Part (i) says that, in a separating equilibrium, the leader would not propose an agenda different from y_1^{L*} in the low state. Any agenda other than optimal strategic policy would give her a lower payoff as long as the low state is correctly inferred by citizens.

Part (ii) establishes a contrasting result in the high state. For any leader with $x^m > x^\dagger$, proposing an agenda equal to y_1^{H*} in the high state will not be an equilibrium (Lemma 2). To prevent her from overstating the favorability of the state, the equilibrium policy $\hat{y}_1^H$ must exceed y_1^{H*}.[17] Such an equilibrium agenda is so high that it is to the right of the leader's ideal policy $x^m + \theta_L$ in the low state, and it significantly reduces the probability of success, which makes proposing $\hat{y}_1^H$ unattractive in the low state. In other words, in the high state, to convince citizens of the state, a radical leader must pursue a policy that is too radical and too costly at the expense of losing too many followers.

The reform agenda in the high state under the least-cost separating equilibrium that satisfies the D1 refinement is the $\hat{y}_1^H$ (greater than y_1^{H*}), such that the leader's incentive

[17] For some parameter values, it is possible that there exists a $\tilde{y} < \hat{y}_1^L$ such that the leader is indifferent between choosing $\tilde{y}$ and inducing citizens to believe the state is high, and choosing $\hat{y}_1^L$ and inducing them to believe the state is low. However, $\hat{y}_1^H = \tilde{y}$ cannot be an equilibrium. The D1 criterion requires that an off-equilibrium agenda $y' \in (x^m + \theta_L, x^m + \theta_H)$ be ascribed to a high type, which would then give an incentive to the high type to deviate from $\tilde{y}$ to y'. See the proof of Proposition 2.

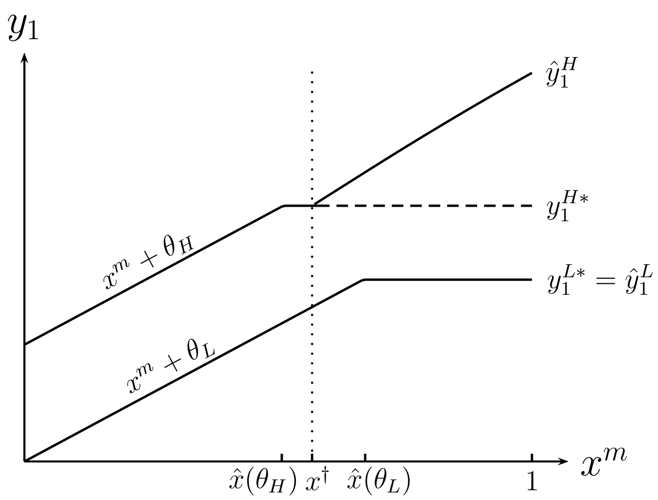


Figure 3. *The strategic agenda pair is (y_1^{L*}, y_1^{H*}) under full information. When $x^m > x^\dagger$, the leader chooses $\hat{y}_1^L = y_1^{L*}$ and $\hat{y}_1^H > y_1^{H*}$ in a separating equilibrium, and $\hat{y}_1^H$ strictly increases in x^m.*

constraint in the low state is just binding:

$$\hat{G}^L \left(u(\hat{y}_1^L, \theta_L, x^m) - u(y_0, \theta_L, x^m)\right) - \hat{G}^H \left(u(\hat{y}_1^H, \theta_L, x^m) - u(y_0, \theta_L, x^m)\right) = 0. \tag{6}$$

When the leader is indifferent between $\hat{y}_1^L$ and $\hat{y}_1^H$ in the low state, she strictly prefers $\hat{y}_1^H$ to $\hat{y}_1^L$ in the high state.[18] Figure 3 illustrates the equilibrium outcome.

In the least-cost separating equilibrium, the chances of success in the high state must be lower than that in the full information setting (i.e., $\hat{G}^H < G^{H*}$). That is because the reform agenda in the high state is more radical, and the mass of attackers decreases with a more radical agenda (Lemma 1). In the high state, the leader cannot moderate her position, even though she knows that her radical agenda discourages citizens: citizens would interpret an agenda less radical than $\hat{y}_1^H$ as an opportunistic deviation by a leader who attempts to exaggerate the state by choosing a high agenda in the low state.

Part (iii) of Proposition 2 characterizes the comparative statics of the equilibrium $\hat{y}_1^H$ with respect to x^m. As the leader becomes more radical, her incentive to exaggerate the state and mislead citizens increases. As a result, she needs to propose even more radical

[18] The single-crossing condition for the separating equilibrium is obtained from the fact that the utility function $u(y_1, \theta, x^m)$ is supermodular in y_1 and θ. Because the reward from success is higher in the high state than in the low state, the leader strictly prefers $\hat{y}_1^H$ to $\hat{y}_1^L$ in the high state if she is indifferent between these two agendas in the low state.

agendas in the high state in order to remain credible. Interestingly, when the leader is very radical, the more favorable state does not necessarily guarantee a better chance of success because the agenda chosen to signal the state is too radical.

A corollary to Proposition 2 is that there exists $x^{\dagger\dagger} > x^{\dagger}$ such that, if the leader's preference is more radical than $x^{\dagger\dagger}$, then the chances of success are smaller in the high state than in the low state. Because $\hat{y}_1^H$ increases with x^m at a nonvanishing rate for a large enough x^m, and $\hat{y}_1^L$ is constant with respect to x^m for $x^m > \hat{x}(\theta_L)$, we must have $\hat{y}_1^H - 2\theta_H > \hat{y}_1^L - 2\theta_L$ when x^m exceeds a threshold value $x^{\dagger\dagger}$. By the indifference condition (2) for the marginal attacker (in Section 3.2), this inequality implies that $x^f(\hat{y}_1^H;\theta_H) > x^f(\hat{y}_1^L;\theta_L)$, and hence $\hat{G}^H < \hat{G}^L$. Such an equilibrium outcome under asymmetric information is qualitatively different from that under full information, where we show that $G^{H*} > G^{L*}$. That is, the leader who does not suffer from the citizens' suspicion always proposes a more radical agenda given a higher state, but she never "over-reacts" by choosing a policy so radical that it hurts the chances of success.

4. Endogenous Leadership

In the previous section, we showed that a radical leader may have to radicalize her agenda to signal that she is not lying to gather support, while a moderate leader is not plagued by such suspicion. A relevant question is, under what conditions would radicals rather than moderates assume the leadership role and propose the reform agenda? Before dealing with this question, one must acknowledge that there can be plenty of plausible routes to opposition leadership, including by sheer chance in many historical events. In this section, we analyze one particular mechanism of how the leadership of a movement may endogenously emerges: citizens can choose to join the leadership, and a reform agenda is proposed as a compromise among leaders with diverse interests. Interestingly, this mechanism of leadership selection and the mechanism of signaling reinforce each other, which, under some circumstances, produces an equilibrium outcome whereby radicals indeed become leaders.

4.1. Leader selection

We enrich the model in Section 3 by allowing the opposition leadership to comprise a group of citizens instead of a single leader and by making the choice to join the leadership group endogenously. Specifically, before observing the state, each citizen only knows its distribution and can make an additional decision to join the leadership of the mass movement ($l_i = 1$) or not ($l_i = 0$). In each state, the payoff to any citizen who chooses to be a leader is given by equation (1) in Section 3.1.[19]

If no citizens choose to join the leadership, then the status quo will prevail, and therefore there will be no revolt. If a group of citizens decide to join the leadership, their preferences within the leadership group must be diverse. We assume that it is the one with the median preference among the leadership group who will formulate and propose the reform agenda for the mass movement. In other words, the task of drafting the policy proposal is delegated to the median citizen in the group, who will be referred to as the *median leader* in the following. We continue to use x^m to denote the preference of the median leader.

Citizens choose to join the leadership before they know the state. After the median leader emerges from the leadership group, she observes the state, makes a proposal, and announces it to all the other citizens. Upon hearing the proposal, each citizen makes inferences about the state and decides whether to participate in attacking the regime or not. That is, once the median leader acquires information about the state, the rest of the game is similar to the one in Section 3. The timing of the game with endogenous leadership is summarized in Figure 4. The equilibrium of this static game is a fixed point, in which the preference of the median leader is endogenously determined and known to other citizens.

The assumptions related to the emergence of leaders deserve discussion. First, the assumption that the median leader chooses the agenda captures the idea that the proposed agenda should reflect the preferences of the leadership group as a whole. This can be for-

[19] In this paper, by joining the leadership group, citizens pay higher costs but derive the pleasure in agency of stronger intensity. Further, there is a continuum of citizens who participate in the leadership in equilibrium. Therefore, we rule out the possibility that citizens may manipulate the formation of the leadership for instrumental benefits. Strategic leadership formation is an interesting topic to pursue on its own.

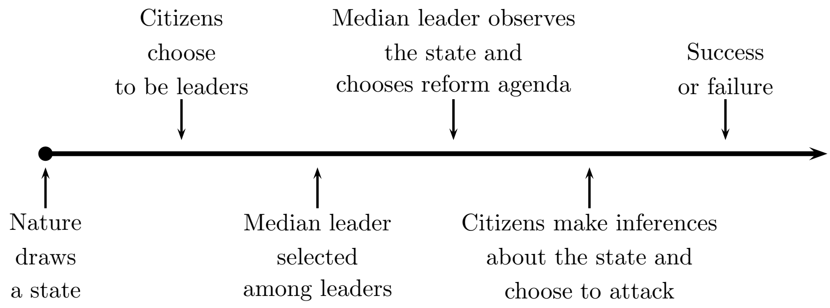


Figure 4. *Timeline of the game with asymmetric information and endogenous leadership.*

mally justified by a democratic decision-making process within the leadership group, in which the median leader's preference prevails in any pairwise vote between reform agendas. This assumption appears to be rather specific but is a useful shortcut for representing two general conditions: (i) the policy proposed by the leadership group is more radical than the policy preferred by the least radical leader in the group, and (ii) the proposed policy becomes more radical if a fraction of the least radical leaders are removed from the group. Conceptually, the precise policy formulation mechanism matters little for the subsequent analysis and results, as long as conditions (i) and (ii) hold, but our setting is more tractable.[20]

Second, in this model, there will be a continuum of leaders. Only the median leader chooses the agenda; what she proposes affects the payoffs of all leaders but not the other way around, because each individual leader is atomistic and her decision to become a leader or not has no influence on who will be selected as the median leader. This assumption is made to simplify the characterization. Even in a setting with a finite number of leaders, the agenda-making mechanism can still satisfy conditions (i) and (ii), in which case our results and analysis continue to hold.[21]

Finally, regarding the timing, we assume that citizens do not observe the state before

[20] For example, if the reform agenda is chosen by someone with a preference equal to the mean rather than the median, or if the leader is selected at a quantile different from the median, or if the representative leader is chosen purely randomly, our results will remain largely intact.

[21] A more realistic approach is that the number of proposal-drafting leaders is finite and their policy preferences have impacts on the final reform proposal. Consider the setting where the proposal is generated from a Nash bargaining game with a finite number of leaders, a standard strategic game for collective decisions. Given that, in our model, the leader's premium is monotonic and single-crossing from below, both conditions are satisfied. However, even richer possibilities that violate the conditions, such as some forms of reconciliation or even struggle among leaders, are not captured by this extension.

they choose to join the leadership, which is reasonable but not crucial for our results. This is reasonable because political mass movements are often sudden outbreaks, and individuals who become leaders may acquire information about the political situation after they assume their role.[22] It is not crucial, because we can also allow a small fraction of randomly drawn citizens to observe the state before they decide whether to participate in the leadership group. In a separating equilibrium where the rest of citizens still infer the state from the reform agenda, the main results and properties of leadership formation will be similar.

4.2. Equilibrium analysis

Because citizens do not know the state at the time that they choose to become leaders, the determination of endogenous leadership is based on the expected value of the leader's premium. Specifically, if the preference of the median leader is conjectured to be equal to x^m, citizens expect that the proposed agenda is $\hat{y}_1^L(x^m)$ or $\hat{y}_1^H(x^m)$, depending on the state. The expected leader's premium for a citizen with preference x_i is

$$\hat{LP}(x_i; x^m) := \sum_{j=L,H} \pi_j LP(x_i; \hat{y}_1^j(x^m), \theta_j),$$

where

$$LP(x_i; y_1, \theta) = \begin{cases} k_l G\big(1 - x^f(y_1; \theta)\big)(2x_i + 2\theta - y_1 - y_0) - c_l & \text{if } x_i + \theta < y_1, \\ k_l G\big(1 - x^f(y_1; \theta)\big)(y_1 - y_0) - c_l & \text{if } x_i + \theta \geq y_1. \end{cases}$$

For a given x^m, citizen i chooses to join the leadership group if and only if the expected leader's premium is nonnegative. Since $LP(x_i; y_1, \theta)$ is weakly increasing in x_i, $\hat{LP}(x_i; x^m)$ is also weakly increasing in x_i, which justifies a monotone strategy. That is, if there exists a *marginal leader* x^l such that $\hat{LP}(x^l; x^m) = 0$, then citizen i chooses $l_i = 1$ if and only if $x_i \geq x^l$. If $\hat{LP}(0; x^m) > 0$, then all citizens choose to become leaders; and if $\hat{LP}(1; x^m) < 0$, then no citizen chooses to be a leader. Of course, the preference of the marginal leader

[22] For example, the student leaders of the 1989 Tiananmen democracy movement were ordinary students before the outbreak of protests. Once they became leaders of the movement, they started connecting with various intellectuals and public officials and collecting information about the situation and the movement, and then made political demands that set course of the movement (Ogden et al. 1992).

depends on x^m, and we write $x^l(x^m)$ to emphasize this dependence.

Because the distribution of preferences x_i is uniform, if the marginal leader has preference $x^l(x^m)$, then the median preference of the leadership group is:

$$M(x^m) = \frac{1}{2}\left(x^l(x^m) + 1\right).$$

The equilibrium in this model is characterized by a 6-tuple, $(x^m_*, x^l_*, x^f_{L*}, x^f_{H*}, \hat{y}^L_{1*}, \hat{y}^H_{1*})$, that satisfies the following requirements:

1. The median leader's preference satisfies $x^m_* = M(x^m_*)$.
2. The marginal leader's preference satisfies $\hat{LP}(x^l_*; x^m_*) = 0$.
3. The reform agenda in the low state solves the maximization problem (3) and is given by $\hat{y}^L_{1*} = y^*_1(x^m_*; \theta_L)$ in equation (5).
4. If $x^m_* \leq x^\dagger$, the reform agenda is given by $\hat{y}^H_{1*} = y^*_1(x^m_*; \theta_H)$ in equation (5); otherwise it is given by the solution to the binding incentive constraint (6) for $x^m = x^m_*$ and $\hat{y}^L_1 = \hat{y}^L_{1*}$.
5. The marginal attacker in state $j \in \{H, L\}$ satisfies $FP(x^f_{j*}; 1 - x^f_{j*}, \hat{y}^j_{1*}, \theta_j) = 0$.

In this equilibrium, citizen i chooses $l_i = 1$ if and only if $x_i \geq x^l_*$, and she chooses $a_i = 1$ in state $j \in \{H, L\}$ if and only if $x_i \geq x^f_{j*}$. The probability of success in state j is $G(1 - x^f_{j*})$.

The key to our model is that the leadership premium $\hat{LP}(x_i; x^m)$ is weakly increasing in x_i, and any citizen with preference $x_i \geq x^l(x^m)$ will choose to be a leader. That means that the median leader x^m in general is more radical than the marginal leader x^l (i.e., condition (i) holds); and the median leader x^m is more radical if the marginal leader x^l is more radical (i.e., condition (ii) holds).

The two conditions combined give rise to a mechanism for the determination of the leadership, in which the mapping $M(x^m)$ is increasing in x^m. To understand the positive slope, consider the following comparative statics on an increase in the preference of the median leader (i.e., a larger x^m). The median leader would propose more radical agendas;

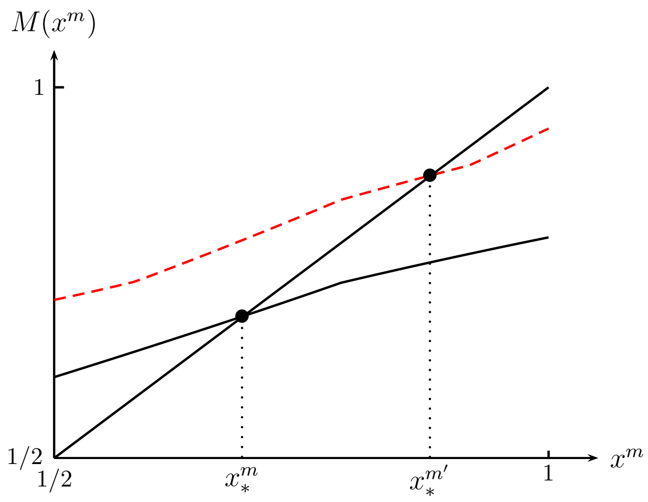


Figure 5. *The solid black curve represents $M(x^m)$, and x^m_* is the fixed point. When the cost of joining the leadership is higher, $M(x^m)$ shifts up and the equilibrium x^m_* increases.*

i.e., $\hat{y}_1^L(x^m)$ and $\hat{y}_1^H(x^m)$ are both higher (Propositions 1 and 2), and this, in turn, hurts the marginal leader who has less radical preferences than the median one. As a result, a more radical citizen would become the marginal leader; i.e., $x^l(x^m)$ shifts to the right as x^m increases. The distribution of preferences of the leadership group therefore becomes more radical, which, consequentially, gives rise to a more radical median leader. In other words, the leadership group will be more radical if the conjectured median leader becomes more radical. The solid line in Figure 5 illustrates.

Proposition 3. *Under some sufficient conditions guaranteed by neither c_f/k_f nor c_l/k_l being too large, the mapping $M(\cdot)$ is continuous and increasing on $[1/2, 1]$, and there exists an equilibrium with $x^m_* \in (1/2, 1)$ such that a positive mass of citizens choose to attack in each state. Further, if $\theta_H - \theta_L \le 1/3$, the equilibrium is unique.*

To ensure the existence of a nontrivial equilibrium in the protest game, we require that c_f/k_f be smaller than a threshold.[23] To ensure that the leadership is not empty in equilibrium, we require that c_l/k_l be smaller than a threshold, so that a citizen with preference $x_i \in [1/2, 1]$ will choose to be a leader if she expects herself to be the median leader who proposes the reform agendas.[24] Under these conditions, Proposition 3 establishes that

[23] In the proof of Proposition 3, we impose the condition $y_{\min}(\theta_L) - \theta_L < 1/2$, which implicitly pins down the upper bound for c_f/k_f.

[24] In the proof of the proposition, we impose the condition $\min\{\hat{LP}(1/2; 1/2), \hat{LP}(1; 1)\} > 0$. It guarantees

$M(x^m)$ is increasing and that a fixed point of the mapping exists.

The fixed point of $M(\cdot)$ is unique and stable if its slope is less than one. Intuitively, $M'(x^m)$ will be smaller if fewer leaders on the left end of the leadership drop out when the median leader becomes more radical. This is true when the proposed policy $\hat{y}_1^H(x^m)$ does not increase too much when the leader's preference x^m becomes more radical. Proposition 3 shows that the condition $\theta_H - \theta_L \leq 1/3$ is sufficient to ensure that $\partial \hat{y}_1^H(x^m)/\partial x^m$ is so small that the slope of $M(\cdot)$ is less than one. Therefore, the fixed point is also a globally stable one: convergence to the fixed point from any likely perturbations is guaranteed.

We have characterized the determination of the endogenous leadership group. In the following section, we investigate the conditions under which citizens with radical preferences take on the leadership roles and how this mechanism of leadership selection and the mechanism of signaling reinforce each other.

5. Discussion

5.1. Suppression and the opposition leadership

Facing the challenge of mass movements, regimes typically resort to harsher punishments to thwart the revolts, which may secure a higher chance of regime survival but may also radicalize the opposition leadership. Therefore, successful revolts can be less likely but more radical. This observation is consistent with the comparative statics of our model: a greater suppression of the mass movement by the regime may radicalize the leadership in equilibrium (i.e., the preference of the leader x_*^m may rise above $x^\dagger$, in response to harsher punishments), and the radical leader is therefore compelled to radicalize her agenda out of the signaling concern.

Consider first an increase in the cost of joining the leadership c_l/k_l, which reflects the harsher punishment imposed by the regime targeting leaders specifically. Other things equal, a higher c_l/k_l reduces the leader's premium and causes the marginal (and moderate) leaders to drop out. As the preference of the marginal leader x^l increases, the preference

that $\hat{LP}(x^m; x^m) > 0$ for all $x^m \in [1/2, 1]$, so that the mapping $M(\cdot)$ is well-defined in the relevant domain.

of the median leader of the remaining group, $(x^l + 1)/2$, also increases. In Figure 5, the curve representing $M(x^m)$ shifts up, from the solid line to the dashed line. In the new equilibrium, a more radical leader will be chosen, and the agendas chosen in both states will be more radical as well.

This result is consistent with a common conjecture about radicalism from the perspective of leader selection: high costs of leading movements filter out moderates, and only radicals remain. In addition, our model shows that the impact of higher c_l/k_l on the selection of leaders is amplified through the endogenous selection of leaders, which is absent in the aforementioned conjecture. This discussion also highlights why it is important to introduce an endogenous mechanism to determine leadership in our model. When the leader is exogenous and imposed, the cost of leadership does not matter for the equilibrium agendas. When citizens can choose to join the leadership or not, such a cost becomes important because it helps determine how radical the leadership group is.

Consider next an increase in the cost of participating in the protests c_f/k_f, which reflects the harsher punishment imposed by the regime targeting all participants. A higher participation cost gives rise to two effects. The direct effect is that fewer followers join the movement, reducing the chances of success and hence the expected payoff to the marginal leader. Therefore, the more moderate marginal leaders drop out and the median leader becomes more radical. The indirect effect is that for a fixed x^m, the median leader responds to lower participation by adjusting her strategic reform agenda. In particular, she may propose a more moderate agenda to partially compensate for the reduced chances of success, which benefits the marginal leader. The overall impact is theoretically ambiguous. However, for a large set of parameters, we find that the direct effect dominates the indirect effect: an increase in participation cost c_f/k_f leads to a more radical leadership group in equilibrium.[25]

The above comparative statics result is consistent with one popular view about why the leadership of democracy movement of China in 1989 further radicalized, after the Chinese

[25]Shadmehr (2015) predicts a similar result with a different mechanism: the leader may radicalize the agenda even further to induce more efforts from radicals, in response to the harsher punishment.

regime imposed martial law, which raised the cost of all participants. According to Ogden et al. (1992, p. 123), "only those willing to risk everything for a political cause would come forth for leadership. Inevitably, this meant that the more radical, more daring students would, at each crucial juncture, control the course of the student movement, and move it onto ever more precarious ground."

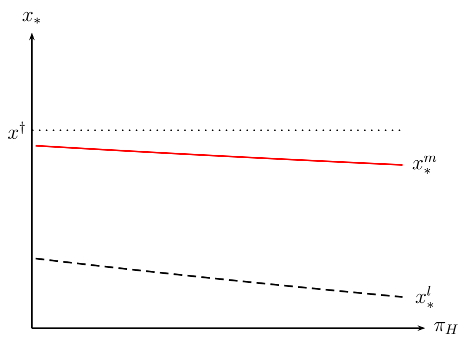
The new mechanism characterized in this section enriches the literature about the relationship between state repression and protests. On the one hand, the direct impact of harsher punishments on radicalization of opposition leadership in this model corresponds to "the short term effect" found empirically in Rasler (1996) that state repression discourages followers from participating in the Iran revolution. On the other hand, Rasler (1996) shows that there exists an indirect effect that state repression increases protests through spatial diffusion. Our model also characterizes a similar indirect effect but through a different mechanism—the leader responds to lower participation by strategically adjusting her reform agenda. It is likely that the strength of two opposing effects varied in different settings: while state repression triggered overall larger protests in the context of the Iran revolution, repressive measures might have radicalized the leadership of democracy movement of China in 1989.[26]

5.2. Structural roots hypothesis

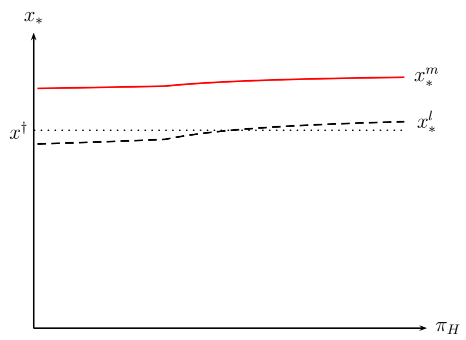
In this section, we turn to the comparative statics of equilibrium outcomes with respect to π_H, the probability of favorable state for a revolt. We show that how π_H affects the equilibrium chances of success depends on whether the median leader is moderate or radical. Such a comparison is particularly interesting, because predictions of the model are qualitatively different if $x_*^m < x^\dagger$ (with the leader choosing her agenda strategically) than if $x_*^m > x^\dagger$ (with the leader "irrationally radicalizing" her agenda to signal the state).

If the equilibrium leader is moderate (i.e., $x_*^m \leq x^\dagger$), then an increase in π_H reduces x_*^m and raises the chances of success in both states. Given a moderate leader, the adopted

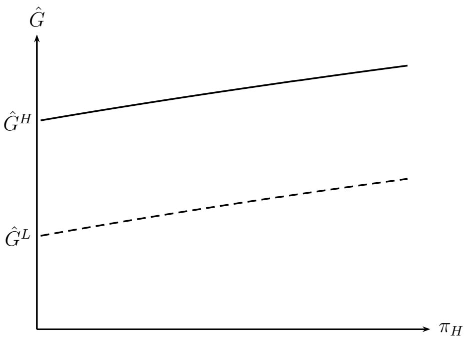
[26] In similar ways, our model can be related to discussions on the relationship between repression and terrorism (e.g., Dugan and Chenoweth 2012).



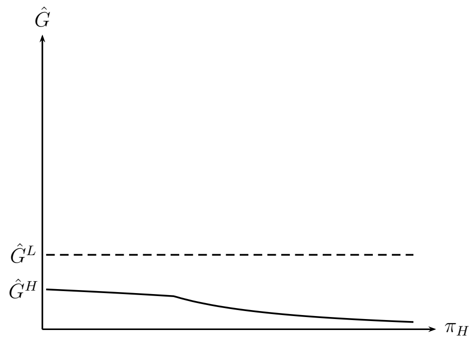
(a) c_l/k_l is low and leader is moderate



(b) c_l/k_l is high and leader is radical



(c) c_l/k_l is low and leader is moderate



(d) c_l/k_l is high and leader is radical

Figure 6. *Contrasting comparative statics with respect to π_H when the equilibrium leader is moderate or radical. In this example, the success determination technology is $G(A) = \Phi(1.5A - 1)$, where Φ is the standard normal distribution function. The other parameters are: $y_0 = 0$, $\theta_L = 0$, $\theta_H = 0.08$, and $c_f/k_f = 0.15$. In panels (a) and (c), we set $c_l/k_l = 0.1072$, and the leader is moderate in equilibrium. In panels (b) and (d), we set $c_l/k_l = 0.2573$, and the leader is radical in equilibrium.*

agendas in both states are the strategic ones, i.e., y_1^{L*} and y_1^{H*}. The marginal leader obtains a larger leader's premium in the high state than in the low state, because both the reward from success and the chances of success are larger in the high state. Therefore, an increase in π_H raises her expected leader's premium and encourages more citizens to become leaders, making the leadership group more moderate. Both the equilibrium x_*^l and x_*^m decrease in π_H. Further, the chances of success improve in both states, because the equilibrium agendas become less radical. Panels (a) and (c) of Figure 6 illustrate.

In contrast, when the equilibrium leader is radical (i.e., $x_*^m > x^\dagger$), such a prediction can be reversed: an increase in π_H raises x_*^m and reduces the chances of success in both states. Two conditions are sufficient for this reversed prediction: (i) $x_*^m + \theta_L > x_*^l + \theta_H$, and (ii) $x_*^l + \theta_L > \hat{y}_{1*}^L$. Condition (i) captures the situation that the interests of marginal and median leaders are sufficiently far apart. Together with the binding incentive constraint (6), this condition implies:

$$\hat{G}^L\left(\hat{y}^L_{1*} - y_0\right) = \hat{G}^H\left(2x^m_* + 2\theta_L - \hat{y}^H_{1*} - y_0\right) \geq \hat{G}^H\left(2x^l_* + 2\theta_H - \hat{y}^H_{1*} - y_0\right).$$

Further, condition (ii) implies that $\hat{y}^L_{1*}$ is to the left of the ideal policies of the marginal leader. Therefore, according to the definition of $LP(\cdot)$ in Section 4.2, the above inequality is equivalent to:

$$LP(x^l_*; \hat{y}^L_{1*}, \theta_L) \geq LP(x^l_*; \hat{y}^H_{1*}, \theta_H).$$

This inequality says that the reform agenda in the high state is radicalized to such an extent that it is too far from the marginal leader's ideal policy, so that her leader's premium in the high state is *smaller* than that in the low state. As a result, an increase in π_H (and a corresponding decrease in π_L) will reduce her expected leader's premium. This will cause the marginal leader to drop out of the leadership, further causing $M(x^m)$ to increase as the remaining leaders become more radical. The equilibrium response is that x^m_* will rise. Therefore, both $\hat{y}^L_{1*}$ and $\hat{y}^H_{1*}$ are non-decreasing in π_H; the probabilities of success in both states are (weakly) smaller.

We illustrate the case of $x^m_* > x^\dagger$ in panel (b) and (d) of Figure 6, by choosing a higher value of c^l/k^l than the one used in panels (a) and (c). In this example, conditions (i) and (ii) specified above are also satisfied. In such an equilibrium, the preference of the leader x^m_* increases in π_H and the success probabilities in the two states weakly decreases. Observe also that $\hat{G}^H < \hat{G}^L$ in this example: when the leader is very radical, the probability of success in the high state is lower than that in the low state.

This set of results also shed some new light on the debate about the role of structural factors in mass movements. The structural-roots hypothesis suggests that the fundamental social, economic, and political structures of society are the key determinants of the likelihood of regime change. In other words, mass political movements are more likely to succeed when society is riper for change. But Geddes (1990) and Goldstone (2001) observe that in many societies which are plagued by structural roots of instability, upheavals with massive attacks do not occur, and therefore call into question the hypothesis.[27] But our analysis in this section shows that the probability of success may increase or decrease, when the external environment is more favorable for revolt or riper for change (i.e., $\hat{G}^H$ can increase or decrease in π_H). Pooling these cases together without heeding the subtlety of the leadership's role, an econometrician might conclude that the factors that can trigger political instability are observed more often than actual political upheavals themselves. According to our theory, it may be less puzzling than it appears.

Another line of critique of the structural roots hypothesis is that leaders' characteristics matter for the outcome of political movements (Goldstone 2001). Indeed, our work provides an example where leaders' characteristics (i.e., their preferences for reform) matter. But we further demonstrate that structural factors may help determine who ends up playing the leadership role. Discussion in this section particularly highlights that both the punishment structure (i.e., c_l/k_l and c_f/k_f) and the uncertain environment (i.e., π_H) affect the leaders' characteristics in equilibrium.

6. Conclusion

It is often said that desperate times call for desperate measures. The flip side to this is that desperate measures are a sign of desperate times. In this paper we study how political leaders choose radical reform agendas as a signal to inform citizens that society badly needs to change. Our model goes beyond this basic insight, however. Specifically, moderate leaders are less inclined to radical ideals, which makes it easier for them to credibly convey their private information to the public. Radical leaders are known to prefer radical ideals; therefore their claims and proposals are difficult to be credible even when their solutions are suitable for the situation. That helps explain why sometimes opposition leaders have to pursue very radical agendas even though they are aware that they will lose support.

[27] For example, Geddes (1990) shows that external threats, one of the structural factors identified by Skocpol (1979) may not necessarily lead to revolutions in Latin American countries. In a review article, Goldstone (2001) reports that the literature has not produced a consensus concerning the degree to which inequality leads to social unrest. Bueno de Mesquita (2010) argues that such an empirical critique may not be fatal to the notion that structural factors are important for regime changes and that when multiple equilibria are present, the impact of structural factors may be obscured by historical and cultural factors that determine equilibrium selection.

Although the focus of this paper is on radical reform agendas rather than radical tactics in social movements, with suitable modifications a similar logic can be applied to explain the latter as well. Further, to the best of our knowledge, our model is the first attempt to analyze explicitly the signaling role of radicalism and endogenous leadership in the context of collective actions.

Appendix

Proof of Lemma 1. For $y_1 \in [y_0, 1+\theta]$, define

$$\tilde{F}(y_1) := FP(y_1-\theta; 1-(y_1-\theta), y_1) = k_f G(1-(y_1-\theta))(y_1-y_0)-c_f.$$

Log-concavity of G implies that $\tilde{F}$ is quasi-concave. Since $\tilde{F}(y_0) < 0$, and the assumption that $G(0) > c_f/k_f$ implies $\tilde{F}(1+\theta) > 0$, the function $\tilde{F}$ must be single-crossing from below. Let $y_{\min} \in (y_0, 1+\theta)$ represent such a crossing point. Because $\tilde{F}$ must be increasing at the crossing point, we have $\tilde{F}'(y_{\min}) > 0$. Since $\tilde{F}$ is quasi-concave, $\tilde{F}'(y_{\min}) > 0$ also implies $\tilde{F}'(y_1) > 0$ for all $y_1 < y_{\min}$.

For $y_1 \in [y_0, 1+\theta]$ and $x^f \in [0, y_1-\theta]$, define

$$F(x^f; y_1) := FP(x^f; 1-x^f, y_1) = k_f G(1-x^f)(2x^f+2\theta-y_1-y_0)-c_f.$$

Log-concavity of G again implies that $F(\cdot; y_1)$ is quasi-concave. Moreover, it is easy to verify that $\tilde{F}'(y_1) > 0$ implies $\partial F(x^f; y_1)/\partial x^f > 0$ when evaluated at $x^f = y_1-\theta$.

(a) Suppose $y_1 < y_{\min}$. Since $F(0; y_1) < 0$ and $F(y_1-\theta; y_1) = \tilde{F}(y_1) < 0$, and since $\partial F(y_1-\theta; y_1)/\partial x^f > 0$, the quasi-concavity of $F(\cdot; y_1)$ implies that $F(x^f; y_1) < 0$ for any $x^f \in [0, y_1-\theta]$. For $x^f > y_1-\theta$, we have

$$FP(x^f; 1-x^f, y_1) = k_f G(1-x^f)(y_1-y_0)-c_f < \tilde{F}(y_1) < 0.$$

Because $FP(x^f; 1-x^f, y_1) < 0$ for all x^f, the only equilibrium is one in which no one attacks.

(b) Suppose $y_1 \geq y_{\min}$. Then we have $F(0; y_1) < 0$ and $F(y_1-\theta; y_1) = \tilde{F}(y_1) \geq 0$. Moreover, since $F(\cdot; y_1)$ is quasi-concave in the relevant domain, it follows that there exists a unique $x^f \in (0, y_1-\theta]$ which satisfies the equilibrium condition $F(x^f; y_1) = 0$.

For the last part of the proposition, we let $\lambda(\cdot) := G'(\cdot)/G(\cdot)$, and prove the following result.

Claim 1. For $y_1 \geq y_{\min}(\theta)$,

$$\frac{\partial x^f}{\partial y_1} = \frac{k_f G(1-x^f)}{2k_f G(1-x^f) - \lambda(1-x^f)c_f} > \frac{1}{2},$$
$$\frac{\partial x^f}{\partial \theta} = -2\frac{\partial x^f}{\partial y_1} < -1.$$

Proof. Use implicit differentiation of the relation $F(x^f(y_1;\theta); y_1) = 0$ to obtain:

$$\frac{\partial x^f}{\partial y_1} = \frac{1}{2 - \lambda(1-x^f)(2x^f + 2\theta - y_1 - y_0)}.$$

Multiplying both the denominator and the numerator by $k_f G(1-x^f)$ and applying the equilibrium condition, we get the expression given in this claim. Note that the denominator has the same sign as $\partial F(x^f(y_1); y_1)/\partial x^f$, which is positive because $F(\cdot; y_1)$ is single-crossing from below for $x^f \in [0, y_1 - \theta]$. Moreover, since both $\lambda(\cdot)$ and c_f are strictly positive, we have $\partial x^f/\partial y_1 > 1/2$. A similar exercise gives $\partial x^f/\partial \theta = -2\partial x^f/\partial y_1 < -1$. □

Since $A = 1 - x^f(y_1;\theta)$, the comparative statics for A has opposite signs to those for x^f. The second part of the proposition then follows immediately from Claim 1. ■

Proof of Proposition 1. Define

$$I(y_1;\theta) := -\frac{\partial}{\partial y_1}\log G(1-x^f(y_1;\theta)) = \lambda(1-x^f(y_1;\theta))\frac{\partial x^f(y_1;\theta)}{\partial y_1}.$$

Claim 2. For $y_1 > y_{\min}$, $I(y_1;\theta)$ is increasing in y_1 and decreasing in θ. Moreover, $I(y_{\min};\theta) < \lambda(1-(y_{\min}-\theta))$.

Proof. From Claim 1, we see that $\partial x^f/\partial y_1$ depends on y_1 only through x^f. Since $\partial x^f/\partial y_1$ increases in x^f, and x^f increases in y_1, $\partial x^f/\partial y_1$ is increasing in y_1. Moreover, log-concavity of G implies that $\lambda(1-x^f(y_1;\theta))$ is increasing in y_1. It then follows that $I(y_1;\theta)$ increases in y_1. Similarly, $I(y_1;\theta)$ depends on θ only through x^f. By Claim 1, we have $\partial I/\partial\theta = -2\partial I/\partial y_1 < 0$. For the last part of this claim, it suffices to show that

$\partial x^f(y_{\min}(\theta);\theta)/\partial y_1 < 1$. Recall that, in the proof of Lemma 1, $\tilde{F}(y_{\min}) = 0$ means that the marginal attacker corresponding to $y_1 = y_{\min}$ is $x^f = y_{\min} - \theta$. Using the expression for $\partial x^f/\partial y_1$ provided therein, we obtain:

$$\frac{\partial x^f(y_{\min}(\theta);\theta)}{\partial y_1} = \frac{1}{2 - \lambda(1-(y_{\min}-\theta))(y_{\min}-y_0)}.$$

But $\tilde{F}'(y_{\min}) > 0$ implies $1 - \lambda(1-(y_{\min}-\theta)(y_{\min}-y_0) > 0$, and the claim follows. □

Since $I(y_1;\theta)$ increases in y_1 by Claim 2, the maximization problem (3) is quasi-concave for $y_1 \in [y_{\min}, x^m + \theta]$. Furthermore, Claim 2 implies that

$$1 - I(y_{\min};\theta)(y_{\min}-y_0) > 1 - \lambda(1-(y_{\min}-\theta))(y_{\min}-y_0) > 0.$$

This contradicts the necessary condition for the corner solution $y_1 = y_{\min}$ to be optimal. The first-order condition for optimality therefore gives either an interior solution or a corner solution at $y_1 = x^m + \theta$. For the corner solution $y_1^* = x^m + \theta$ to be optimal, we require:

$$1 - I(x^m+\theta;\theta)(x^m+\theta-y_0) \geq 0.$$

The left-hand-side of the above is strictly positive at $x^m = y_{\min} - \theta$, and is strictly decreasing in x^m. There exists a unique $\hat{x}(\theta) > y_{\min} - \theta$ such that the left-hand-side is equal to 0 when $x^m = \hat{x}(\theta)$. When $x^m > \hat{x}(\theta)$, $y_1^* = \hat{x}(\theta) + \theta$ satisfies the first-order condition for an interior solution.

At an interior solution, we have

$$1 - I(\hat{x}(\theta)+\theta;\theta)(\hat{x}(\theta)+\theta-y_0) = 0.$$

Differentiate the above equation to get $\partial \hat{x}(\theta)/\partial\theta \in (-1,1)$. Since $y_1^* = \min\{x^m, \hat{x}(\theta)\} + \theta$, we have $\partial y_1^*/\partial\theta \in (0,2)$. Moreover, $\partial y_1^*/\partial\theta < 2$ and Claim 1 together imply that $x^f(y_1^*(x^m;\theta);\theta)$ decreases in θ. It follows that $G(1 - x^f(y_1^*(x^m;\theta);\theta))$ increases in θ. ■

Proof of Lemma 2. For $x^m \geq y_{\min}(\theta_L) - \theta_L$, let the difference in payoff between choosing

y_1^{H*} and y_1^{L*} in state θ_L be represented by

$$\tilde{\Delta}_L(x^m) := G^{L*}\left(u(y_1^{L*},\theta_L,x^m) - u(y_0,\theta_L,x^m)\right) - G^{H*}\left(u(y_1^{H*},\theta_L,x^m) - u(y_0,\theta_L,x^m)\right),$$

where y_1^{H*} and y_1^{L*} depend on x^m. We show that $\tilde{\Delta}_L(x^m)$ decreases in x^m. There are four cases to consider:

(a) If $x^m \le \min\{\hat{x}(\theta_H),\hat{x}(\theta_L)\}$, then $y_1^{H*} = x^m + \theta_H$ and $y_1^{L*} = x^m + \theta_L$. This gives:

$$\tilde{\Delta}_L(x^m) = G^{L*}(x^m + \theta_L - y_0) - G^{H*}(2x^m + 2\theta_L - (x^m + \theta_H) - y_0).$$

The derivative with respect to x^m is

$$\tilde{\Delta}_L'(x^m) = \left[G^{L*}\left(1 - I^{L*}(x^m + \theta_L - y_0)\right) - G^{H*}\left(1 - I^{H*}(x^m + \theta_L - y_0)\right)\right] - G^{H*}I^{H*}(\theta_H - \theta_L),$$

where $I^{H*} := I(y_1^{H*};\theta_H)$ and $I^{L*} := I(y_1^{L*};\theta_L)$. The term in brackets is negative, because $G^{L*} < G^{H*}$ (Proposition 1) and $I^{L*} > I^{H*}$ (which follows from Claim 1 and the fact that I depends on y_1 and θ only through $x^f(y_1;\theta)$). We therefore have $\tilde{\Delta}_L'(x^m) < 0$.

(b) If $x^m \in [\min\{\hat{x}(\theta_H),\hat{x}(\theta_L)\}, \max\{\hat{x}(\theta_H),\hat{x}(\theta_L)\}]$, then $y_1^{H*} = \min\{x^m,\hat{x}(\theta_H)\} + \theta_H$ and $y_1^{L*} = \min\{x^m,\hat{x}(\theta_L)\} + \theta_L$. This gives:

$$\tilde{\Delta}_L(x^m) = G^{L*}(\min\{x^m,\hat{x}(\theta_L)\} + \theta_L - y_0) – G^{H*}(2x^m + 2\theta_L - (\min\{x^m,\hat{x}(\theta_H)\} + \theta_H) - y_0).$$

Therefore,

$$\tilde{\Delta}_L'(x^m) = \begin{cases} G^{L*}\left(1 - I^{L*}(x^m + \theta_L - y_0)\right) - 2G^{H*} & \text{if } \hat{x}(\theta_H) < \hat{x}(\theta_L), \\ -G^{H*}\left(1 - I^{H*}(x^m + \theta_L - y_0) + I^{H*}(\theta_H - \theta_L)\right) & \text{if } \hat{x}(\theta_H) > \hat{x}(\theta_L). \end{cases}$$

In either case, we have $\tilde{\Delta}_L'(x^m) < 0$.

(c) If $x^m \in (\max\{\hat{x}(\theta_H),\hat{x}(\theta_L)\}, \hat{x}(\theta_H) + \theta_H - \theta_L)$, then $y_1^{H*} = \hat{x}(\theta_H) + \theta_H$ and $y_1^{L*} =$

$\hat{x}(\theta_L) + \theta_L$. This gives:

$$\tilde{\Delta}_L(x^m) = G^{L*}\left(\hat{x}(\theta_L) + \theta_L - y_0\right) - G^{H*}\left(2x^m + 2\theta_L - (\hat{x}(\theta_H) + \theta_H) - y_0\right)$$

with $\tilde{\Delta}_L'(x^m) = -2G^{H*} < 0$.

(d) If $x^m \geq \hat{x}(\theta_H) + \theta_H - \theta_L$, then $y_1^{H*} = \hat{x}(\theta_H) + \theta_H$, $y_1^{L*} = \hat{x}(\theta_L) + \theta_L$, and

$$\tilde{\Delta}_L(x^m) = G^{L*}\left(\hat{x}(\theta_L) + \theta_L - y_0\right) - G^{H*}\left(\hat{x}(\theta_H) + \theta_H - y_0\right) < 0.$$

In this case, $\tilde{\Delta}_L(x^m)$ is constant with respect to x^m.

Because $\tilde{\Delta}_L(x^m)$ is continuous in x^m, these four cases show that it is decreasing in x^m for all $x^m \geq y_{\min}(\theta_L) - \theta_L$. Moreover,

$$\begin{aligned}
\tilde{\Delta}_L(\hat{x}(\theta_H)) &= G^{H*}\left(y_1^{L*} - \hat{x}(\theta_H) - 2\theta_L + \theta_H\right) + (G^{L*} - G^{H*})\left(y_1^{L*} - y_0\right) \\
&> G^{H*}\left(y_1^{L*} - \hat{x}(\theta_H) - 2\theta_L + \theta_H + (\log G^{L*} - \log G^{H*})\left(y_1^{L*} - y_0\right)\right),
\end{aligned}$$

where the inequality follows from the fact that $t - 1 > \log t$ for any positive t. Now,

$$\begin{aligned}
\log G^{L*} - \log G^{H*} &= \log G(1 - x^f(y_1^{L*}; \theta_L)) - \log G(1 - x^f(y_1^{H*} - 2(\theta_H - \theta_L); \theta_L)) \\
&= -\int_{y_1^{H*} - 2(\theta_H - \theta_L)}^{y_1^{L*}} I(y_1; \theta_L))\,\mathrm{d}y_1 \\
&> -I^{L*}\left(y_1^{L*} - \hat{x}(\theta_H) - 2\theta_L + \theta_H\right),
\end{aligned}$$

where the first equality follows from Claim 1 and the inequality follows from Claim 2. This gives:

$$\tilde{\Delta}_L(\hat{x}(\theta_H)) > G^{H*}\left(y_1^{L*} - \hat{x}(\theta_H) - 2\theta_L + \theta_H\right)\left(1 - I^{L*}(y_1^{L*} - y_0)\right) \geq 0,$$

where the last inequality follows from the first-order condition (4). In part (d) above, we have also shown that $\tilde{\Delta}_L(\hat{x}(\theta_H) + \theta_H - \theta_L) < 0$. Thus, there exists a unique $x^\dagger \in (\hat{x}(\theta_H), \hat{x}(\theta_H) + \theta_H - \theta_L)$ such that $\tilde{\Delta}_L(x^m) \geq 0$ if $x^m \leq x^\dagger$ and $\tilde{\Delta}_L(x^m) < 0$ if $x^m > x^\dagger$.

For $x^m \leq x^\dagger$, the leader prefers choosing y_1^{L*} to choosing y_1^{H*} in state θ_L because $\tilde{\Delta}_L(x^m) \geq 0$. In state θ_H, the leader prefers choosing y_1^{H*} to choosing y_1^{L*} because the former is closer to the leader's ideal policy ($x^m + \theta_H \geq y_1^{H*} > y_1^{L*}$) and because the probability of success is higher ($G^{H*} > G^{L*}$). Given such choices, citizens correctly infer the state based on the agenda chosen by the leader, and therefore the full-information outcome constitutes an equilibrium outcome.

For $x^m > x^\dagger$, the full-information outcome cannot be supported in equilibrium because $\tilde{\Delta}_L(x^m) < 0$ implies that the leader would deviate to choosing y_1^{H*} in state θ_L. ■

Proof of Proposition 2. *Part (i).* Suppose $\hat{y}_1^L \neq y_1^{L*}$. If x^m deviates to y_1^{L*}, the worst inference that citizens can make is that the state is low, which is the same as the equilibrium inference. But by construction, y_1^{L*} is preferred to $\hat{y}_1^L$, a contradiction.

Part (ii). In any separating equilibrium, the policy y_1^H chosen in the high state has to satisfy the incentive constraint for the leader in the low state:

$$\begin{aligned}\Delta_L(x^m; y_1^H) = {} & G(1 - x^f(\hat{y}_1^L; \theta_L))\big(u(\hat{y}_1^L, \theta_L, x^m) - u(y_0, \theta_L, x^m)\big) \\ & - G(1 - x^f(y_1^H; \theta_H))\big(u(y_1^H, \theta_L, x^m) - u(y_0, \theta_L, x^m)\big) \geq 0,\end{aligned}$$

where $\hat{y}_1^L = y_1^{L*}$ depends on x^m. For $x^m > x^\dagger$, Lemma 2 establishes that $\Delta_L(x^m; y_1^{H*}) = \tilde{\Delta}_L(x^m) < 0$. There are three cases.

(a) If $y_1^{H*} > x^m + \theta_L$, then for any $y_1^H \in (x^m + \theta_L, y_1^{H^*})$, $\Delta_L(x^m; \cdot)$ is strictly increasing, implying that $\Delta_L(x^m; y_1^H) < \Delta_L(x^m; y_1^{H*}) < 0$ for such y_1^H.

(b) For any $y_1^H \in [\hat{y}_1^L, \min\{y_1^{H*}, x^m + \theta_L\}]$, $\Delta_L(x^m; \cdot)$ is strictly decreasing. Therefore, $\Delta_L(x^m; y_1^H) < \Delta_L(x^m; \hat{y}_1^L) < 0$.

(c) For $y_1^H < \hat{y}_1^L$, $\Delta_L(x^m; \cdot)$ is strictly decreasing. There exists $\tilde{y} \in [y_{\min}(\theta_H), \hat{y}_1^L)$ such that $\Delta_L(x^m; y_1^H) < 0$ for all $y_1^H \in (\tilde{y}, \hat{y}_1^L)$. Such y_1^H cannot be part of a separating equilibrium because it violates the incentive constraint in the low state. For $y_1^H < \tilde{y}$, we have $\Delta_L(x^m; y_1^H) > 0$. But because the reward from success weakly increases in the state, this implies

$$\Delta_H(x^m; y_1^H) := G(1 - x^f(y_1^H; \theta_H))\big(u(y_1^H, \theta_H, x^m) - u(y_0, \theta_H, x^m)\big) \\ - G(1 - x^f(\hat{y}_1^L; \theta_L))\big(u(\hat{y}_1^L, \theta_H, x^m) - u(y_0, \theta_H, x^m)\big) < 0,$$

which violates the incentive constraint in the high state. If $y_1^H = \tilde{y}$, it is possible to have a knife-edge equilibrium in which the leader chooses $\hat{y}_1^L$ in the low state and chooses $\tilde{y} < \hat{y}_1^L$ in the high state, and the incentive constraints in both states are satisfied with equality. However, such a knife-edge equilibrium does not satisfy the D1 refinement. Consider a deviation to some $y' \in (x^m + \theta_L, x^m + \theta_H]$. For any given belief, the high type gains more from such deviation than does the low type. According to the D1 criterion, citizens should assign off-equilibrium belief that such deviation comes from the high type. Given such off-equilibrium belief, the high type indeed could profitably deviate from $\tilde{y}$ to y', which means that the knife-edge case does not satisfy the D1 equilibrium refinement.

We conclude that any separating equilibrium that satisfies the D1 refinement must satisfy $\hat{y}_1^H > y_1^{H*}$. Because $\Delta_L(x^m; y_1^H)$ is strictly increasing for $y_1^H > y_1^{H*}$ and is positive when y_1^H is sufficiently large, there exists a unique $\hat{y}_1^H > y_1^{H*}$ that satisfies $\Delta_L(x^m; \hat{y}_1^H) = 0$. The pair $(\hat{y}_1^L, \hat{y}_1^H)$ constitute a separating equilibrium because the leader weakly prefers $\hat{y}_1^L$ to $\hat{y}_1^H$ in the low state. Moreover, because $\Delta_L(x^m; \hat{y}_1^H) = 0$ implies $\Delta_H(x^m; \hat{y}_1^H) > 0$, the leader strictly prefers $\hat{y}_1^H$ to $\hat{y}_1^L$ in the high state. This equilibrium can be supported by off-equilibrium beliefs which assign probability 1 that the state is high when the policy $y_1 \geq \hat{y}_1^H$ and probability 0 otherwise.

To show that these beliefs satisfy the D1 criterion, we need the following result.

Claim 3. For any $x^m \geq x^\dagger$, $\hat{y}_1^H(x^m) \in (x^m + \theta_L, x^m + \theta_H)$.

Proof. Let $x_L^f := x^f(\hat{y}_1^L; \theta_L)$ and $x_H^f := x^f(x^m + \theta_H; \theta_H) = x^f(x^m + 2\theta_L - \theta_H; \theta_L)$. By

Claim 1, we see that $x^f_L \geq x^f_H$ (and $G^L \leq G^H$) if and only if $\hat{y}^L_1 \geq x^m + 2\theta_L - \theta_H$. We have

$$\begin{aligned}
\Delta_L(x^m; x^m + \theta_H) &= G^H \left(\hat{y}^L_1 - x^m - 2\theta_L + \theta_H\right) + (G^L - G^H)(\hat{y}^L_1 - y_0) \\
&> G^H \left(\hat{y}^L_1 - x^m - 2\theta_L + \theta_H + (\log G^L - \log G^H)(\hat{y}^L_1 - y_0)\right) \\
&> G^H \left(\hat{y}^L_1 - x^m - 2\theta_L + \theta_H + I^L(x^m + 2\theta_L - \theta_H - \hat{y}^L_1)(\hat{y}^L_1 - y_0)\right) \\
&= G^H \left(\hat{y}^L_1 - x^m - 2\theta_L + \theta_H\right)\left(1 - I^L(\hat{y}^L_1 - y_0)\right),
\end{aligned}$$

where $I^L := I(\hat{y}^L_1; \theta_L)$. If $x^m < \hat{x}(\theta_L)$, $\hat{y}^L_1 - x^m - 2\theta_L + \theta_H > 0$ and $1 - I^L(\hat{y}^L_1 - y_0) \geq 0$. If $x^m \geq \hat{x}(\theta_L)$, $1 - I^L(\hat{y}^L_1 - y_0) = 0$. In either case, $\Delta_L(x^m; x^m + \theta_H) > 0$.

Now, we let x^f_H, G^H, and I^H represent the values of the relevant variables at the point $y^H_1 = x^m + \theta_L$. We have

$$\begin{aligned}
\Delta_L(x^m; x^m + \theta_L) &= G^L \left(\hat{y}^L_1 - x^m - \theta_L\right) - (G^H - G^L)(x^m + \theta_L - y_0) \\
&< G^L \left(\hat{y}^L_1 - x^m - \theta_L + (\log G^L - \log G^H)(x^m + \theta_L - y_0)\right).
\end{aligned}$$

If $\hat{y}^L_1 \in [x^m + \theta_L - 2(\theta_H - \theta_L), x^m + \theta_L]$, then $G^L \leq G^H$, and therefore the above expression is negative. If $\hat{y}^L_1 < x^m + \theta_L - 2(\theta_H - \theta_L)$, then $G^L > G^H$, and we have

$$\begin{aligned}
\Delta_L(x^m; x^m + \theta_L) &< G^L \left(\hat{y}^L_1 - x^m - \theta_L + I^H(x^m + \theta_L - 2(\theta_H - \theta_L) - \hat{y}^L_1)(x^m + \theta_L - y_0)\right) \\
&< -G^H(\theta_H - \theta_L),
\end{aligned}$$

where the second inequality follows because $x^m > x^\dagger > \hat{x}(\theta_H)$ implies that $I^H(x^m + \theta_L - y_0) < 1$. We conclude that $\Delta_L(x^m; x^m + \theta_L) < 0$.

Since $\Delta_L(x^m; x^m + \theta_H) > 0 > \Delta_L(x^m; x^m + \theta_L)$, and $\Delta_L(x^m; \cdot)$ is strictly increasing in the relevant range, the $\hat{y}^H_1$ that satisfies the binding incentive constraint $\Delta_L(x^m; \hat{y}^H_1) = 0$ is unique and satisfies $\hat{y}^H_1 \in (x^m + \theta_L, x^m + \theta_H)$. □

Consider an off-equilibrium policy $y' \in [x^m + \theta_L, x^m + \theta_H]$. The minimum success probability G' that would induce the high type to deviate to y' requires

$$G' > \frac{\hat{G}^H(\hat{y}_1^H - y_0)}{y' - y_0};$$

and the minimum G' needed to induce the low type to deviate requires

$$G' > \frac{\hat{G}^L(\hat{y}_1^L - y_0)}{2x^m + 2\theta_L - y' - y_0}.$$

Using the binding incentive constraint (6), the set of beliefs that would support deviation by the high type strictly contains the set of beliefs that would support deviation by the low type if and only if

$$\frac{2x^m + 2\theta_L - \hat{y}_1^H - y_0}{\hat{y}_1^H - y_0} > \frac{2x^m + 2\theta_L - y' - y_0}{y' - y_0},$$

which is true if and only if $y' > \hat{y}_1^H$. Thus the D1 criterion requires assigning probability 1 that the state is high if $y' > \hat{y}_1^H$, and probability 0 if $y' < \hat{y}_1^H$.

For deviations $y' < x^m + \theta_L$, the corresponding comparison requires

$$\hat{G}^L(\hat{y}_1^L - y_0) > \hat{G}^H(\hat{y}_1^H - y_0),$$

which contradicts the binding incentive constraint (6). Therefore citizens assign probability 0 that the state is high upon observing such an agenda. For $y' > x^m + \theta_H$, the comparison requires

$$\frac{2x^m + 2\theta_L - \hat{y}_1^H - y_0}{\hat{y}_1^H - y_0} > \frac{2x^m + 2\theta_L - y' - y_0}{2x^m + 2\theta_H - y' - y_0},$$

which is always true. In this case, the off-equilibrium belief that the state is high with probability 1 is again consistent with the D1 criterion.

Part (iii). Use implicit differentiation of $\Delta_L(x^m; \hat{y}_1^H(x^m)) = 0$ to get:

$$\frac{\partial \hat{y}_1^H}{\partial x^m} = \begin{cases} \frac{2\hat{G}^H - \hat{G}^L\left(1 - I^L(x^m + \theta_L - y_0)\right)}{\hat{G}^H\left(1 + I^H(2x^m + 2\theta_L - \hat{y}_1^H - y_0)\right)} & \text{if } x^m < \hat{x}(\theta_L), \\ \frac{2}{1 + I^H(2x^m + 2\theta_L - \hat{y}_1^H - y_0)} & \text{if } x^m > \hat{x}(\theta_L). \end{cases}$$

By Lemma 2, $x^m < \hat{x}(\theta_L)$ implies $\hat{G}^H > \hat{G}^L$. Therefore, $\partial \hat{y}_1^H / \partial x^m > 0$. For comparative statics respect to θ_H, we have

$$\frac{\partial \hat{y}_1^H}{\partial \theta_H} = \frac{2I^H(2x^m + 2\theta_L - \hat{y}_1^H - y_0)}{1 + I^H(2x^m + 2\theta_L - \hat{y}_1^H - y_0)} > 0. \qquad \blacksquare$$

Proof of Proposition 3. Two conditions are sufficient for this proposition: $y_{\min}(\theta_L) - \theta_L < 1/2$ and $\min\{\hat{LP}(1/2; 1/2), \hat{LP}(1; 1)\} > 0$. We first establish the following result.

Claim 4. $\min\{\hat{LP}(1/2; 1/2), \hat{LP}(1; 1)\} > 0$ implies $\hat{LP}(x^m; x^m) > 0$ for all $x^m \in [1/2, 1]$.

Proof. Consider the leader's premium in the low state. We have

$$\frac{\mathrm{d}}{\mathrm{d}x^m} LP(x^m; \hat{y}_1^L(x^m), \theta_L) = \begin{cases} k_l \hat{G}^L \left(1 - I^L(x^m + \theta_L - y_0)\right) & \text{if } x^m < \hat{x}(\theta_L), \\ 0 & \text{if } x^m > \hat{x}(\theta_L). \end{cases}$$

This shows that $LP(x^m; \hat{y}_1^L(x^m), \theta_L)$ is increasing for $x^m < \hat{x}(\theta_L)$ and is constant for $x^m > \hat{x}(\theta_L)$. In the high state,

$$\frac{\mathrm{d}}{\mathrm{d}x^m} LP(x^m; \hat{y}_1^H(x^m), \theta_H) = k_l \hat{G}^H \left(1 - I^H(\hat{y}_1^H - y_0)\right) \frac{\partial \hat{y}_1^H}{\partial x^m}.$$

If $x^m > x^\dagger$, $I^H(\hat{y}_1^H - y_0) > I^H(y_1^{H*} - y_0) > 1$. Therefore $LP(x^m; \hat{y}_1^H(x^m), \theta_H)$ is increasing for $x^m < \hat{x}(\theta_H)$, constant for $x^m \in (\hat{x}(\theta_H), x^\dagger)$, and decreasing for $x^m > x^\dagger$. Because $\hat{LP}(x^m; x^m)$ is a weighted average of the leader's premium in the two states, its minimum on the interval $[1/2, 1]$ must be at either corner of the interval. $\square$

Claim 4 implies that the mapping M is well-defined on $[1/2, 1]$. It is continuous because $x^m \geq 1/2$ implies that $x^m \geq 1/2 > y_{\min}(\theta_L) - \theta_L$. Moreover, $x^l(1/2) > 0$ implies $M(1/2) > 1/2$ and $x^l(1) < 1$ implies $M(1) < 1$. Thus a fixed point of the the mapping M exists.

We next show that $M'(x^m) > 0$. Note that Claim 2 implies that $LP(x^l; y_1, \theta_L)$ is quasi-concave in y_1, and is strictly decreasing in y_1 for all $y_1 > y_1^*(x^l; \theta_L)$. Therefore, $\hat{y}_1^L =$

$y_1^*(x^m;\theta_L) \geq y_1^*(x^l;\theta_L)$ implies that

$$\frac{\partial LP^L}{\partial x^m} = \frac{\partial LP(x^l;\hat{y}_1^L,\theta_H)}{\partial y_1}\frac{\partial \hat{y}_1^L(x^m)}{\partial x^m} \leq 0,$$

with equality if only if $x^m > \hat{x}(\theta_L)$. Similarly, $LP(x^l;y_1,\theta_H)$ is strictly decreasing in y_1 for all $y_1 > y_1^*(x^l;\theta_H)$. Therefore, $\hat{y}_1^H \geq y_1^*(x^m;\theta_H) \geq y_1^*(x^l;\theta_H)$ implies that $\partial LP^H/\partial x^m \leq 0$, with equality if and only if $x^m \in (\hat{x}(\theta_H), x^\dagger)$. Since both LP^L and LP^H decreases in x^m, a higher x^m lowers the expected leader's premium $\hat{LP}(x_i;x^m)$ for every citizen i, and therefore raises $x^l(x^m)$. This implies that $M(x^m)$ also becomes higher.

Claim 5. If $\theta_H - \theta_L < 1/3$, then $M'(x_*^m) < 1$.

Proof. For $j \in \{H,L\}$, let $LP^j = LP(x^l;\hat{y}_1^j(x^m),\theta_j)$. We have

$$\frac{\partial LP^L}{\partial x^m} + 2\frac{\partial LP^L}{\partial x^l} = \begin{cases} k_l\hat{G}^L\left[3 - I^L(2x^l + 2\theta_L - (x^m+\theta_L) - y_0)\right] & \text{if } x^m < \hat{x}(\theta_L), \\ 2k_l\hat{G}^L & \text{if } x^m > \hat{x}(\theta_L) > x^l, \\ 0 & \text{if } x^l > \hat{x}(\theta_L). \end{cases}$$

Because $x^l < x^m$, the bracketed term for the first case is greater than $3 - I^L(x^m + \theta_L - y_0)$, which is positive by the first-order condition (4).

Furthermore, we have

$$\frac{\partial LP^H}{\partial x^m} + 2\frac{\partial LP^H}{\partial x^l} = k_l\hat{G}^H\left[4 - \left(1 + I^H(2x^l + 2\theta_H - \hat{y}_1^H - y_0)\right)\frac{\partial \hat{y}_1^H}{\partial x^m}\right].$$

From the proof of Proposition 2 (i.e., part (iii)), we see that the bracketed term is greater than or equal to

$$4 - 2\frac{1 + I^H(2x^l + 2\theta_H - \hat{y}_1^H - y_0)}{1 + I^H(2x^m + 2\theta_L - \hat{y}_1^H - y_0)}.$$

A sufficient condition for the above expression to be positive when evaluated at the fixed point $x^m = x_*^m$ and $x^l = x^l(x_*^m) = 2x_*^m - 1$ is that

$$\frac{2x^l(x_*^m) + 2\theta_H - \hat{y}_{1*}^H - y_0}{2x_*^m + 2\theta_L - \hat{y}_{1*}^H - y_0} \leq 2.$$

The above inequality is true if and only if

$$[1-3(\theta_H-\theta_L)]+\left[x_*^m+\theta_H-\hat{y}_{1*}^H\right]+\left[x_*^m-x^l(x_*^m)\right]+[\theta_L-y_0]\geq 0.$$

By Claim 3, the second bracketed term is positive. Therefore, $\theta_H-\theta_L\leq 1/3$ implies that the inequality is true.

We have shown that for $j\in\{H,L\}$,

$$\sum_{j=H,L}\pi_j\left(\frac{\partial LP^j}{\partial x^m}+2\frac{\partial LP^j}{\partial x^l}\right)>0$$

when evaluated at $(x^m,x^l)=(x_*^m,x_*^l)$. By the implicit function theorem, this implies that $\partial x^l(x_*^m)/\partial x^m<2$, and hence $M'(x_*^m)<1$. □

Claim 5 implies that $M(x^m)-x^m$ is single-crossing from above when $\theta_H-\theta_L\leq 1/3$. This shows that the fixed point x_*^m is unique. ■

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2013). "A Political Theory of Populism." *Quarterly Journal of Economics 128*(2), 771–805.

Banks, Jeffrey (1990). "Monopoly Agenda Control and Asymmetric Information." *Quarterly Journal of Economics 105*(2), 445–464.

Banks, Jeffrey and Joel Sobel (1987). "Equilibrium Selection in Signaling Games." *Econometrica 55*(3), 647–661.

Bils, Peter (2019). "Overreacting and Posturing: How Accountability and Ideology Shape Executive Policies." Working Paper, Princeton University.

Boleslavsky, Raphael and Christopher Cotton (2015). "Information and Extremism in Elections." *American Economic Journal: Microeconomics 7*(1), 165–207.

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

Cho, In-Koo and David M. Kreps (1987). "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics 102*(2), 179–221.

Dapiran, Antony (2017). *City of Protest: A Recent History of Dissent in Hong Kong*. Penguin Group Australia.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao (2017). "Voting to Tell Others." *Review of Economic Studies 84*(1), 143–181.

Dewan, Torun and David P. Myatt (2007). "Leading the Party: Coordination, Direction, and Communication." *American Political Science Review 101*(4), 827–845.

Dugan, Laura and Erica Chenoweth (2012). "Moving Beyond Deterrence: The Effectiveness of Raising the Expected Utility of Abstaining from Terrorism in Israel." *American Sociological Review 77*(4), 597–624.

Fu, Qiang, Ming Li, and Xue Qiao (2020). "On the Paradox of Mediocracy." Working Paper, Concordia University.

Geddes, Barbara (1990). "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis 2*, 131–150.

Goldstone, Jack Andrew (2001). "Toward a Fourth Generation of Revolutionary Theory." *Annual Review of Political Science 4*(1), 139–187.

Hermalin, Benjamin E. (1998). "Toward an Economic Theory of Leadership: Leading by Example." *American Economic Review 88*(5), 1188–1206.

Landa, Dimitri and Scott A. Tyson (2017). "Coercive Leadership." *American Journal of Political Science 61*(3), 559–574.

Lindbeck, Assar and Jörgen W. Weibull (1987). "Balanced-Budget Redistribution as the Outcome of Political Competition." *Public Choice 52*(3), 273–297.

Michaeli, Moti and Daniel Spiro (2018). "The Dynamics of Revolutions." Working Paper, University of Oslo.

Morris, Stephen and Mehdi Shadmehr (2020). "Inspiring Regime Change." Working Paper, Massachusetts Institute of Technology.

Ogden, Suzanne, Kathleen Hartford, Nancy Sullivan, and David Zweig (1992). *China's Search for Democracy: The Students and Mass Movement of 1989*. Armonk: M.E. Sharpe.

Rasler, Karen (1996). "Concessions, Repression, and Political Protest in the Iranian Revolution." *American Sociological Review* , 132–152.

Shadmehr, Mehdi (2015). "Extremism in Revolutionary Movements." *Games and Economic Behavior 94*, 97–121.

Shadmehr, Mehdi and Dan Bernhardt (2019). "Vanguards in Revolution." *Games and Economic Behavior 115*, 146–166.

Skocpol, Theda (1979). *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. New York: Cambridge University Press.

Sunada, Ichiro (1969). "The Thought and Behavior of Zengakuren: Trends in the Japanese Student Movement." *Asian Survey 9*(6), 457–474.

Tsurumi, Kazuko (1970). "Some Comments on the Japanese Student Movement in the Sixties." *Journal of Contemporary History 5*(1), 104–112.

Wittman, Donald (1983). "Candidate Motivation: A Synthesis of Alternative Theories." *American Political Science Review 77*(1), 142–157.

Wood, Elisabeth Jean (2003). *Insurgent Collective Action and Civil War in El Salvador*. Cambridge: Cambridge University Press.