Radicalism in Mass Movements:

Asymmetric Information and Agenda Escalation*

HENG CHEN

University of Hong Kong

WING SUEN

University of Hong Kong

September 6, 2019

Abstract. Asymmetric information and diverse preferences for reform create an agency problem between opposition leaders and followers. Dissatisfied citizens are unsure how bad the current situation is and can infer it from the leader's reform agenda. They understand that the leader has incentive to exaggerate and mislead them. Therefore, the leader has to radicalize the agenda as a way of signaling the necessity of change. Radicalism is costly as it reduces the chances of success, but is necessary for maintaining credibility. Radicalism also discourages moderate citizens from joining the leadership, thus further radicalizing the leadership group and their agenda.

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

JEL Classification. D74; D82

*Heng Chen: hengchen@hku.hk, tel 852-28578506. Wing Suen: wing.suen@hku.hk, tel 852-28578605.
We thank Navin Kartik, Balazs Szentes, Xianwen Shi, and seminar participants at the Asian Meeting of the Econometric Society and Political Selection Workshop (2019) 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).

1. Introduction

Leaders of mass movements provide a 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, so that they rally support from dissatisfied citizens to replace the status quo. Yet, the interests of leaders and their followers are seldom one and the same. For one thing, preferences of different individuals are naturally different: some prefer radical solutions while others prefer more moderate ones. For another, 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 proposals of the 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. China had promised in 2007 that the Chief Executive of Hong Kong in 2017 would be selected via "universal suffrage," without stating precisely what that meant. By about 2013 a number of different electoral reforms were proposed to determine the process by which the Chief Executive would be selected. The more moderate proposals would attempt to squeeze the greatest degree of democracy within the strictures of the Basic Law, Hong Kong's mini-constitution. But leaders of pro-democratic 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 an Occupy Central movement and proposed to use civil disobedience to signal their resolve to achieve a full-fledged "genuine democracy." While the majority of Hong Kong citizens endorse the values of modern democracy, "many people consider that

Occupy Central is too radical a move to strive for true democracy," as acknowledged by one of the leaders, Benny Tai.[1] Occupy Central as originally envisioned by its leaders was supposed to be no more than an unauthorized protest that would block the streets over a weekend. But other events (class boycott by students and the use of tear gas by police against protesters) accelerated the onset of the movement. Student leaders took over the center stage of the movement, and escalated the event to paralyze traffic in key parts of the city for 77 days. In the end, the movement eventually ran out of steam without any political achievement, and the undemocratic electoral system remains largely intact.

Even if leaders have ideologically extreme preferences, 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 and raises the likelihood of successful reform. Why did radical leaders often refuse to settle for less radical but more realistic agendas so as to boost support? Why was it common that the leadership eventually was taken over by more radical groups who escalated their agenda to the far end of the spectrum?

In this paper, we develop a theory to analyze the rise of 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 they need a change but disagree about 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 the protest if it is sufficiently attractive for them. 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 stimulate followers, given that protestors' actions are costly. But when the agenda is very radical, relatively moderate citizens may choose to become bystanders rather than followers. Not surprisingly, the leader strategically chooses an agenda by trading off the chances of success against his own policy preference. Section 3 elaborates on this simple framework.

Two key elements will be added to this benchmark model. In Section 4, we introduce

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

asymmetric information by assuming that 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 one. But when citizens make inferences about the state based on the scale of reform proposed by the opposition leader, a larger fraction of citizens will be mobilized in the high state. Therefore, the leader may have an incentive to exaggerate 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 his own policy preference. Instead he may have to resort to "irrational radicalism" by choosing a scale of reform that is even larger than what is warranted by the bad situation. Although he knows that radicalism would not be popular with citizens, he cannot soften his position; otherwise, citizens would interpret that he is choosing an agenda to shore up support.

If the leader is known to have a moderate preference, he 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 so as to convince the citizens that the current situation is indeed bad. This type of signaling is costly for the leader: the probability of success is reduced under asymmetric information. If the leader's ideology preference happens to be very radical, his agenda would be distorted so much that the winning probability in the high state (when conditions are favorable for success) is even lower than that in the low state.

In Section 5, we further enrich the model by introducing endogenous determination of leadership. Conflict of interests between "marginal leaders" and the "representative leader" provides a mechanism that leads to the escalation of radicalism. If the leadership group is more radical, they prefer a relatively more radical agenda. This discourages the moderates in the leadership, because the agenda is too radical to their tastes and the chances of success are smaller. As a result, the most moderate leaders in the group drop out, which makes the preferences of the remaining leaders more radical on average, and they would in turn propose even more radical agendas. We show that such an escalation mechanism is stable, which gives rise to a unique endogenous leadership group in equilibrium.

Asymmetric information and endogenous escalation interact and reinforce each other. First, under some circumstances, it is indeed the radical citizens that form the leadership and take extra efforts to lead. The radical leader suffers from citizens' suspicion that he may be bluffing to shore up support, which compels him to radicalize his proposal to signal. Second, if the identity of the leader is exogenously fixed, the problem of asymmetric information makes the reform agenda more radical in the high state but has no effect on the reform agenda in the low state, because there is no need to signal a low state to citizens. However, once citizens can choose whether to join the leadership, this is not the case anymore. The mere prospect that the state may be high and that accordingly the reform agenda would be irrationally radical tends to repel moderates from joining the ranks of the opposition leadership. This mechanism ensures that even the reform agenda chosen in the low state will be relatively radical, because citizens who choose to become leaders are relatively radical.

In Section 6, we illustrate agenda escalation with historical examples through the lens of our model. We also relate our theory to an empirical puzzle: in societies with structural roots of political instability, political upheavals with mass support are not observed as aften (Geddes 1990; Goldstone 2001). According to our model, this puzzle is less surprising than it seems, because a society ripe for revolts is also a breeding ground for radical leaders, whose radical agendas would undermine the prospect of success.

To be sure, our model can account for certain types of radicalism but not all. If leaders have other means to credibly transmit the information they possess, asymmetric information between leaders and citizens would not cause leaders to use radicalism as a signaling device. In fact, one contribution to the literature of regime change made in this paper is that we 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 the leaders. Agency issues arise because their interests are not perfectly aligned and because leaders possess superior information. In contrast to a standard principal-agent problem, however, the choice of taking up the leadership role and becoming "agents" is endogenous. In addition, comparing to a standard signaling game, the cost of signaling in this model is endogenously determined by a protest game, rather than exogenously specified.

2. Review of Literature

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 efforts of participation; and the leader proposes a revolutionary agenda to attract followers, but also to induce their efforts. Therefore, the leader strategically chooses the degree of radicalism in his agenda by trading off between extensive and intensive margins. Michaeli and Spiro (2018) analyze under what conditions radicals rather than moderates would initiate dissent, and the strength of dissent 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 in three aspects. First, we only allow for 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. Third, asymmetric information plays a large role in our model but not in the aforementioned works.

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 or the aggregate state. Shadmehr and Bernhardt (2019) use their model to explain why the regime may gain by radicalizing the revolutionary vanguards: very radical vanguards are too eager to revolt, which de-legitimize these vanguards in the eyes of the followers. Bueno de Mesquita (2010) studies a coordination game where vanguards can conduct violence, which acts as a public signal of aggregate sentiment to followers. Our model is a signaling game, i.e., leaders propose an agenda that signals the mismatch between the state and the status quo policy, which is known to the leaders but not to the followers.

Some other studies focus on the motivating role of leadership in regime change games. In a global game setting, Morris and Shadmehr (2018) allow rewards received by followers to vary according to their efforts, and the leader wants to maximize the likelihood of regime change by choosing a reward function. In our work, we focus on the mechanism that generates the leadership, the associated policy proposal and its impact on regime change. Majumdar and Mukand (2010) highlight the importance of complementarity in efforts taken by followers and the leader in social movements, which illustrates why leaders can be the catalyst. Our focus instead is on the asymmetric information friction and unaligned interests within the opposition group.[2]

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 median voter's preference) as a way to signal that they are not captured by right-wing special interests. Kartik and McAfee (2007) study how politicians choose political platforms to signal their "character," an attribute that voters value. In our collective action setting, signaling is about a common factor (the "state") that affects the payoff to all citizens, rather than about the personal qualities of the leader. 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.[3] 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.

The key mechanism that drives radicalism in mass movements in our model is asymmetric information between leaders and followers. In electoral politics, there are studies that explore how informational friction between voters and political candidates can lead

[2] Our work is also broadly related to the literature concerning the coordinating role of leadership in environments where followers wish to align their actions (Dewan and Myatt 2007; 2008; 2012). Bolton, Brunnermeier, and Veldkamp (2012) show that the right characteristics of leaders facilitate the trade off between commitment and adaptation in a corporate environment. Landa and Tyson (2017) argue that leaders with strong coercive power can improve the transmission of information to followers.

[3] See also Fu, Li, and Qiao (2018) for the signaling approach to leadership in organizations.

to extremism. For example, Boleslavsky and Cotton (2015) show that voters may tolerate extremism policies more when they are better informed about the quality of candidates during political campaigns.

3. Strategic Reform Agenda in Protest Game

Our model embeds two mechanisms that capture salient features in mass political movement: (1) asymmetric information between opposition leaders and their followers exists; and (2) the opposition leadership is endogenous. To lay the groundwork, in this section, we describe a benchmark model in which we abstract from asymmetric information and fix the leadership preference. We specify the payoff structures of leaders and followers, and analyze how an exogenous leader strategically chooses the political agenda and how citizens make participation decisions. In Section 4, we extend this model to allow for asymmetric information. In Section 5, we further develop it by allowing the leadership to form endogenously.

3.1. The benchmark model

Consider a society populated by a unit mass of citizens, indexed by i, who are not satisfied with the status quo but have heterogeneous preferences regarding the appropriate policy for society. The preference of citizen i is parameterized by x_i, which is uniformly distributed on $[0,1]$. In general, citizens prefer to align the policy, denoted by y, to the current situation in society, which we refer to as the "state" and denote by θ. The payoff to citizen 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 for citizen i is $x_i + \theta$, which we refer to as his *ideal policy*. We say that citizen i is *more radical* if his preference x_i is higher. The ideal policy for the most radical citizen is $1+\theta$. In the benchmark model, we maintain the assumption that the state θ is fixed and commonly known.

The status quo policy is y_0. We assume that $y_0 < \theta$, so that there exists a mismatch between the status quo policy and the state, which drives the demand for reform. If the state θ is high, the grievance against the regime is strong. Since every citizen's ideal policy is to the right of y_0, all citizens agree that some reform is desirable but they disagree on which reform is best. Indeed, if a reform policy is farther from a citizen's ideal policy than the status quo is, this citizen will prefer maintaining the status quo. The assumption that different citizens have different ideal policies captures the notion that the interests of citizens are not perfectly aligned.[4]

In the benchmark model, we assume that there exists an opposition leader, with preference x^m, who formulates and proposes an alternative policy y_1 to replace the status quo. 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. Once the policy proposal of the leader is announced, each citizen decides 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*.

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. 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 pre-revolutionary promises with impunity. In this paper, we abstract from the issue of commitment.[5]

[4] The assumption that all citizens have ideal policies to the right of y_0 is made only for simplicity of exposition. If there are citizens whose ideal policies are to the left of the status quo, they will be supportive of the existing regime and will not choose to be an opposition leader or choose to attack.

[5] To accommodate the situation where opposition leaders do not have full commitment power, we may consider an alternative model in which, upon success, the leader implements the proposed policy only with some probability, but chooses to re-optimize with a different policy with complementary probability. This alternative model is qualitatively similar to ours.

The utility from regime change and the cost incurred differ across three types of citizens: bystanders, followers, and leaders. In the benchmark model, the decision to become a *leader* ($l_i = 1$) or not ($l_i = 0$) is not considered; we just take the identity of the leader as fixed.

Under the assumption that each citizen is atomistic, his participation decision has no influence on the total size of the attack (i.e., A) in this model. The payoff to a bystander (i.e., $a_i = 0$ and $l_i = 0$) 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*, and is equal to $u(y_1, \theta, x_i) - u(y_0, \theta, x_i)$. 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 in Wood (2003) and formalized in Morris and Shadmehr (2018). It refers to "the value they [revolutionaries] put on being part of the making of history" (Wood 2003, p. 38). Morris and Shadmehr (2018) 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.[6] We denote $FP(x_i) := U_f(x_i) - U_b(x_i)$ as the "follower's premium." Given θ, A, and y_1, citizen i chooses $a_i = 1$ if and only if $FP(x_i) \geq 0$.

Individuals who choose to be part of the leadership group are involved in organizing the opposition movement and formulating a policy alternative y_1, and incur a cost of c_l for these leadership activities. The leadership role is also more rewarding: the 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:

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

We denote $LP(x_i) := U_l(x_i) - \max\{U_b(x_i), U_f(x_i)\}$ as the "leader's premium." The leader's premium plays a role in determining whether a citizen chooses to be a leader or not, a decision which we analyze in Section 5. In the benchmark model considered in this section, we assume that the identity of the leader is fixed, with preference x^m. The exogenous leader x^m chooses a reform agenda y_1 to maximize $U_l(x^m)$.

Throughout this paper, we maintain the following assumption on the success determination function $G(A)$.

Assumption 1. 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 then decreasing returns. Assumption 1 can accommodate this type of success determination technology, because log-

[6] 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 successful reform is a common device (e.g., Bueno de Mesquita 2010) used to rationalize the motive for taking collective actions and to abstract from free-riding issues when modeling citizens as atomistic agents. DellaVigna et al. (2017) examines the hypothesis that people vote because they derive pride from telling others that they voted. This behavioral motive in the context of voting is a counterpart to the pleasure in agency from participation in large social movements.

concavity is consistent with increasing returns as long as $G'(A)/G(A)$ is non-increasing. The second part of Assumption 1 (i.e., $G(0) > c_f/k_f$) is a sufficient condition that guarantees the existence of non-trivial equilibria with positive mass of attack. If it is a dominant strategy for the most radical citizen $x_i = 1$ to attack when the reform agenda is his ideal policy (i.e., if $FP(1; A = 0, y_1 = 1 + \theta) > 0$), then this condition holds. This condition requires that the cost of attack cannot be too high relative to the pleasure in agency motive. It also 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 revolution leaders and followers. We can interpret $G(0)$ as the probability that these "background factors" can potentially bring down the regime.[7]

3.2. Equilibrium of the protest game

Given the status quo policy y_0 and the reform agenda y_1, citizens decide whether to attack the regime or not in a simple protest game. For any expected size of attack A, citizen i chooses $a_i = 1$, if and only if $FP(x_i; A, y_1) \geq 0$. The function $FP(\cdot; A, y_1)$ is weakly increasing; therefore there exists a *marginal attacker*, x^f, satisfying $FP(x^f; A, y_1) = 0$ such that citizen i attacks if and only if $x_i \geq x^f$. Figure 1 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 the dependence of outcome of the protest game on the reform agenda and the state, we use $x^f(y_1; \theta) \in (0, 1)$ to denote the equilibrium marginal attacker that sat-

[7] The condition $G(0) > c_f/k_f$ is made for convenience only, because we want to focus on the interesting case where equilibrium of the protest game is not trivial. Also note that this condition is sufficient but not necessary for the existence of equilibria with positive mass of attack. For example, we can show that such equilibria exist if $c_f/k_f < 1/4$ when $G(A) = A$, even though the second part of Assumption 1 fails to hold.

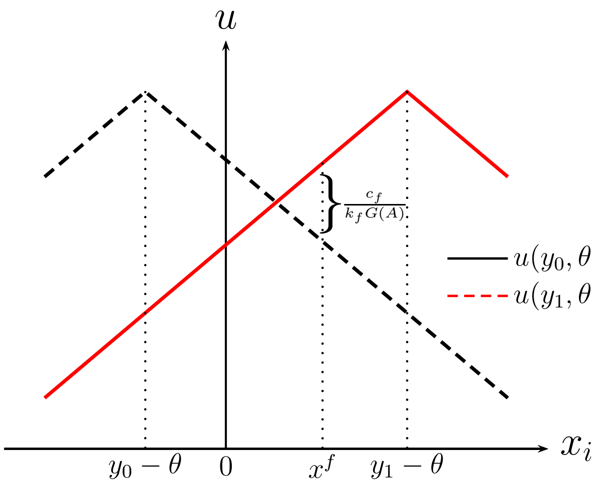


Figure 1. *The determination of the marginal attacker x^f. The utility difference between the solid line and the dashed line represents the reward from success for citizen with preference x_i. The follower's premium is non-negative if and only if $x_i \geq x^f$.*

isfies 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.

Proposition 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 with $x^f(y_1;\theta) = 1$; (b) if $y_1 \geq y_{\min}(\theta)$, then there exists an equilibrium with positive mass of attack and with $x^f(y_1;\theta) \leq y_1 - \theta$; further, the equilibrium marginal attacker $x^f(y_1;\theta)$ increases in y_1 and decreases in θ.*

In case (a) of Proposition 1, the reward from success is small as 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 equilibrium. In case (b), the reform agenda y_1 is sufficiently far from the status quo. Because a successful mass movement would bring about a more sizeable change in society, at least a fraction of the relatively more radical citizens would find it worthwhile to attack, i.e., a non-trivial equilibrium exists. In this case, there may be multiple equilibria, one with $x^f \leq y_1 - \theta$ and another with $x^f > y_1 - \theta$.[8] Throughout the paper, we focus on the equilibrium with the largest equilibrium attack size (i.e., the one with $x^f \leq y_1 - \theta$), for the following reasons. First, it 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.[9]

[8] 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.

[9] Consider the other possible equilibrium with $x^f < y_1 - \theta$. Even when this equilibrium exists, it requires

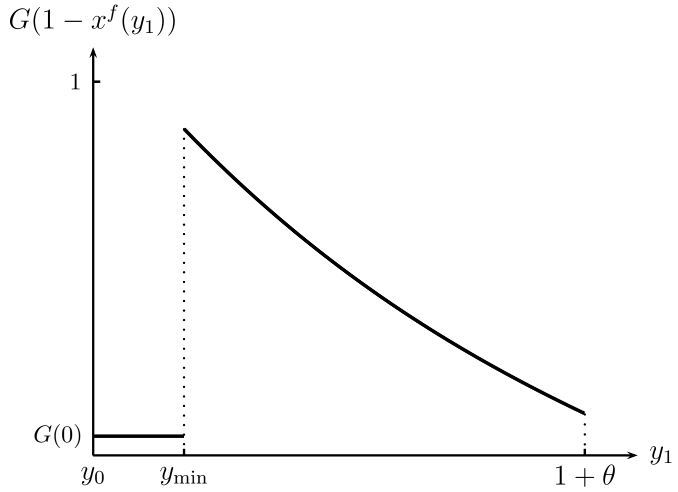


Figure 2. *The probability of success in equilibrium is $G(0)$ when $y_1 < y_{\min}$, jumps up at $y_1 = y_{\min}$, and then decreases as y_1 increases.*

Second, it is the only stable equilibrium with meaningful comparative statics.[10] Third, the main focus of this paper is the determination of the reform agenda rather than the protest game itself.

When $y_1 \geq y_{\min}$, a more radical reform agenda y_1 suppresses participation in equilibrium. First, because $x^f \leq y_1 - \theta$, the reform agenda y_1 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 causes the marginal attacker to shift to the right even further. Consequentially, the marginal attacker becomes more radical and the probability of success falls correspondingly. See Figure 2 for illustration.

Proposition 1 also implies that, for given y_1, the probability of success $G(1-x^f(y_1;\theta))$ increases in θ. That is because a larger mismatch between the state and status quo policy causes more citizens to prefer changing the status quo, as the reward from success is increasing in θ for any $y_1 > y_0$.

citizens with preference slightly lower than x^f to choose not to attack and citizens with preferences above x^f to choose to attack, despite the fact that these two groups of citizens obtain exactly the same reward from success.

[10] The proof of Proposition 1 establishes that $FP(x^f; 1-x^f, y_1)$ is locally increasing at this equilibrium value of x^f. Hence, this equilibrium is locally stable in the sense of best-response dynamics.

3.3. Strategic choice of reform agenda

Now we turn to the agenda-making stage. The leader's preference x^m is exogenously given and he chooses y_1 to maximize his own payoff $U_l(x^m)$, taking into account its impact on the protest game. 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 he 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 $x^m + \theta$ is to the right of $y_{\min}$.[11] For $x^m + \theta > y_{\min}$, the leader would never find it optimal to choose $y_1 > x^m + \theta$, because he can always choose a smaller policy which is closer to his ideal policy and brings a higher probability of success. He would not choose $y_1 < y_{\min}$ either, because such agenda would induce no attack in the protest game. For $y_1 \in [y_{\min}, x^m + \theta]$, we show that 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:

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

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 chances of getting the reward from success is lowered as $\partial x^f / \partial y_1$ is negative. The first-order condition (4) optimally balances this trade-off. Because such an 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.

[11] 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.

Proposition 2. *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:*

$$y_1^*(x^m;\theta) = \begin{cases} x^m + \theta & \text{if } x^m \in [y_{\min}(\theta) - \theta, \hat{x}(\theta)), \\ \hat{x}(\theta) + \theta & \text{if } x^m \geq \hat{x}(\theta). \end{cases} \tag{5}$$

When x^m is relatively small (i.e., $x^m \leq \hat{x}(\theta)$), the optimal solution is a corner solution, with the leader choosing his own ideal policy. When x^m is large, the leader cannot afford picking his ideal policy, because this would discourage people from following the mass movement, reducing the chances of success. So the leader has to compromise by choosing the agenda $y_1^* = \hat{x}(\theta) + \theta$, which is less radical than his ideal policy. In general, the strategic agenda $y_1^*(x^m;\theta)$ is weakly increasing in x^m. The fact that the optimal policy y_1^* is constant for $x^m \geq \hat{x}(\theta)$ is a consequence of the piecewise linear utility function.[12]

Proposition 2 shows that for any $x^m \geq y_{\min} - \theta$, we have $x^m + \theta \geq y_1^* \geq y_{\min}$. In Proposition 1, we have shown that $y_1^* \geq y_{\min}$ implies $y_1^* \geq x^f + \theta$. Together, they imply $x^m \geq x^f$, meaning that the leader always chooses to attack in equilibrium.

Proposition 3. *For any leader with $x^m \geq y_{\min}(\theta) - \theta$, the strategic agenda $y_1^*(x^m;\theta)$ increases in the state θ; so does the probability of success, $G(1 - x^f(y_1^*(x^m;\theta);\theta))$.*

For a fixed reform agenda, the chances of success in a protest game are higher when the state is higher, causing the opposition leader to put more weight on the reward from success. Proposition 3 shows that the opposition leader chooses a more radical agenda in response to a higher state, but would never raise y_1^* to the point where it hurts the chances of success. In other words, $G(1 - x^f(y_1^*(x^m;\theta);\theta))$ always increases in θ.

The result that the leader chooses a more radical policy when the state is high is important, because it provides the leader an incentive to mislead citizens when the leader

[12] Whenever $x^m + \theta > y_1 > y_0$, the reward from success for the leader x^m is $y_1 - y_0$, which does not depend on x^m locally. Therefore, this strategic reform policy y_1^* also does not depend on x^m. In an alternative setup with quadratic utility function, the reward from success strictly increases in the leader's preference x^m even locally. The two specifications deliver very similar results, but the linear specification is more tractable.

knows but citizens do not know the state. In the next section, we will elaborate on this point formally. We will also show that, under some circumstances, asymmetric information may cause the leader to "over-react" to the state by choosing a policy so radical that it reduces the probability of success, despite having a high state that is favorable to the mass movement.

4. Asymmetric Information and Radicalism

4.1. Model description: information asymmetry

In this section we extend the benchmark model to study the effects of asymmetric information between leaders and citizens by allowing for uncertainty about the state. Specifically, assume that there are two states: θ_H and θ_L, with $\theta_H > \theta_L$. In the high state θ_H, the discrepancy between the current circumstances and the status quo is larger. This environment is more favorable for a revolt, because for any $y_1 > y_0$ the reward from success is larger when θ is higher. The prior probability of the two states are π_H and π_L, with $\pi_H + \pi_L = 1$.

The key assumption is that the leader observes θ but other citizens do not. Moreover, the leader cannot produce objective verifiable evidence that credibly transmits his 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. Citizens observe the agenda y_1, make an inference about the state, and decide whether to attack or not (choose a_i).

Our assumption about information asymmetry is reasonable. Although citizens may derive pleasure in agency from joining a mass movement, each individual follower has little influence on the outcome. Hence the incentive for citizens to obtain accurate information about the aggregate state in society (and thus the relative merits of different agendas) is small. Revolution leaders, on the other hand, have a much greater stake in their cause. They have to formulate a policy alternative, organize the masses, and convince them that their alternative is superior to the status quo. Because their choice of agenda has a material effect on the outcome of the mass movement, they tend to spend more time and effort to learn about the environment. In this paper, we abstract from the information acquisition decision and simply 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 signaling literature (e.g., see Hermalin 1998).

4.2. Equilibrium analysis: separating equilibrium

To fix notation, we let $y_1^{L*} := y_1^*(x^m; \theta_L)$ and $y_1^{H*} := y_1^*(x^m; \theta_H)$ represent the optimal agenda chosen by x^m in the two states when there is complete information about θ, which are given by Proposition 2 in Section 3 for the benchmark model. We conduct our analysis for an exogenous leader, with preference

$$x^m \geq y_{\min}(\theta_L) - \theta_L.$$

This ensures that $y_1^{L*} > y_{\min}(\theta_L)$ and $y_1^{H*} > y_{\min}(\theta_H)$, so that in each state there will be a positive mass of attackers in the equilibrium of the protest game.[13] 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 3 establishes that $y_1^{H*} > y_1^{L*}$ and $G^{H*} > G^{L*}$. These reform agendas and success probabilities are the full-information outcomes for the relevant states.

For the same reform agenda, the chances of success are higher if citizens believe the state is high (Proposition 1). This creates an incentive for an informed leader to mislead citizens. Even though the state is low, the leader may be better off telling followers that the state is high in order to shore up support for a mass attack. But citizens would be rightly skeptical of such rhetorics, given that such messages are unverifiable. This makes cheap talk about the state ineffective in transmitting the information possessed by the leader. When a leader lacks an effective means to convey information about the state, he has to rely on the choice of reform agenda to credibly signal his private information. Such a signaling mechanism is costly in the sense that the full-information outcome becomes infeasible if the preference of the leader is relatively high.

Proposition 4. *There exists a unique threshold $x^\dagger \in (\hat{x}(\theta_H), \hat{x}(\theta_H) + \theta_H - \theta_L)$ such that the*

[13] Because a higher θ increases the payoff from attacking, the least radical agenda $y_{\min}(\theta)$ that can bring about positive attack is decreasing in θ. Therefore, $x^m \geq y_{\min}(\theta_L) - \theta_L$ implies $x^m > y_{\min}(\theta_H) - \theta_H$.

full-information outcome (y_1^{L*}, y_1^{H*}) is an equilibrium outcome under asymmetric information if and only if $x^m \leq x^\dagger$.

In state θ_L, a leader with a higher x^m has greater incentive to mislead citizens into believing that the state is high. First, regardless of his preference x^m, the leader benefits from choosing y_1^{H*} so as to mimic the high state, because the chances of success would increase from G^{L*} to G^{H*}. Second, by choosing y_1^{H*} instead of y_1^{L*}, the reward from success would change from $y_1^{L*} - y_0$ to $2x^m + 2\theta_L - y_1^{H*} - y_0$. The net change is increasing in x^m. Moreover, it is negative when x^m is low and positive when x^m is high.

In other words, a leader with small x^m faces a trade-off between a higher chances of success and a lower reward from success when he lies, while both are higher for a leader with large x^m. Proposition 4 shows that there exists a unique $x^\dagger$ such that the second force dominates if and only if $x \leq x^\dagger$. If $x^m \leq x^\dagger$ and the leader tries to mislead, he has to adopt the agenda y_1^{H*}, which is far to the right of his ideal policy. Even though the chances of success are larger, it is still not worth it. 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 based on the agenda chosen by the leader.

Suppose the leader's preference is radical ($x^m > x^\dagger$). The leader finds that y_1^{H*} is not too far from or even closer to his ideal policy in the low state, and the odds of success are also better. This radical leader would have incentive to mimic the high state by proposing y_1^{H*}. Of course, citizens would not be systematically fooled into believing that the state is high when they see 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.[14]

Proposition 4 is a core result of our theory. When the leader is moderate, citizens believe that he does not have incentive to mislead, and he indeed behaves "optimally" by striking a balance between the probability of success and the reward from success in each state. In contrast, when the leader is radical, he cannot convince citizens that the state is high

[14] The threshold $x^\dagger$ may be greater than 1 if θ_H is very large compared to θ_L. In this case, even the most radical leader with preference $x^m = 1$ would not prefer to mimic the high state because $y_1^*(1; \theta_H)$ is far to the right of his ideal point $1 + \theta_L$ in the low state. In general, as long as θ_H is not too large, we have $x^\dagger < 1$, and the full-information outcome will not be always achievable.

by proposing the strategic policy y_1^{H*}. To credibly convey the information that the state is high, the agenda chosen must be even more radical than y_1^{H*}, so that the leader would not have 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; he has to resort to "irrational radicalism" to separate from leaders who may otherwise bluff about the state in order to boost support.

In the following proposition, we characterize the least-cost separating equilibrium which satisfies the D1 criterion (Banks and Sobel 1987; Cho and Kreps 1987). 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 5. *Suppose $x^m > x^\dagger$. In the least-cost separating equilibrium that satisfies the D1 refinement, we have $\hat{y}_1^L = y_1^{L*}$ and $\hat{G}^L = G^{L*}$ in the low state, and $\hat{y}_1^H > y_1^{H*}$ and $\hat{G}^H < G^{H*}$ in the high state. Furthermore, if $x^m > \hat{x}(\theta_L) + \theta_H - \theta_L > x^\dagger$, then $\hat{G}^H < \hat{G}^L$.*

In a separating equilibrium, the leader would not propose an agenda different from y_1^{L*} in the low state, because this strategic agenda optimally balances his policy preference and the chances of success. In other words, we must have $\hat{y}_1^L = y_1^{L*}$. However, for any leaders with $x^m > x^\dagger$, proposing an agenda equal to y_1^{H*} in the high state will not be an equilibrium, because the leader would have incentive to mislead citizens by proposing y_1^{H*} even when the state is low. To prevent him from overstating the state to be more favorable than it is, the equilibrium policy $\hat{y}_1^H$ must exceed y_1^{H*}.[15] Such 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. The reform agenda in the high state under the least-cost separating equilibrium which satisfies the D1 refinement is the smallest $y_1^H > y_1^{H*}$ such that the leader's incentive constraint in

[15] 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 $\hat{y}_1^L$ in each state. However, 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 for the high type to deviate from $\tilde{y}$ to y'. See the proof of Proposition 5.

the low state is just binding:

$$\hat{G}^L \left(u(\hat{y}_1^L, \theta_L, x^m) - u(y_0, \theta_L, x^m) \right) - \hat{G}^H \left(u(y_1^H, \theta_L, x^m) - u(y_0, \theta_L, x^m) \right) = 0. \tag{6}$$

When the leader is indifferent between $\hat{y}_1^L$ and $\hat{y}_1^H$ in the low state, he strictly prefers $\hat{y}_1^H$ to $\hat{y}_1^L$ in the high state. The single-crossing condition for 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 he is indifferent between these two agendas in the low state. Figure 3 illustrates.

The result that $\hat{y}_1^H > y_1^{H*}$ implies $\hat{G}^H < G^{H*}$, because the probability of success decreases in the reform agenda. Informational asymmetry between the leader and citizens drives the leader to choose a more radical agenda than he would prefer under complete information, so as to signal the state is indeed high. Such signaling is costly, as it reduces the equilibrium probability of success.[16] In the high state, the leader cannot moderate his position, even though he knows that his radical agenda discourages citizens. In the separating equilibrium described in Proposition 5, 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 policy in the low state.

Interestingly, when the leader is very radical (specifically, $x^m > \hat{x}(\theta_L) + \theta_H - \theta_L$), the agenda $\hat{y}_1^H$ that separates the high state is so radical that the probability of success is even lower than that in the low state, i.e., $\hat{G}^H < \hat{G}^L$. Such an equilibrium outcome is qualitatively different from the complete information case. When information is complete, Proposition 3 establishes that a higher state leads to a higher equilibrium chance of success. The leader proposes a higher agenda in a higher state, but he never "over-reacts" to a more favorable environment by choosing a radical agenda to the extent that it hurts the chances of success. In contrast, under asymmetric information, a very radical leader can propose an agenda so radical that the equilibrium chances of success become even lower in the high state than in

[16] In this signaling game the "cost" of proposing a radical agenda is not exogenously imposed, but is endogenously determined in the equilibrium of the protest game.

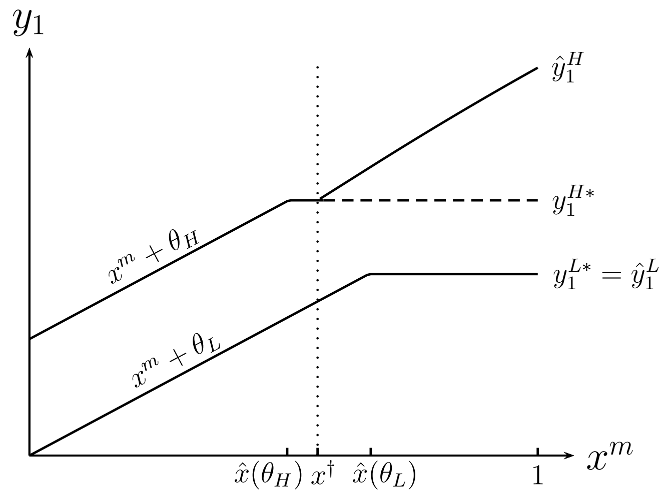


Figure 3. *The optimal agenda pair is (y_1^{L*}, y_1^{H*}) under complete information. When $x^m > x^\dagger$, the leader chooses $\hat{y}_1^L = y_1^{L*}$ and $\hat{y}_1^H > y_1^{H*}$ in a separating equilibrium and $\hat{y}_1^H$ strictly increases in x^m.*

the low state, despite an external environment that is more favorable to success.

To emphasize the dependence of the equilibrium agendas on the preference of the leader, we sometimes write $\hat{y}_1^L(x^m)$ and $\hat{y}_1^H(x^m)$. The equilibrium $\hat{y}_1^L(x^m)$ is the same as $y_1^*(x^m; \theta_L)$; Propositions 2 and 3 show that it weakly increases in x^m and strictly increases in θ_L. The following result provides comparative statics for the equilibrium $\hat{y}_1^H(x^m)$.

Proposition 6. *For $x^m > x^\dagger$, the equilibrium agenda $\hat{y}_1^H(x^m)$ in the high state is strictly increasing in x^m and θ_H; so is the gap $\hat{y}_1^H(x^m) - \hat{y}_1^L(x^m)$.*

As the leader becomes more radical, his incentive to exaggerate the state and mislead citizens increases. As a result, he needs to propose more radical agendas in the high state in order to remain credible. This explains the first part of Proposition 6. Recall that Proposition 2 shows that both y_1^{H*} and y_1^{L*} do not depend on x^m for $x^m > \max\{\hat{x}(\theta_H), \hat{x}(\theta_L)\}$. In contrast, in the least-cost separating equilibrium, $\hat{y}_1^H$ strictly increases in x^m for all $x^m > x^\dagger$, and the gap between $\hat{y}_1^H$ and $\hat{y}_1^L$ increases in x^m too. Figure 3 illustrates. The result that $\hat{y}_1^H(x^m)$ increases in x^m, is not only intuitive, but also important for the escalation mechanism that we will elaborate in the next section.

5. Leadership Radicalization and Agenda Escalation

In the previous section, we show that a radical leader may have to radicalize his agenda to signal that he is not lying for the purpose of gathering support, while a moderate leader does not have incentive to do so. A relevant question is, under what conditions would radicals rather than moderates assume the leadership role to formulate the reform agenda? While acknowledging that there are multiple routes to opposition leadership, including by sheer chances in many historical events, we analyze in this section how the leadership endogenously emerges in equilibrium. We characterize one particular mechanism in which self-selection into the leadership group and the divergence of interests within this group systematically cause the leadership to radicalize and the reform agenda to escalate. This mechanism of leadership selection and the mechanism of signaling reinforce each other, which, under some circumstances, produces an equilibrium outcome that radicals become leaders and propose even more radical agendas which cannot be justified by their policy preferences.

5.1. Model description: endogenous leadership

In this section we further enrich the model in Section 4 by allowing the opposition leadership to comprise a group of citizens instead of a single leader, and by making the choice of joining the leadership group endogenous.

Specifically, we assume that each citizen independently makes two decisions sequentially. Citizen i first chooses to be a leader of the mass movement ($l_i = 1$) or not ($l_i = 0$). When citizens decide whether to become leaders (choose l_i), they do not observe the state and only know its distribution. In each state, the payoff to any citizen who chooses to be a leader is given by equation (1) in Section 3.

Given diverse preferences within the leadership group, the leaders disagree about the policy to be proposed. We assume that it is the citizen with the median preference among the leadership group who will formulate and propose the reform agenda y_1 for the mass movement. In other words, the task of drafting the policy proposal is delegated to the

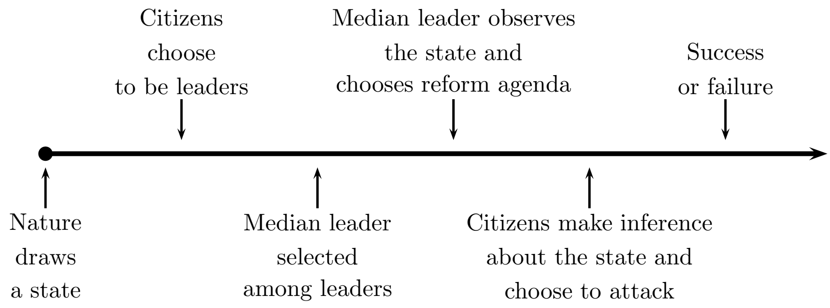


Figure 4. *Timeline of the game with asymmetric information and endogenous leadership.*

median citizen in the group, who will be referred as the *median leader* in the following. The median leader then observes the state, makes a decision on what to propose, and announces it to all the other citizens. We assume that the status quo y_0 would be the policy proposal if nobody chooses to join the leadership. Upon knowing the proposal, each citizen decides whether to participate in attacking the regime ($a_i = 1$) or not ($a_i = 0$). 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 or identity of the median leader is endogenously determined and known to other citizens.

The assumptions related to the emergence of leaders deserve discussion. First, in this model there will be a continuum of leaders. Only the median leader chooses the agenda; what he proposes affects the payoffs to all leaders but not the other way around, because each individual leader is atomistic and his 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.[17] Second, the assumption that the median leader chooses the agenda is a useful shortcut. On the one hand, it captures the idea that the proposed agenda should reflect the preferences of the leadership group as a whole. On the other hand, it abstracts away the process of conciliation or even struggle among leaders with diverse opinions on reform proposals, which is not the focus of this paper.[18]

[17] In our model, there exist infinitely many leaders, while in Shadmehr (2015) and other works, there is only one leader. The two approaches capture two ends of possible modeling choices. Having a finite number of leaders strategically interacting with each other and with the large crowd of citizens will be an interesting topic to pursue on its own.

[18] Our assumption can be formally justified by a majority voting process within the leadership group. In fact, the precise policy formulation mechanism matters little for our subsequent analysis and results. For example, if the reform agenda is chosen by someone with preference parameter equal to the mean rather

5.2. Equilibrium analysis: escalation

Because citizens do not know the state at the time 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

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

For given x^m, citizen i chooses to join the leadership group if and only if the expected leader's premium is non-negative. 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. Therefore, 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 identity of the marginal leader 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:

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

Equilibrium in this model is characterized by a 6-tuple, $(x_*^m, x_*^l, x_{L*}^f, x_{H*}^f, \hat{y}_{1*}^L, \hat{y}_{1*}^H)$, that

than the median, or if the pivotal leader is selected at a quantile different from the median, our results will remain largely intact as long as the "representative leader" has preferences that broadly reflect the distribution of preferences among leaders.

satisfies the following requirements:

1. The median leader satisfies $x_*^m = M(x_*^m)$.
2. The marginal leader 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 $\hat{y}_{1*}^L = y_1^*(x_*^m; \theta_L)$ in equation (5).
4. The reform agenda in the high state is given by $\hat{y}_{1*}^H = y_1^*(x_*^m; \theta_H)$ in equation (5) if $x_*^m \leq x^\dagger$; it is given by the solution to the binding incentive constraint (6) for $x^m = x_*^m$ and $\hat{y}_1^L = \hat{y}_{1*}^L$ if $x_*^m > x^\dagger$.
5. The marginal attacker in state $j \in \{H, L\}$ satisfies $FP(x_{j*}^f; 1 - x_{j*}^f, \hat{y}_{1*}^j, \theta_j) = 0$.

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

In our model, because $\hat{LP}(x_i; x^m)$ is weakly increasing in x_i, any citizen with preference $x_i \geq x^l(x^m)$ will choose to be a leader. This means that the median leader x^m in general is more radical than the marginal leader $x^l(x^m)$. This feature of the model provides a mechanism for the radicalization of leadership. Such a mechanism can be best understood in terms of best-response dynamics. If the median leader becomes more radical (x^m shifts to the right), he proposes more radical agendas ($\hat{y}_1^L(x^m)$ and $\hat{y}_1^H(x^m)$ are both weakly higher, as shown by Propositions 3 and 6). However, the more radical reform agendas proposed by the median leader hurts the marginal leader, as the latter has less radical preferences than the former. In other words, for $x^l < x^m$, $\hat{LP}(x^l; x^m)$ decreases as x^m increases, which implies that the marginal leader $x^l(x^m)$ will shift to the right as x^m shifts to the right. But as the marginal leader shifts to the right, the distribution of preferences of the leadership group becomes more radical, giving rise to a median leader who is more radical. In other words, the leadership group will be radicalized if the conjectured median leader becomes more radical, and the reform agendas will escalate too. This escalation mechanism is reflected in the fact that the mapping $M(x^m)$ is increasing in x^m.

Proposition 7. *Suppose* $\min\{\hat{LP}(1/2;1/2), \hat{LP}(1;1)\} > 0$ *and* $y_{\min}(\theta_L) - \theta_L < 1/2$. *The mapping* $M(\cdot)$ *is continuous and increasing on* $[1/2, 1]$. *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* $\theta_H - \theta_L \leq 1/3$*, the equilibrium is unique.*

To ensure the existence of a non-trivial equilibrium with positive mass of attackers in both states, we impose the condition $y_{\min}(\theta_L) - \theta_L < 1/2$, which is sufficient (but not necessary). In general, the minimum reform agenda $y_{\min}(\theta)$ is low if citizens are likely to attack. This condition can alternatively be stated as requiring c_f/k_f to be smaller than a threshold.

To ensure that the leadership is not empty in equilibrium, we impose the condition that $\min\{\hat{LP}(1/2;1/2), \hat{LP}(1;1)\} > 0$. It guarantees 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. This condition simply says that a citizen will choose to be a leader if he expects himself to be the one who proposes the reform agendas. This condition can alternatively be stated as requiring c_l/k_l to be smaller than a threshold.

Proposition 7 establishes that $M(x^m)$ is increasing when it is well-defined, and a fixed point of the mapping exists. The solid line in Figure 5 illustrates. Moreover, $M(x^m) - x^m$ is single-crossing from above when θ_H is not too large relative to θ_L, which guarantees that the fixed point is unique. Intuitively, $M'(x^m)$ will be smaller if x^l does not increase that much when x^m rises. In other words, fewer leaders on the left end of the leadership drop out when the median leader is more radical. This is true when $\hat{y}^H_{1*}(x^m)$ is less responsive to x^m, which is indeed the case when θ_H is small.[19]

[19] The proof of Proposition 6 shows that $\partial \hat{y}^H_{1*}(x^m)/\partial x^m$ strictly increases in x^m, and x^m_* is small when θ_H is small.

6. Discussion

6.1. Radicalization and suppression

In our model, the radicalization of leadership relies on a mechanism in which relatively moderate leaders retreat from the leadership when radical ones take on the role of proposing reform agendas. Such a mechanism is useful for interpreting the trajectory of leadership radicalization in historical and contemporary mass movements.

A good case in point is the experience of the Democracy Movement of China in 1989. At the early stage of the movement, the demands of protestors were modest, practical and even evidently strategic. The major requests mainly involved cleaning up corruption and having a fairer media coverage. For example, workers demanded to know Zhao Ziyang's golf expenses and the financial sources behind the gambling hobby of Deng Xiaoping's son.[20] Students demanded to have a "fair and faithful news coverage of this patriotic democracy movements" and to "promulgate the press law as soon as possible."[21] Students also suggested that they should "choose the ones [demands] that can be easily accepted by the government as the goals for the present demonstration, so as to achieve a substantial victory and to win the people's understanding and support."[22]

However, as the movement unfolded, the moderate voices and strategic proposals gradually retreated and yielded to radical ones. Consistent with predictions of the escalation mechanism in our model, relatively moderate leaders who proposed compromises (such as evacuating the square and returning to campuses) were removed from their positions or marginalized. They dropped out from key decision-making processes, either voluntarily or involuntarily. The leadership radicalized itself, when only radicals chose to remain and led the movement (Schock 2005, p. 107; Ogden et al. 1992, p. 246). On May 31, Chai Ling, the newly elected "commander-in-chief" and leader of the radicals, announced the

[20] "Ten Questions," April 20, 1989, Beijing Workers' Autonomous Union, translated and collected in Ogden et al. (1992).

[21] "A Peaceful Petition of Seven Demands," April 21, 1989 (Ogden et al. 1992).

[22] "Impeach the Existing Students' Union and Strive for Campus Democracy," undated, Rational and Ardent Youth of Class of 1988, Department of Mathematics, Peking University (Ogden et al. 1992).

updated their agenda with much more impractical demands, including ending the martial law, withdrawal of the army, amnesty to participants of the movement, and a complete end to press censorship (Ogden et al. 1992, p. 247). None of the demands seemed practical or realistic in the context of the political situation of the late 1980s 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 radical leadership did not maintain the appeal to the popular mass: "From May 28 to June 3, the student presence in Tiananmen Square subsided considerably" (Ogden et al. 1992, p. 238).

Our model also predicts that a greater suppression of the mass movement by the regime may radicalize the leadership even further. Consider an increase in cost of joining the leadership, which reflects the harsher punishment imposed by the government. Other things equal, a higher c_l/k_l reduces the leader's premium and causes the marginal leaders to drop out. As x^l increases $(x^l+1)/2$ also increases. In Figure 5, the curve representing $M(x^m)$ shifts up. The fact that $M'(x^m) > 0$ suggests that the adjustment process is cumulative: A more radical leader chooses more radical agendas, which hurt the marginal leaders and cause them to drop out, thus further radicalizing the leadership. The escalation mechanism ensures that a more radical leader will be chosen in the new equilibrium and the agendas chosen in both states will be more radical as well. Figure 5 illustrates such best-response dynamics.

On the one hand, 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. On the other hand, our model shows that the impact of higher c_l/k_l on the selection of leaders is amplified through escalation, which is absent in the aforementioned conjecture.[23] This model prediction is consistent with one popular view about why the leadership of Democratic Movement of China in 1989 further radicalized especially after the Chinese regime imposed martial law. According to Ogden et al. (1992,

[23] Similarly, x^m_* also increases with c_f/k_f. That is because a higher c_f/k_f makes it less likely for citizens to attack the regime, which reduces the probability of success, lowers the leader's premium, and causes $M(x^m)$ to increase.

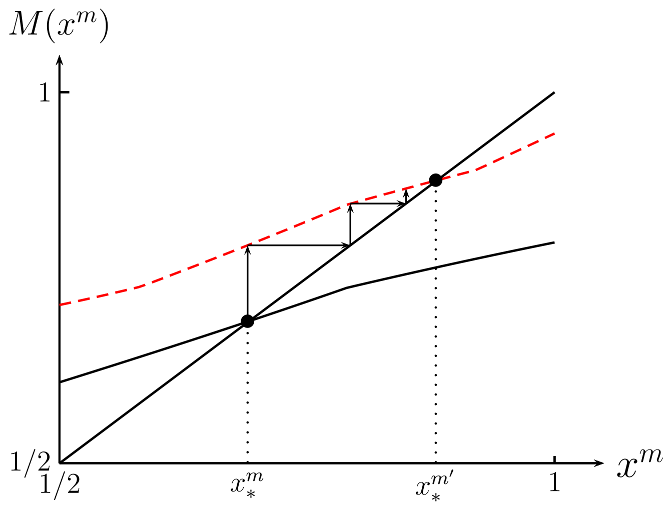


Figure 5. *The solid black curve represents $M(x^m)$, and x_*^m is the fixed point. When the cost of joining the leadership is higher, $M(x^m)$ shifts up and the equilibrium x_*^m increases.*

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."

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 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.

6.2. 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. Since x_*^m increases in c_l/k_l, the equilibrium preference of the median leader will exceed $x^\dagger$ when c_l/k_l becomes sufficiently high (and yet low enough so that the condition specified in Proposition 7 still holds). In Figure 6, we provide a numerical example to demonstrate that there indeed is a non-empty set of parameters that give rise to a unique equilibrium with $x^m_* > x^\dagger$.

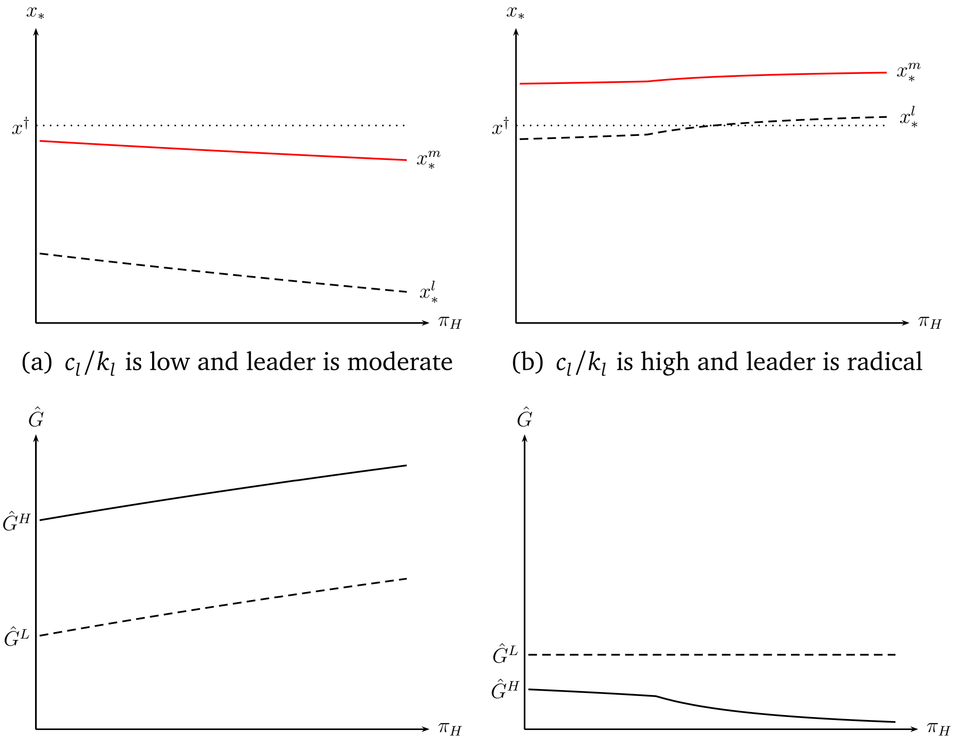
Such an equilibrium is particularly interesting for two reasons. First, it shows that the mechanisms of information asymmetry and endogenous leadership can reinforce each other. Without information asymmetry, the reform agenda in each state (i.e., y_1^{H*} and y_1^{L*}) becomes constant for $x^m > \max\{\hat{x}(\theta_H), \hat{x}(\theta_L)\}$, and therefore $M'(x^m) = 0$ for such x^m. This would eliminate the escalation mechanism described earlier. With information asymmetry, as shown in Proposition 6, $\hat{y}_1^H$ strictly increases in x^m for all $x^m > x^\dagger$. This ensures that $M'(x^m) > 0$, and the escalation mechanism remains in operation throughout.

Second, predictions of the model may be qualitatively different if $x^m_* < x^\dagger$ (with the equilibrium leader choosing his agenda strategically) than if $x^m_* > x^\dagger$ (with the leader "irrationally radicalizing" his agenda to signal the state). The following proposition showcases the contrast between these two cases when we consider the impact of a higher π_H.

Proposition 8. *If in equilibrium we have $x^m_* < x^\dagger$, then an increase in π_H reduces x^m_* and raises the chances of success in both states. If in equilibrium we have (i) $x^m_* > x^\dagger$, (ii) $x^m_* + \theta_L > x^l_* + \theta_H$, and (iii) $x^l_* + \theta_L > \hat{y}^L_{1*}$, then an increase in π_H raises x^m_* and (weakly) reduces the chances of success in both states.*

The formal proof is omitted as it is provided in the following discussion. Suppose the equilibrium leader is moderate (i.e., $x^m_* < x^\dagger$). The adopted agendas in both states will be 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 his expected leader's premium and will encourage more citizens to become leaders, making the leadership group more moderate. In other words, 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. Condition (ii) of Proposition 8 captures the situation that the interests of



(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 equilibrium leader is moderate. In panels (b) and (d), we set $c_l/k_l = 0.2573$, and the equilibrium leader is radical.*

marginal and median leaders are sufficiently far apart. Together with the binding incentive constraint (6), this condition implies:

$$\hat{G}^L\left(\hat{y}_{1*}^L - y_0\right) = \hat{G}^H\left(2x_*^m + 2\theta_L - \hat{y}_{1*}^H - y_0\right) \geq \hat{G}^H\left(2x_*^l + 2\theta_H - \hat{y}_{1*}^H - y_0\right). \qquad (7)$$

Further, condition (iii) implies that $\hat{y}_{1*}^L$ is to the left of the ideal policies of the marginal leader. Therefore, according to the definition of $LP(x_i; y_1, \theta)$ in Section 5.2, the comparison between the left-hand-side and right-hand-side of (7) is equivalent to the comparison between $LP(x_*^l; \hat{y}_{1*}^L, \theta_L)$ and $LP(x_*^l; \hat{y}_{1*}^H, \theta_H)$. In other words, the inequality (7) is equivalent to:

$$LP(x_*^l; \hat{y}_{1*}^L, \theta_L) \geq LP(x_*^l; \hat{y}_{1*}^H, \theta_H).$$

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 and his 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 his 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 (ii) and (iii) specified in Proposition 8 are also satisfied. We show that x^m_* increases in π_H, while the equilibrium success probabilities in the two states weakly decreases.[24] Also observe that, in this example, we have $\hat{G}^H < \hat{G}^L$, which is consistent with the last part of Proposition 5: when the leader is very radical, the probability of success in the high state is lower than that in the low state.

Proposition 8 highlights another difference between a model with exogenous and one with endogenous leadership. With exogenous leadership, the reform agendas chosen (and hence the success probabilities) in the two states are independent of the probabilities π_H and π_L. Endogenous leadership breaks this independence, because the distribution of states affects citizens' decisions to join the leadership or not.

Further, Proposition 8 can also shed some new lights on the debate about the role of structural factors in mass movements. It is intuitive that mass political movements should be easier to succeed when society is riper for a 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.[25] They therefore call into question the so-called structural-roots hypothesis, which holds that the fundamental social, economic, and

[24] In panel (d), $\hat{G}^L$ is constant with respect to π_H because $x^m_* > \hat{x}(\theta_L)$ for all $\pi_H \in (0,1)$ implies that the equilibrium reform agenda $\hat{y}^L_{1*}$ remains unchanged as the equilibrium leader becomes more radical.

[25] 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 revolutionary unrest.

political structures of society are the key determinants of the likelihood of regime change.[26]

Our theory provides a new angle to reconcile the empirical puzzle, considering the subtle role of leadership in regime changes. In our model, the "state" is meant to capture a common factor that affects all citizens' preferences. Given any fixed reform agenda and the status quo (with $y_1 > y_0$), a "high state" is a situation in which the underlying conditions in society are ripe for a change, because the reward from success is increasing in the state. Our benchmark case shows that "root causes" can be important for the outcomes of a regime change game. Proposition 3 demonstrates that when the aggregate grievance is larger (i.e., the discrepancy between the current state θ and status quo policy y_0) and commonly known, the chances of success are indeed larger.

However, our model also shows that such a monotonic relationship is weakened or even reversed when citizens may not know for sure the aggregate state. In Section 4, Proposition 4 demonstrates that when the selected leader happens to be moderate, the aggregate grievance is still the key determinant of the success. But when the leader happens to be radical, he would propose an even more radical policy which cannot be justified by his preference, to credibly signal the state and inform citizens, at a cost of reducing the support. When he is very radical, Proposition 5 shows that the probability of success $\hat{G}^H$ becomes even lower than the probability $\hat{G}^L$ in the low state when he does not need to distort his strategic reform agenda. Likewise, Proposition 6 shows that an increase in π_H can cause $\hat{G}^H$ to fall further. Pooling these cases together, an econometrician who overlooked the subtlety of the leadership's role would conclude that the factors that can trigger political instability are observed more often than actual political upheavals themselves.

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 (in the sense of their preferences for reform) matter. But we further demonstrate that who ends up playing the leadership role is not purely accidental and structural factors can also come into play. Proposition 8 in partic-

[26] 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.

ular highlights the fact that both the cost of joining the leadership c_l/k_l and the likelihood of riper state π_H affect the leaders' preference in equilibrium.

7. 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. First, because the cost of signaling is endogenously determined rather than exogenously imposed, we can study how asymmetric information intertwines with the political agency problem. 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. Second, because there are diverse interests even within the leadership group itself, the reform agendas proposed by a "representative" leader out of signaling concerns may be so radical that it discourages marginal (and moderate) citizens from joining the leadership, thus further radicalizing the "representative" leader and his agendas.

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. Our approach to the analysis of leadership resembles principal-agent theory: the opposition leader of political movements can be interpreted as an "agent" and citizens as "principals." Furthermore, diverse preferences within the leadership group produce further agency problems that can lead to radicalization of the group.

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. A prominent example of a radical tactic is the use of hunger strike or even self-immolation as a form of political protest. Milder forms of protest could in principle communicate the demand for change at a lower cost. But precisely because the cost is low, they may not be sufficiently credible. The willingness to sacrifice through a hunger strike may send a powerful signal to convince the public that the status quo is intolerable and that change is desperately needed.

A limitation of our approach is that the status quo policy is taken as fixed. In reality the regime may adjust its own policies in response to the demands of the opposition. Also absent in our analysis is the possibility of compromise or bargaining between the two sides. We simply assume that the opposition fully commits to implementing its reform agenda, and either the opposition's agenda or status quo prevails, depending on the outcome of the mass protest. Extending the model to allow for these richer possibilities will be an interesting task for future research.

Appendix

Proof of Proposition 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 second part of Assumption 1 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 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 lemma, 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 $F(x^f(y_1;\theta); y_1) = 0$ with respect to y_1 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 in the domain $x^f \in [0, y_1 - \theta]$. 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 $\partial x^f/\partial \theta = -2\partial x^f/\partial y_1 < -1$. ■

Proof of Proposition 2. 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 Proposition 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. ■

Proof of Proposition 3. At a corner solution, $y_1^* = x^m+\theta$ obviously increases with θ. At an interior solution, y_1^* satisfies the first order condition, $1-I(y_1^*;\theta)(y_1^*-y_0) = 0$. Since $I(y_1;\theta)$ decreases in θ by Claim 2, we have $\partial y_1^*/\partial\theta > 0$.

Claim 3. $\partial y_1^*(x^m;\theta)/\partial\theta < 2$.

Proof. The claim is obviously true when $y_1^* = x^m+\theta$. At an interior solution, we have

$y_1^* = \hat{x}(\theta) + \theta$, where $\hat{x}(\theta)$ satisfies:

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

Implicit differentiation of the above shows that $\partial \hat{x}/\partial\theta < 1$, and hence $\partial y_1^*/\partial\theta < 2$. $\square$

Claim 1 and Claim 3 together imply that $x^f(y_1^*(x^m;\theta);\theta)$ decreases in θ. This in turn implies that the probability of success, $G(1 - x^f(y_1^*(x^m;\theta);\theta))$, increases in θ. $\blacksquare$

Proof of Proposition 4. 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 $x^m \leq \min\{\hat{x}(\theta_H), \hat{x}(\theta_L)\}$, 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 3) 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 constitute 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 5. 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.

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:

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

where $\hat{y}_1^L = y_1^{L*}$ depends on x^m. For $x^m > x^\dagger$, Proposition 4 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

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

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, the market 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 4. 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 $x_L^f \geq x_H^f$ (and $G^L \leq G^H$) if and only if $\hat{y}_1^L \geq x^m + 2\theta_L - \theta_H$. We have

$$\begin{aligned}
\Delta_L(x^m; x^m + \theta_H) &= G^H\left(\hat{y}_1^L - x^m - 2\theta_L + \theta_H\right) + (G^L - G^H)(\hat{y}_1^L - y_0) \\
&> G^H\left(\hat{y}_1^L - x^m - 2\theta_L + \theta_H + (\log G^L - \log G^H)(\hat{y}_1^L - y_0)\right) \\
&> G^H\left(\hat{y}_1^L - x^m - 2\theta_L + \theta_H + I^L(x^m + 2\theta_L - \theta_H - \hat{y}_1^L)(\hat{y}_1^L - y_0)\right) \\
&= G^H\left(\hat{y}_1^L - x^m - 2\theta_L + \theta_H\right)\left(1 - I^L(\hat{y}_1^L - y_0)\right),
\end{aligned}$$

where $I^L := I(\hat{y}_1^L; \theta_L)$. If $x^m < \hat{x}(\theta_L)$, $\hat{y}_1^L - x^m - 2\theta_L + \theta_H > 0$ and $1 - I^L(\hat{y}_1^L - y_0) \geq 0$. If $x^m \geq \hat{x}(\theta_L)$, $1 - I^L(\hat{y}_1^L - y_0) = 0$. In either case, $\Delta_L(x^m; x^m + \theta_H) > 0$.

Now, we let x_H^f, G^H, and I^H represent the values of the relevant variables at the point $y_1^H = x^m + \theta_L$. We have

$$\begin{aligned}
\Delta_L(x^m; x^m + \theta_L) &= G^L\left(\hat{y}_1^L - x^m - \theta_L\right) - (G^H - G^L)(x^m + \theta_L - y_0) \\
&< G^L\left(\hat{y}_1^L - x^m - \theta_L + (\log G^L - \log G^H)(x^m + \theta_L - y_0)\right).
\end{aligned}$$

If $\hat{y}_1^L \in [x^m + \theta_L - 2(\theta_H - \theta_L), x^m + \theta_L]$, then $G^L \leq G^H$, and therefore the above expression is negative. If $\hat{y}_1^L < x^m + \theta_L - 2(\theta_H - \theta_L)$, then $G^L > G^H$, and we have

$$\begin{aligned}
\Delta_L(x^m; x^m + \theta_L) &< G^L\left(\hat{y}_1^L - x^m - \theta_L + I^H(x^m + \theta_L - 2(\theta_H - \theta_L) - \hat{y}_1^L)(x^m + \theta_L - y_0)\right) \\
&< -G^H(\theta_H - \theta_L),
\end{aligned}$$

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 $\hat{y}_1^H$ that satisfies the binding incentive constraint $\Delta_L(x^m; \hat{y}_1^H) = 0$ is unique and satisfies $\hat{y}_1^H \in (x^m + \theta_L, x^m + \theta_H)$. □

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.

Finally, from condition (2) that determines the marginal attacker, we have $x^f(\hat{y}_1^L; \theta_L) = x^f(y_1^H; \theta_H)$ if and only if $y_1^H = \hat{y}_1^L + 2(\theta_H - \theta_L)$. Note that this value of y_1^H is greater than

$x^m + \theta_L$ if and only if $x^m < \hat{x}(\theta_L) + 2(\theta_H - \theta_L)$. The value of Δ_L at $y_1^H = \hat{y}_1^L + 2(\theta_H - \theta_L)$ is:

$$\begin{cases} \hat{G}^L(\theta_H - \theta_L) > 0 & \text{if } x^m < \hat{x}(\theta_L), \\ 2\hat{G}^L(\hat{x}(\theta_L) + \theta_H - \theta_L - x^m) & \text{if } x^m \in [\hat{x}(\theta_L), \hat{x}(\theta_L) + 2(\theta_H - \theta_L)], \\ -2\hat{G}^L(\theta_H - \theta_L) < 0 & \text{if } x^m > \hat{x}(\theta_L) + 2(\theta_H - \theta_L). \end{cases}$$

Therefore, $\Delta_L(x^m; \hat{y}_1^L + 2(\theta_H - \theta_L)) < 0$ if and only if $x^m > \hat{x}(\theta_L) + \theta_H - \theta_L$. Because $\Delta_L(x^m; \cdot)$ is increasing in the relevant range, $\Delta_L(x^m; \hat{y}_1^L + 2(\theta_H - \theta_L)) < 0$ implies that $\hat{y}_1^H > \hat{y}_1^L + 2(\theta_H - \theta_L)$. By condition (2), we obtain $x^f(\hat{y}_1^H; \theta_H) > x^f(\hat{y}_1^L; \theta_L)$, and therefore $\hat{G}^H < \hat{G}^L$, if and only if $x^m > \hat{x}(\theta_L) + (\theta_H - \theta_L)$. ■

Proof of Proposition 6. Use implicit differentiation of $\Delta_L(x^m; \hat{y}_1^H(x^m)) = 0$ with respect to x^m to get:

$$\frac{\partial \hat{y}_1^H}{\partial x^m} = \begin{cases} \frac{2\hat{G}^H - \hat{G}^L\left(1 - I^L(x^m + \theta_L - y_0)\right)}{\hat{G}^H\left(1 + I^H(2x^m + 2\theta_L - \hat{y}_1^H - y_0)\right)} > 1 & \text{if } x^m < \hat{x}(\theta_L), \\ \frac{2}{1 + I^H(2x^m + 2\theta_L - \hat{y}_1^H - y_0)} > 0 & \text{if } x^m > \hat{x}(\theta_L). \end{cases}$$

By Proposition 5, $x^m < \hat{x}(\theta_L)$ implies $\hat{G}^H > \hat{G}^L$. Moreover, $x^m < \hat{x}(\theta_L)$ implies $I^L = I(x^m + \theta_L; \theta_L) > I^H = I(\hat{y}_1^H - 2(\theta_H - \theta_L); \theta_L)$, because $\hat{y}_1^H < x^m + \theta_H$ (Claim 4). Therefore, $\partial \hat{y}_1^H / \partial x^m > 1$. Recall that $\partial \hat{y}_1^L / \partial x^m$ is equal to 1 when $x^m < \hat{x}(\theta_L)$ and is equal to 0 when $x^m > \hat{x}(\theta_L)$. This shows that $\hat{y}_1^H - \hat{y}_1^L$ strictly increases in x^m.

For comparative statics respect to θ_H, we have

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

Proof of Proposition 7. We first establish the following result.

Claim 5. $\min\{\hat{LP}(1/2; 1/2), \hat{LP}(1; 1)\} > 0$ 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 5 implies that the mapping M is well-defined on $[1/2, 1]$. It is continuous because $x^m \geq 1/2$ implies that $x^m > y_{\min}(\theta_L) - \theta_L$. 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 6. 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 6, 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 4, 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 $M'(x^m_*) < 1$. $\square$

Claim 6 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. $\blacksquare$

References

Acemoglu, D., G. Egorov, and K. Sonin (2013). "A political theory of populism." *Quarterly Journal of Economics 128*(2), 771–805.

Banks, J. S. and J. Sobel (1987). "Equilibrium selection in signaling games." *Econometrica 55*(3), 647–661.

Boleslavsky, R. and C. Cotton (2015). "Information and extremism in elections." *American Economic Journal: Microeconomics 7*(1), 165–207.

Bolton, P., M. K. Brunnermeier, and L. Veldkamp (2012). "Leadership, coordination, and corporate culture." *Review of Economic Studies 80*(2), 512–537.

Bueno de Mesquita, E. (2010). "Regime change and revolutionary entrepreneurs." *American Political Science Review 104*(3), 446–466.

Cho, I.-K. and D. M. Kreps (1987). "Signaling games and stable equilibria." *The Quarterly Journal of Economics 102*(2), 179–221.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2017). "Voting to tell others." *Review of Economic Studies 84*(1), 143–181.

Dewan, T. and D. P. Myatt (2007). "Leading the party: Coordination, direction, and communication." *American Political Science Review 101*(4), 827–845.

Dewan, T. and D. P. Myatt (2008). "The qualities of leadership: Direction, communication, and obfuscation." *American Political Science Review 102*(3), 351–368.

Dewan, T. and D. P. Myatt (2012). "On the rhetorical strategies of leaders: Speaking clearly, standing back, and stepping down." *Journal of Theoretical Politics 24*(4), 431–460.

Fu, Q., M. Li, and X. Qiao (2018). "On the paradox of mediocracy." Working Paper, Department of Economics, Concordia University.

Geddes, B. (1990). "How the cases you choose affect the answers you get: Selection bias in comparative politics." *Political Analysis 2*, 131–150.

Goldstone, J. A. (2001). “Toward a fourth generation of revolutionary theory.” *Annual Review of Political Science 4*(1), 139–187.

Hermalin, B. E. (1998). “Toward an economic theory of leadership: Leading by example.” *American Economic Review 88*(5), 1188–1206.

Kartik, N. and R. P. McAfee (2007). “Signaling character in electoral competition.” *American Economic Review 97*(3), 852–870.

Landa, D. and S. A. Tyson (2017). “Coercive leadership.” *American Journal of Political Science 61*(3), 559–574.

Lindbeck, A. and J. W. Weibull (1987). “Balanced-budget redistribution as the outcome of political competition.” *Public Choice 52*(3), 273–297.

Majumdar, S. and S. W. Mukand (2010). “The leader as catalyst: on mass movements and the mechanics of institutional change.” Working Paper, Department of Economics, University of Warwick.

Michaeli, M. and D. Spiro (2018). “The dynamics of revolutions.” Working Paper, Department of Economics, University of Oslo.

Morris, S. and M. Shadmehr (2018). “Inspiring Regime Change.” Working Paper, Department of Economics, Princeton University.

Ogden, S., K. Hartford, N. Sullivan, and D. Zweig (1992). *China’s Search for Democracy: The Students and Mass Movement of 1989*. Armonk: M.E. Sharpe.

Schock, K. (2005). *Unarmed Insurrections: People Power Movements in Nondemocracies*. Minneapolis: University of Minnesota Press.

Shadmehr, M. (2015). “Extremism in revolutionary movements.” *Games and Economic Behavior 94*, 97–121.

Shadmehr, M. and D. Bernhardt (2019). “Vanguards in revolution.” *Games and Economic Behavior 115*, 146–166.

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

Wittman, D. (1983). "Candidate motivation: A synthesis of alternative theories." *American Political Science Review 77*(1), 142–157.

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