

# Inspiring Regime Change<sup>1</sup>

Stephen Morris<sup>2</sup>

Mehdi Shadmehr<sup>3</sup>

<sup>1</sup>We are grateful for comments of Nemanja Antic, Dan Bernhardt, Raphael Boleslavsky, Georgy Egorov, Alex Jakobsen, Stefan Krasa, Elliot Lipnowski, Dimitri Migrow, Konstantin Sonin, seminar participants at the 2016 Columbia Conference on Political Economy, 2017 North American Summer Meeting of the Econometric Society, 2017 Wallis Institute Annual Conference on Political Economy, 2018 North American Winter Meeting of the Econometric Society, 2019 Midwest Political Science Association (MPSA) Meeting, 2020 AEA Annual Meeting, University of Chicago, University of Illinois, University of Michigan, University of Central Florida, University of Miami and Queen's University, and the discussions of Brenton Kenkel at MPSA, Mehmet Ekmekci at ASSA, Scott Tyson at the Wallis conference, and Alex Debs at the Columbia conference. We have received excellent research assistance from Elia Sartori. An earlier version of this paper circulated under the title "Reward and Punishment in a Regime Change Game."

<sup>2</sup>Department of Economics, MIT. Email: [semorris@mit.edu](mailto:semorris@mit.edu)

<sup>3</sup>Harris School of Public Policy, University of Chicago, and Department of Economics, University of Calgary. E-mail: [mshadmeh@gmail.com](mailto:mshadmeh@gmail.com)

## Abstract

We conceptualize the process of inspiring regime change, characterize the optimal inspiration strategy, and study its consequences. Drawing from the literature, we formalize the process of inspiring regime change as a mechanism, in which a leader assigns psychological rewards to different anti-regime actions. Citizens face a coordination problem in which each citizen has a private, endogenous degree of optimism about the likelihood of regime change. Because more optimistic citizens are easier to motivate, optimal inspiration entails optimal screening: a mechanism to implicitly parse citizens based on their degree of optimism. This leads to a distribution of anti-regime actions. A key result is the emergence of a vanguard, consisting of citizens who engage in the endogenous, maximum level of anti-regime action. Other citizens participate at varying degrees, with less optimistic citizens contributing less. We show that more heterogeneity (e.g., higher inequality) among potential revolutionaries reduces the likelihood of regime change.

*Keywords:* Regime Change, Inspiration, leadership, Vanguard, Mechanism Design, Global Games

*JEL Classification:* D74, D82, D83

# 1 Introduction

The American Revolutionary Army was at the verge of disintegration in December of 1776 as enlistments were about to expire. Despite appeals by officers, soldiers refused to re-enlist. Then, Washington addressed the troops directly. “My brave fellows,” Washington said, “you have done all I asked you to do, and more than you could be reasonably expected; but your country is at stake, your wives, your houses, and all that you hold dear.... If you will consent to stay only one month longer, you will render that service to the cause of liberty and to your country which you probably never can do under any other circumstances” (McCullough 2005, p. 285). They stayed. This is an episode of what we mean by inspiring regime change: a leader makes a speech, convenes a meeting, writes a treatise or a pamphlet that appeals to, amplifies, or modifies the people’s values, leading the people to contribute to the cause. We develop a model to analyze a leader’s optimal inspiration strategy, and study its consequences.

People can contribute to a movement at various levels; they can make small or large donations, participate in protests at varying frequencies, wear a wristband with a particular color to show solidarity, or engage in armed resistance—Tilly (2008) called these *contentious performances*. A key question for a leader is how to maximize the people’s contribution. To answer that, one must know what motivates people to participate in a movement. The key form of motivation in our context is the psychological rewards that a citizen receives from participating in a movement that succeeds. In particular, we adopt Wood’s (2003) notion of *pleasure in agency*, which refers to psychological benefits corresponding to “the positive effect associated with self-determination, autonomy, self-esteem, efficacy, and pride that come from the successful assertion of intention.” It is a warm glow payoff that one gets from “being part of the making of history.” Different contribution levels correspond to potentially different degrees of pleasure in agency rewards. Where do such contribution-reward mappings come from? From the culture and individual experiences and characteristics. A key part of those experiences and cultural elements is the leaders who influence the link from contributions to pleasure in agency rewards. They do so via a variety of mechanisms from re-framing the nature of conflict to magnifying some values over others (Snow et al. 1986). They have been called transformative, charismatic, or people-oriented leaders, and they appear frequently in major movements (Burns 1978, 2003; Bass 1985; Snow et al. 1986; Goldstone 2001; Ahlquist and Levi 2011). The question of this paper is: What happens when *transformative leaders* manipulate *pleasure in agency* rewards to influence the people’s contributions (*contentious performances*) to maximize the likelihood of *regime change*?

Our paper makes several contributions. (1) We conceptualize and formalize the process of inspiring regime change as a mechanism, in which a leader assigns pleasure in agency rewards to different anti-regime actions. (2) We uncover that optimal inspiration entails optimal screening. The process of inspiring regime change is *like* an exchange: a leader gives pleasure in agency rewards to citizens in exchange for their contributions. Because pleasure in agency rewards are received if the movement succeeds, more optimistic citizens are easier to motivate. Thus, if the leader knew citizens' degrees of optimism, for a given contribution level, she would give more rewards to a more pessimistic citizen than to a more optimistic one. However, citizens' degrees of optimism are their private information. Thus, the leader must design a mechanism to screen for the citizens' beliefs about the chances of success. The trade offs that arise in this screening mechanism reveals the complex strategic nature of optimal inspiration. This merges the seemingly distinct categories of idealist leaders who inspire, and strategic, "task-oriented" leaders who "plot a strategy suitable to resources and circumstances" (Goldstone 2001, p. 157).

(3) We show that, regardless of the leaders' charisma and skills or societal culture, when leaders act optimally and strategically, a vanguard arise naturally. That is, there will be a group of citizens distinct from others in that they engage in the maximum (and endogenous) level of anti-regime activities. This group corresponds to the vanguard in Lenin's treatise, *What Is to be Done?*, Sons of Liberty in the American Revolution, and similar groups that emerge in many other revolutions. Vanguards are more optimistic than other non-vanguard citizens about the chances of success. However, even though they all engage at the same maximal level of action, vanguard citizens are not identical to each other: some of them are more optimistic than others. Non-vanguard citizens contribute at lower levels, with less optimistic citizens contributing less. This generates a distribution of contribution levels in the population that captures the diversity of contentious performances in social movements (Tilly 1978, 2006, 2008; Tarrow 2011).

Extending our model to allow for heterogeneity beyond information (e.g., different exposures to the regime's unjust practices), (4) we show that more heterogeneity within the population of potential contributors reduces the likelihood of regime change and the depth of the revolutionary coalition, defined as the fraction of participating citizens. The distribution of anti-regime activities in the citizenry may have important consequences for the post-revolution regime. For example, if regime change is mainly the result of high effort by a narrow subset of the population, the post-revolution regime is likely to reflect the preferences and ideological convictions of that narrow group. Movements that rely on intense contribution of a small subset of the population

often have to rely on insurgencies (Bueno de Mesquita 2013), and successful movements based on insurgencies often lead to non-democratic outcomes (Wantchékon and García-Ponce 2017). Moreover, the literature typically focuses on heterogeneity (e.g., inequality) in the whole population, showing that more heterogeneity increases the likelihood of instability (Acemoglu and Robinson 2001, 2006; Boix 2003). By contrast, we show that more heterogeneity within potential revolutionaries reduces the likelihood of regime change. (5) Our methodological contribution is to combine mechanism design with global games to analyze a setting in which coordination and screening are intertwined. We disentangle this seemingly intractable problem into two separate and tractable problems of a continuous-action global game and a screening problem.

Our model involves two forms of interactions: a vertical interaction between the leader and citizens, and a horizontal interaction among citizens. The first, vertical form of interaction corresponds to a leader’s decision how to assign psychological rewards to different contribution levels. Through speeches, meetings, and writings, a leader induces these contribution-reward mappings in the citizens’ payoffs. For example, by elevating the values of honor, justice, or patriotism for contributions to the revolution, a leader increases a citizen’s payoff from such contributions. But even very charismatic and skillful leaders can induce only so much contribution by invoking psychological mechanisms. Thus, we require that a leader is constrained to induce up to some maximum level of rewards.<sup>1</sup> The second, horizontal form of interactions corresponds to the coordination aspect of regime change. Citizens’ payoffs from participating in the revolution depend on whether there is a regime change, and a large number of citizens must contribute for a regime change to happen. As Wood (2003) describes, pleasure in agency is “a frequency-based motivation: it depends on the likelihood of success, which in turn increases with the number participating” (p. 235-6). These features imply that citizens care about others’ contributions, making coordination a key strategic element of anti-regime movements. To maximize the likelihood of regime change, a leader must account for citizen coordination that follows her design of pleasure in agency rewards.

We model the horizontal interactions among citizens as a coordination game of regime change, where there is a continuum of ex-ante identical citizens who must decide how much to contribute to the movement. Higher contributions are more costly, capturing both economic opportunity costs and state repression. If the revolution succeeds, participants receive (pleasure

---

<sup>1</sup>This means, for example, that in a religious context, even unlimited comforts and pleasures in heaven in return for some contributions translates into some maximum psychological benefit from those contributions in the faithful’s payoffs that determine their actions.

in agency) rewards that depend on their contribution levels. The revolution succeeds (regime change) if the aggregate contribution exceeds a critical level, which captures, e.g., the regime's strength. We model inspiration by a leader as the leader choosing how contribution levels map into pleasure in agency rewards. There is an upper bound on the level of rewards that the leader can induce for any level of contribution, capturing the limits of the leader's skills as well as other cultural and societal limits to inspiration. We study this problem in a global game model (Carlsson and van Damme 1993) of regime change (Morris and Shin 1998), where citizens observe the regime's strength with private noise and perfect coordination is not possible. In sum, first, the leader induces a contribution-reward mapping. Next, citizens observe their private signals and decide how much to contribute. Then, the success or failure of the revolution is realized and payoffs are enjoyed.

Two forces drive our results. First, there is *strategic uncertainty*: citizens have heterogeneous information and are uncertain about each other's contributions. Because regime change depends on the aggregate contribution of a large number of citizens, strategic uncertainty gives rise to heterogeneous beliefs about the likelihood of regime change. These heterogeneous beliefs arise endogenously because each citizen's beliefs depend on what he anticipates others will do, which, in turn, depend on their own beliefs, and so on. Thus, depending on their information, some citizens will be endogenously more optimistic than others about the likelihood of regime change, and their degree of optimism is their private information. Because pleasure in agency rewards are enjoyed only if the revolution succeeds, more optimistic citizens have higher valuations for a level of reward than citizens who are more pessimistic: for the same contribution level, their expected rewards are higher. Consequently, more optimistic citizens are easier to motivate.

The second force is *screening*. If the leader could condition rewards on a citizen's optimism, then she could fine-tune rewards at the individual level. For example, a more optimistic citizen would get a lower reward for the same contribution than a more pessimistic citizen who assigns a lower chance to receiving the same reward. But a citizen's degree of optimism is his private information and is not available to the leader. If the leader offers the same (maximum) rewards for all contributions, all citizens will choose the minimum contribution, and the likelihood of regime change is *minimized*. Thus, the leader has to generate incentives for higher contributions by giving higher rewards to higher contributions, while accounting for the citizens' strategic responses. Here lies the leader's key tradeoff: If the leader assigns a higher (pleasure in agency) rewards for a given level of contribution, it induces higher contributions from those who would

otherwise have chosen lower levels of contribution. But it also induces lower contributions from those who would otherwise have chosen higher levels of contribution, since they can get the reward at lower cost. The optimal design of pleasure in agency rewards (optimal inspiration) trades off these countervailing effects by designing a mechanism that optimally induces citizens with different and private degrees of (endogenous) optimism to make the contributions and receive the rewards that the leader intended for them.

In essence, inspiring regime change is an exchange, but not of standard tangible material. It is an exchange between a leader in possession of pleasure in agency (psychological) rewards and a large number of citizens with private, heterogeneous valuations for those rewards. Each citizen’s valuation is his degree of optimism, arising endogenously due to strategic uncertainty embedded in the citizens’ coordination interaction. That is, when a leader tries to induce a contribution-reward mapping, it is *as if* she is exchanging rewards for contributions in a population with endogenous and private valuations for these rewards.

The formal literature on citizen participation in protests and revolutions implicitly assumes a form of pleasure in agency rewards from participation in a successful revolution (Bueno de Mesquita 2010; Boix and Svolik 2013; Edmond 2013; Casper and Tyson 2014; Chen and Suen 2016; Tyson and Smith 2018; Shadmehr 2019; Barbera and Jackson 2020). However, these papers do not study the nature and sources of these rewards or the role of leaders in influencing those rewards. Because participation decisions are binary in these settings, if the leader had control over rewards, she would give the maximum reward to those who participate, and no reward to others. A related literature studies the role of leaders in political settings (Dewan and Myatt 2007, 2008; Bueno de Mesquita 2010; Loeper et al. 2014; Acemoglu and Jackson 2015; Dewan and Squintani 2018; Shadmehr and Bernhardt 2019; Chen and Suen 2019), focusing on communication and coordinating effect of leaders, or their agenda setting (DeNardo 1985; Shadmehr 2015)—see Ahlquist and Levi (2011) for a review. In contrast, we formalize and analyze a leader’s role in inspiring anti-regime actions, characterizing the optimal design of pleasure in agency rewards in a continuous action coordination setting.

More broadly, the empirical regularities from qualitative and quantitative literatures on the nature of and motivation for participation in anti-regime movements have largely remained outside the focus of formal models—the next section will discuss these regularities. There is a small literature that integrates those empirical regularities into formal models (Shayo 2009; Siegel 2011; Sambanis and Shayo 2013). Our paper is the first that aims to conceptualize and

formalize the mechanisms involved in inspiring regime change and study their consequences.

The analysis of the substantive aspects of inspiring regime change requires methodological progress over the existing literature, which we will detail after our formal analysis. Our key methodological contribution is to analyze a setting where coordination and screening are intertwined; a screening problem arises because of the endogenous heterogeneity that is introduced by strategic uncertainty. We show that this apparently complex problem can be cleanly reduced to solving a continuous action global game and a screening problem with non-transferable utilities.

We next detail our framework and provide evidence for its key elements and results. Section 3 introduces the model. Section 4 studies a benchmark with exogenous pleasure in agency rewards. In Section 5, we present our formal analysis of inspiring regime change. Proofs are in an online appendix.

## 2 Participation, Motivation, and Inspiring Leadership

Our analysis is built on three fundamental elements: the nature of participation in regime change, the nature of motivations for participation, and the role of leaders in influencing motivations. Citizens can contribute at various levels to a movement. Thus, a citizen’s contribution level can be represented by a number  $e \geq 0$ . A citizen who contributes at a level  $e$  pays a cost of  $C(e)$ , which captures state repression and other opportunity costs of anti-regime actions. A citizen’s motivation for contribution is the psychological rewards that he receives if the movement succeeds. These pleasure in agency rewards are represented by a function  $B(e)$ , which maps each contribution level  $e$  to a level of pleasure in agency rewards. Inspiring regime change corresponds to influencing this mapping by leaders. In particular, the leader designs/chooses the contribution-reward mapping,  $B(e)$ , subject to the constraint that  $0 \leq B(e) \leq M$ , for some number  $M > 0$ . The parameter  $M$  reflects the leader’s skills and the society’s culture and experiences, including available “cultural idioms” (Skocpol 1997) or “repertoire of common symbolics” (Dabashi 1993) that facilitate the leader’s task of creating, activating, and manipulating motivations for anti-regime actions. For example, a more skillful or charismatic leader can induce higher levels of pleasure agency rewards, corresponding to a higher  $M$ . The leader’s constraint also reflects the psychological nature of pleasure in agency rewards. Unlike material rewards, pleasure in agency rewards are non-rival: that some citizens receive these rewards does not subtract from the rewards that others could receive. This section expands on these elements,



and provides evidence for this framework and for the results of our analysis that is built on it.

## 2.1 Participation

What is the nature of participation in anti-regime social movements? Citizens can participate at varying degrees. For example, they can make different monetary contributions, donate their goods or expertise, make their homes available for meetings, join demonstrations with different frequencies, participate in boycotts, or take up arms “for a few weeks” or “only a few days,” as Washington asked colonists in the American Revolution (Washington 1890, v. VI, p. 213-4). *Contentious Performances*, as Tilly (2008) dubbed them, take various forms, including arson (Swing Rebellion), cattle maiming (Tithe War), sit-ins in foreign embassies (Iranian Constitutional Revolution), cacerolazo (1989 Venezuelan Caracazo), shouting slogans from roof-tops at night (1979 Iranian Revolution), occupations of government buildings, wearing particular colors (green wristbands in the Iranian Green Movement, orange ribbons in the Ukrainian Orange Revolution, or yellow ribbons in the Yellow Revolution in the Philippines), boycotts (American Revolution), parades (Woman Suffrage Movement), strikes, marches, demonstrations, freedom rides, street blockades, suicide bombings, assassinations, hijackings, and guerrilla wars.

Some contributions require more effort and are more costly than others, e.g., taking up arms versus participating in a demonstration versus wearing a particular color to show solidarity. That is, different contentious performances correspond to different levels of anti-regime contributions. Indeed, Tilly’s celebrated notion of the repertoires of contentious performances (Tilly 1978, 2006, 2008; McAdam 1999; Tarrow 2011) presupposes a variety of anti-regime actions with various levels of participation and contributions. Thus, a fundamental question for an opposition leader is: what kinds of contentious performances can be elicited and how? The answer depends on the nature of the people’s motivations to make those contributions.

## 2.2 Motivation

What motivates citizens to contribute to anti-regime movements, where individual effects are minimal, risks are high, and the material benefits of success are public goods? This question is the essence of Tullock’s (1971) “Paradox of Revolution,” which specialized Olson’s (1965) *Logic of Collective Action* to revolution settings. One possible answer is that participants somehow receive *selective* material rewards. Indeed, evidence suggests selective material incentives were a key factor in the decision of many individuals to join rebel forces in some civil wars, e.g.,

Sierra Leone (Weinstein 2007; Humphreys and Weinstein 2008)—see also Popkin (1979) and Ross (2006).<sup>2</sup> However, as Wood (2003) documents for El Salvador, psychological rewards are the key motivation in other civil wars. As Blattman and Miguel (2010) summarize in their review of the civil war literature, “non-material incentives are thought to be common within armed groups. Several studies argue that a leader’s charisma, group ideology, or a citizen’s satisfaction in pursuit of justice (or vengeance) can also help solve the problem of collective action in rebellion” (p. 15).

Growing evidence persistently indicates that psychological rewards were the primary motivation in a variety of anti-regime movements in Eastern Europe (Petersen 2001), El Salvador (Wood 2003), Morocco (Lawrence 2017), Iran (Shadmehr 2017); Syria (Pearlman 2018), Turkey and Ukraine (Aytaç and Stokes 2019). Material motivations may have been the primary form of motivation in some civil wars, but evidence suggests that logic does not extend to less militarized movements. Hundreds of thousands who protested against election fraud in the Green Movement that followed the contested 2009 Iranian presidential election could not have expected to receive selective material rewards for their participation. Moments before heading to the streets on June 20, 2011, when a protester was asked why she wanted to participate, she responded: “because they get their legitimacy from us by our votes.... I want to, at least, get that legitimacy back from them” (personal communication). In the American Revolution, countless sermons, speeches, letters, and pamphlets invoked cultural values of defending liberty, honor, patriotism, and Christian duties to motivate participation. Thomas Paine’s *The American Crisis* (“These are the times that tries men’s souls.”), Abraham Keteltas’s 1777 sermon, “God pleads his cause” (Sandoz 1998, p. 579-605; see also Breen 2010), and Washington’s speeches to the troops and his letters to the colonies (Washington 1890, vol. VI, p. 213-4, 282-3) are but a few examples.<sup>3</sup>

What is the nature of psychological incentives? Early works on protests and revolutions emphasized expressive motives: some underlying reason causes psychological stress, and protest

---

<sup>2</sup>However, even in those civil wars, the literature distinguishes between opportunists, who participate in anticipation of immediate material rewards, and more committed activists (Weinstein 2007). Moreover, in the context of civil war, receiving material rewards in exchange for joining the rebels may not indicate the primacy of material incentives; individuals often have to give up their economic activities that generate income for their families, and the material rewards may only partially compensate for the foregone income. Further, civil wars with rebel armies are a particular form of conflict that are more prone to rely on mercenary-like recruitment and lootable resources—because arms and equipments are expensive. These features, which are more conducive to generate selective material rewards, are less available in less militarized conflicts.

<sup>3</sup>Such motivations may seem implausible to some today. However, as Middlekauff (2005) puts it in his history of the American Revolution, “Courage, honor, gallantry in service of liberty, all those words calculated to bring a blush of embarrassment to jaded twentieth-century men, defined manhood for the eighteenth century” (p. 515).

generates a cathartic release of those psychological tensions (Davies 1962; Smelser 1962; Geschwender 1967; Gurr 1970). As McAdam (1999) puts it, in such theories, “the social movement is effective not as political action but as therapy” (p. 10). Later studies, pioneered by Tilly, showed that when deciding whether to join a movement, individuals account for the likelihood of success and the costs of participation (Tilly 1978, 2008; McAdam 1999; Tarrow 2011). In particular, movements with no prospect of success do not sustain because participation costs exceed the psychological benefits of participating in a failed movement. We call this the *Lost Cause Principle*. In Washington’s words, “The honor of making a brave defense does not seem to be a sufficient stimulus, when the success is very doubtful and the falling into the Enemy’s hands probable” (Middlekauff 2005, p. 342). Indeed, Washington’s victory in the battle of Trenton after a series of defeats improved recruitment for the revolution by improving the people’s beliefs that the revolution could succeed; as the loyalist Nicholas Cresswell recorded in his diary in 1777, after the battle of Trenton: “The minds of the people are much altered. A few days ago they had given up the cause for lost. Their late successes have turned the scale and now they are all liberty mad again” (Rhodehamel 2001, p. 264).

The Lost Cause Principle and the Olsonian Logic of Collective Action impose restrictions on the nature of psychological rewards in social movements. The Lost Cause Principle implies that expressive motives are not enough for sustained actions in support of lost causes. The Olsonian Logic of Collective Action suggests that individual action cannot be based on the implausible belief that a typical individual’s contributions make a non-negligible difference in the outcome. Wood’s (2003) notion of pleasure in agency is consistent with both these requirements. Wood (2003) develops the notion of pleasure in agency to capture individuals’ motivations for participating in contentious collective actions. Pleasure in agency is “the positive effect associated with self-determination, autonomy, self-esteem, efficacy, and pride that come from the successful assertion of intention” (p. 235). It is “a frequency-based motivation: it depends on the likelihood of success, which in turn increases with the number participating (Schelling 1978; Hardin 1982). Yet the pleasure in agency is undiminished by the fact that one’s own contribution to the likelihood of victory is vanishingly small” (p. 235-6). For example, in her interviews with insurgents during and after the El Salvadoran civil war, she found that insurgents “repeatedly asserted their pride in their wartime activities and consistently claimed *authorship* of the changes that they identified as their work” (p. 231). “They took pride, indeed pleasure, in the successful assertion of their interests and identity...motivated in part by the value they put

on being part of the making of history” (p. 18-9). Other researchers, too, have pointed to the same kind of psychological rewards (Petersen 2001; Pearlman 2018). For example, Pearlman (2018) identifies the role of self-respect and “joy of agency” in recent Syrian protests.

The content of inspiring speeches and writings in social movements are remarkably consistent with the key attributes of pleasure in agency rewards as psychological rewards that one receives for contributing to a movement that succeeds. They have a dual emphasis: (1) they appeal to the people’s values (e.g., justice, honor, patriotism, faith, courage), and (2) they insist that their movement has good chances of success. For example, in the American Revolution, the statements of Washington, Paine, and Keteltas which aimed to inspire contribution (see above), all alternate between highlighting cultural values and arguing that the revolution can succeed. For example, Keteltas who argued in his sermon that “the cause of this American continent...is the cause of God,” also adamantly argued that the revolution will succeed, because “God will effectually plead it; he will plead it by his almighty word, his all conquering spirit, and his over ruling providence,” all of which he interprets in detail based on the Bible and Christian theology. The title of his sermon, “God Pleads His Cause,” is a reflection of this dual emphasis (Sandoz 1998, p. 579-605). A similar dual emphasis also appears in the very different context of the development of Islam in its early years. Muhammad needed believers to contribute to the movement with their money, bodies and weapons. Many Quranic verses aim to get people to make such contributions. Quran repeatedly tells the faithful that success is near, and that God will help them in many hidden ways, e.g., by sending invisible warrior angels or by instilling fear in the hearts of their enemies to weaken their resolve to fight (3:121-7). Thus, like Washington, Paine, and Keteltas, Muhammad had to emphasize that the movement will succeed. Just as in the American Revolution, “the honor of making a brave defense does not seem to be a sufficient stimulus, when the success is very doubtful,” so too, in the Islamic Movement, appealing to the people’s values of justice or God’s love, the promise of heavenly gardens, or the threat of hellish fires do not alone inspire action; leaders have to convince people that their movement can succeed—the Lost Cause Principle holds.

### **2.3 Leadership: Inspiring Regime Change**

Where do pleasure in agency rewards come from? They come from the culture and individual experiences and characteristics, including past interactions with the regime in periods of contention (Gecas 2000; Wood 2003; Pearlman 2018). A key part of these cultures and experiences

is the leaders who frame and interpret events, and organize cultural elements to inspire action. An extensive leadership literature that cuts across the disciplines of political science, sociology, and management argue that some individuals have the abilities, skills, or the charisma to inspire costly actions in others (Burns 1978, 2003; Bass 1985; Snow et al. 1986; Goldstone 2001; Ahlquist and Levi 2011). These leaders have been called transformational, transformative, charismatic, or people-oriented leaders. Goldstone (2001) identifies two distinct types of revolutionary leaders: people-oriented and task-oriented. “People-oriented leaders are those who inspire people, give them a sense of identity and power, and provide a vision of a new and just order” (p. 157). These are leaders who can activate, amplify, or transform people’s feelings and identities, e.g., by innovative framing of events and experiences or by their personality traits, such as their charisma (Snow et al. 1986). In the terminology of organizational behavior, they are “transformational leaders” who have the ability to create inspirational motivations through a variety of psychological mechanisms (Burns 1978, 2003; Bass 1985). Skillful leaders can create, activate, amplify or manipulate pleasure in agency rewards by tapping into the people’s values, emotions, and history. We call this process inspiring regime change. It corresponds to the leader influencing the pleasure in agency rewards  $B(e)$  associated with a level of anti-regime action  $e$ , subject to cultural restrictions and opportunities as well as the leader’s skills and abilities that put an upper bound on the magnitude of pleasure in agency rewards.

In contrast to people-oriented leaders, “task-oriented leaders are those who can plot a strategy suitable to resources and circumstances” (Goldstone 2001, p. 157), including the repressive capacity of the state and available revolutionary skills, to effectively transform anti-regime sentiments into concrete actions. Our analysis shows how a “people-oriented” leader can optimally activate, amplify, and manipulate pleasure in agency rewards to maximize the likelihood of regime change. That even the design of psychological pleasure in agency rewards requires “a strategy suitable to resources and circumstances” merges these seemingly separate categories of people-oriented and task-oriented leadership.

We show the optimal design of pleasure in agency rewards will assign higher pleasure in agency rewards to higher contributions (Proposition 2 (Optimal Inspiration)). Indeed, in the context of the American Revolution, by invoking the people’s sense of honor and patriotism, Washington aimed to assign different rewards to different levels of participation: those who would join the army for a few weeks would be more honorable and better patriots than those who would join for a few days. Short of that, people could contribute their cattle. Even those

who would refuse, could still contribute simply by refusing to sell to the enemy (Washington 1890, vol. VI, p. 213-4, 382-3).

We argued that inspiring regime change can be conceptualized as an exchange between a leader in possession of pleasure in agency (psychological) rewards and a large number of citizens with private, heterogeneous valuations for those rewards. When a leader tries to induce a contribution-reward mapping, it is *as if* she is exchanging rewards for contributions in a population with endogenous and private valuations for these rewards. In the context of the early Islamic movement, Quran uses the language of exchange to describe the the nature of rewards designed to motivate contributions: “God has purchased from the believers their lives and their properties in exchange for Paradise” (9:111). Moreover, consistent with optimal design of pleasure in agency rewards (Proposition 2) and similar to above observations in the context of the American Revolution, higher and more costly contributions would yield higher rewards: “Not equal among you are those who contributed before the conquest, and fought. Those are higher in rank than those who contributed afterwards, and fought. But God promises both a good reward” (57:10). Indeed, rewards increase even with marginal contributions: “Nor do they spend any expenditure, small or large, nor do they cross any valley, but it is recorded to their credit” (9:121). “Not equal are the inactive among the believers—except the disabled—and those who strive in the cause of God with their possessions and their persons. God prefers those who strive with their possessions and their persons above the inactive, by a degree. But God has promised goodness to both. Yet God favors the strivers, over the inactive, with a great reward” (4:95; see also 9:19-22).

If leaders knew the citizens’ beliefs about the likelihood of success, they would condition rewards on those beliefs. Not knowing the beliefs, however, the leaders assign different rewards to different participation levels. But this mechanism makes the leaders’ optimal motivation problem a screening problems, with the accompanying tradeoffs, which we described in the Introduction. For example, raising the value of joining the cause as a militiaman for “only a few days” increases the incentives of those who would not have otherwise joined the militia; but it also reduces the incentive of those who would have otherwise stayed “for a few weeks.”

Remarkably, the optimal design of pleasure in agency rewards leads to the emergence of a group of citizens who engage in the maximum level of anti-regime activities, regardless of the skills and charisma of the leader or other constraints that bound pleasure in agency rewards by  $M$  (Proposition 2 (Optimal Inspiration)). In the American Revolution, this group corresponds

to the Sons of Liberty (Maier 1972), and during the Revolutionary War, perhaps a subset of regulars who remained with the Continental Army during the toughest times. Middlekauff's (2005) description is telling: "By winter 1779-1780 the Continentals were beginning to believe that they had no one save themselves to lean on.... Dissatisfaction in these months slowly turned into a feeling of martyrdom. They felt themselves to be the martyrs to the 'glorious cause' " (p. 513-4).

More broadly, the emergence of this cadre is consistent with the notion of professional revolutionaries in Lenin's treatise, *What Is to be Done?*. Although this group is distinct from other citizens contributing to the cause, there is little difference in the magnitude of duties and the level of revolutionary activity among the members of the group. Contrasting the organization of workers, who engage in various degrees of contentious activity, with the organization of "revolutionary social-democratic party," Lenin insisted that "the organization of the revolutionaries must consist first and foremost of people who make revolutionary activity their profession.... In view of this common characteristic of the members of such an organization, all distinctions as between workers and intellectuals, not to speak of distinctions of trade and profession, in both categories, must be effaced" (p. 71). Of course, the members of this group are not identical; some have stronger beliefs in the likelihood of regime change than others, for example, as a result of their different interpretations of Marxist theory. However, despite these differences in their beliefs, they constitute a distinct group that exert maximum effort in service of the revolution.

### 3 Model

There is a continuum of citizens, indexed by  $i \in [0, 1]$ , who must decide how much effort  $e \geq 0$  to contribute to regime change. Exerting effort  $e$  costs  $C(e)$ , independent of whether there is regime change, and gives a benefit  $B(e)$  only if there is regime change. Thus, if a citizen believes that the regime change occurs with probability  $p$ , his expected payoff from choosing effort  $e$  is  $pB(e) - C(e)$ . Here,  $B(e)$  is the pleasure in agency rewards for contribution  $e$ , and  $C(e)$  is the cost of contributing  $e$ , including economic opportunity costs as well as state repression. We assume  $C(e)$  is strictly increasing and twice continuously differentiable with  $C(0) = 0$ . Regime change occurs if and only if the total contribution  $\int e_i di$  is greater than or equal to the regime's strength  $\theta \in \mathbb{R}$ . The regime's strength  $\theta$  is uncertain, and citizens have an improper common prior that  $\theta$  is distributed uniformly on  $\mathbb{R}$ . Moreover, each citizen  $i \in [0, 1]$  receives a noisy pri-

vate signal  $x_i = \theta + \nu_i$  about  $\theta$ , where  $\theta$  and the  $\nu_i$ s are independently distributed with  $\nu_i \sim f(\cdot)$ .<sup>4</sup>

There is also a leader who aims to maximize the likelihood of regime change. The leader chooses the (pleasure in agency) reward function  $B(e)$  subject to the constraint that  $B(e) \in [0, M]$ , with  $M > 0$ . The leader's choice of a contribution-reward mapping,  $B(e)$ , corresponds to inspiring regime change. Parameter  $M$  captures the limits of pleasure in agency rewards in inducing action. It depends on cultural and economic aspects of the society as well as the charisma and skills of the leader.

The game proceeds as follows. The leader chooses pleasure in agency rewards  $B(e)$  subject to her constraint. Then, the regime's strength is realized. Next, citizens observe their private signals, and simultaneously decide how much to contribute. The success or failure of the revolution is realized and payoffs are received.

## 4 Regime Change with Exogenous Pleasure in Agency

We begin our analysis with a benchmark in which pleasure in agency benefits from participation are exogenously given. In the next section we endogenize these rewards by allowing the leader to choose these rewards subject to her constraints.

### 4.1 A Key Statistical Property: Uniform Threshold Belief

It is useful to first introduce a key statistical property that we call *uniform threshold belief*. Fix any level of regime strength  $\theta = \hat{\theta}$ . A citizen's *threshold belief* about  $\hat{\theta}$  is the probability that he assigns to the event that the regime is weaker:  $\theta \leq \hat{\theta}$ . Any level of regime's strength  $\theta$  gives rise to a distribution of signals ( $x_i$ s), and hence a distribution of threshold beliefs about  $\hat{\theta}$ . The *threshold belief distribution* at  $\theta = \hat{\theta}$  is the distribution of threshold beliefs about  $\hat{\theta}$  in the population when the regime's strength is also  $\hat{\theta}$ . We will show that this threshold belief distribution is always uniform, regardless of the value of  $\hat{\theta}$  and the distribution of noise.

To state the uniform threshold belief property formally, write  $H(\cdot|\hat{\theta})$  for the cdf of the threshold belief distribution at  $\hat{\theta}$ . Thus, the proportion of citizens with threshold belief about  $\hat{\theta}$

---

<sup>4</sup>The assumption of an improper common prior is standard in this literature. At the cost of further notation, one can replace this assumption with a prior that is uniformly distributed on a sufficiently large bounded interval. More importantly, results proved under the uniform prior assumption can also be reproduced with an arbitrary smooth prior when the noise is sufficiently small: see Carlsson and van Damme (1993), Morris and Shin (2003) and Frankel et al. (2003).



below  $p$  when the true state is  $\hat{\theta}$  is  $H(p|\hat{\theta})$ . The key observation is that, because  $\theta$  is distributed uniformly so that there is no prior information about  $\theta$ , one can consider  $\theta$  as a signal of  $x_i$ , writing  $\theta = x_i - \nu_i$ . So the threshold belief about  $\hat{\theta}$  for a citizen observing  $x_i$  is simply the probability that  $x_i - \nu_i \leq \hat{\theta}$ , or  $1 - F(x_i - \hat{\theta})$ . Thus, a citizen observing  $x_i$  has a threshold belief weakly less than  $p$  if  $1 - F(x_i - \hat{\theta}) \leq p$ , or  $x_i \geq \hat{\theta} + F^{-1}(1 - p)$ . Now, if the true state is  $\hat{\theta}$ , the proportion of citizens observing a signal weakly larger than  $\hat{x}$  will be the proportion of citizens with  $\hat{\theta} + \nu_i \geq \hat{x}$ , which is  $1 - F(\hat{x} - \hat{\theta})$ . Combining these observations, we have that  $H(p|\hat{\theta}) = 1 - F(\hat{x} - \hat{\theta})$  where  $\hat{x} = \hat{\theta} + F^{-1}(1 - p)$ . Substituting for  $\hat{x}$ , we have

$$H(p|\hat{\theta}) = 1 - F(\hat{x} - \hat{\theta}) = 1 - F(F^{-1}(1 - p)) = 1 - (1 - p) = p.$$

We have shown the uniform threshold belief property, which also appears in Guimaraes and Morris (2007):

**Lemma 1.** *The threshold beliefs distribution is always uniform on  $[0, 1]$ , so that  $H(p|\hat{\theta}) = p$  for all  $p \in [0, 1]$  and  $\hat{\theta} \in \mathbb{R}$ .*

An intuition for this surprising result is the following. At any given threshold state and for any given citizen, we can identify the citizen's rank in the threshold belief distribution, i.e., the proportion of citizens with higher signals. This rank is necessarily uniformly distributed: since a citizen does not know his rank, his belief about whether the true state is below the threshold state is uniform. More formally, suppose that noise itself was uniformly distributed on the interval  $[0, 1]$ . If the true state was  $\hat{\theta}$ , a citizen observing  $x_i$  in the interval  $[\hat{\theta}, \hat{\theta} + 1]$  would have threshold belief  $\hat{\theta} + 1 - x_i$ , and we would have uniform threshold belief. But if the noise had some arbitrary distribution, we could do a change of variable, replacing the level of a citizen's signal with its percentile in the distribution. Because citizens do not know their own percentiles, the same argument then holds.

## 4.2 Equilibrium

If a citizen assigned a probability  $p$  that regime change succeeds, he chooses his contribution level  $e$  to maximize his expected payoff:

$$\max_{e \geq 0} pB(e) - C(e).$$

Generally, many contribution levels may be optimal. For example,  $B(e)$  and  $C(e)$  may change in a way that the citizen remains indifferent between  $e_1 < e_2$ , so that  $pB(e_1) - C(e_1) =$

$pB(e_2) - C(e_2)$ . Because we aim to consider the leader's design of benefits, we cannot put restrictions on the benefit function to rule this out. Indeed, as we will show in the next section, the leader's design of optimal  $B(e)$  will have strictly *convex* segments.

**Definition 1.** The *optimal effort correspondence* is

$$e^*(p) = \arg \max_{e \geq 0} pB(e) - C(e);$$

we write

$$e_{\min} = \min(e^*(0)) \text{ and } e_{\max} = \max(e^*(1)).$$

We maintain the assumption that the maximum and minimum exist (and are finite).

**Definition 2.** The optimal effort correspondence is *weakly increasing* if  $p_2 > p_1$  and  $e_i \in e^*(p_i)$ ,  $i \in \{1, 2\}$ , imply  $e_2 \geq e_1$ .

That is, a citizen who is strictly more optimistic about the likelihood of success than another will put in at least as much effort as him.

Note that because  $e^*(p)$  is weakly increasing, it is almost everywhere single-valued. A strategy for a citizen  $i$  is a mapping  $s_i : \mathbb{R} \rightarrow \mathbb{R}_+$ , where  $s_i(x_i)$  is the effort level of citizen  $i$  when he observes signal  $x_i$ . We focus on weakly decreasing strategies. Each strategy profile  $(s_i)_{i \in [0,1]}$  will give rise to aggregate behavior

$$\hat{s}(\theta) = \int_{i=0}^1 \left( \int_{\nu_i=-\infty}^{\infty} s_i(\theta + \nu_i) f(\nu_i) d\nu_i \right) di.$$

If all citizens follow weakly decreasing strategies, then  $\hat{s}$  is weakly decreasing, and there is a unique threshold  $\theta^*$  such that  $\hat{s}(\theta^*) = \theta^*$ . Then,  $\hat{s}(\theta) \geq \theta$  and there will be a regime change if and only if  $\theta \leq \theta^*$ . Thus, a citizen observing a signal  $x_i$  will assign probability  $G(\theta^*|x_i)$  to the event that  $\theta \leq \theta^*$  and thus to regime change. We conclude that each citizen must be following the strategy with

$$s^*(x_i) = e^*(G(\theta^*|x_i)).$$

Letting  $p = G(\theta^*|x_i)$ , and recalling that  $H(p|\theta^*)$  is the cdf of  $G(\theta^*|x_i)$  conditional on  $\theta^*$ , this implies that the aggregate effort of citizens when the true state is  $\theta^*$  is

$$\hat{s}(\theta^*) = \int_{p=0}^1 e^*(p) dH(p|\theta^*).$$

By Lemma 1, we know that  $H(p|\theta^*)$  is a uniform distribution. We also know that  $\theta^* = \hat{s}(\theta^*)$ . Thus,

$$\theta^* = \hat{s}(\theta^*) = \int_{p=0}^1 e^*(p) dp.$$

Summarizing this analysis, we have:

**Proposition 1.** *If the optimal effort correspondence  $e^*$  is weakly increasing, then there is a unique monotonic equilibrium where all citizens follow the essentially unique common strategy*

$$s^*(x_i) = e^*(1 - F(x_i - \theta_{e^*})). \quad (1)$$

*In this equilibrium, there is regime change whenever  $\theta \leq \theta_{e^*}$ , where the regime change threshold  $\theta_{e^*}$  is given by*

$$\theta_{e^*} = \int_{p=0}^1 e^*(p) dp. \quad (2)$$

Continuous action global games have been studied by Frankel et al. (2003) in an abstract setting, and by Guimaraes and Morris (2007) in a model of currency attacks. They proved the stronger result that the corresponding equilibrium was unique among all equilibria, but under assumptions implying supermodular payoffs. The above proposition holds under weaker assumptions, requiring only that  $e^*(p)$  is increasing in  $p$ . We have proved uniqueness within the class of monotonic equilibria, but leave open the question of whether non-monotonic equilibria exist.

Given that we do not impose any assumption on  $B(e)$ , it remains to show whether our assumptions about costs are enough to deliver that  $e^*(p)$  is increasing in  $p$ . Lemma 2 shows that our assumption that  $C(e)$  is strictly increasing suffices.

**Lemma 2.** *If  $C(e)$  is strictly increasing in  $e$ , then any selection from optimal effort correspondence is weakly increasing.*

We conclude this section by providing a standard class of examples. Suppose that the exogenous reward scheme is  $B(e) = \sqrt{e}$  and the cost scheme is  $C(e) = e^n$ ,  $n \geq 1$ . In this case,  $e^*(p) = (\frac{p}{2n})^{\frac{2}{2n-1}}$  and

$$\theta_{e^*} = \frac{2n-1}{2n+1} (2n)^{\frac{2}{1-2n}}.$$

Figure 1 illustrates. Proposition 1 shows that the equilibrium regime change threshold  $\theta_{e^*}$  is simply the area under the curve  $e^*(p)$  in the right panel. Here, the pleasure in agency mapping  $B(e) = \sqrt{e}$  was exogenous, and its relevant part did not exceed 1. We will show that when the leader optimally chooses pleasure in agency rewards subject to the same maximum of 1, the pleasure in agency mapping will take a very different form and the area under the corresponding  $e^*(p)$  (which determines the equilibrium regime change threshold) will be larger—see Figure 2.

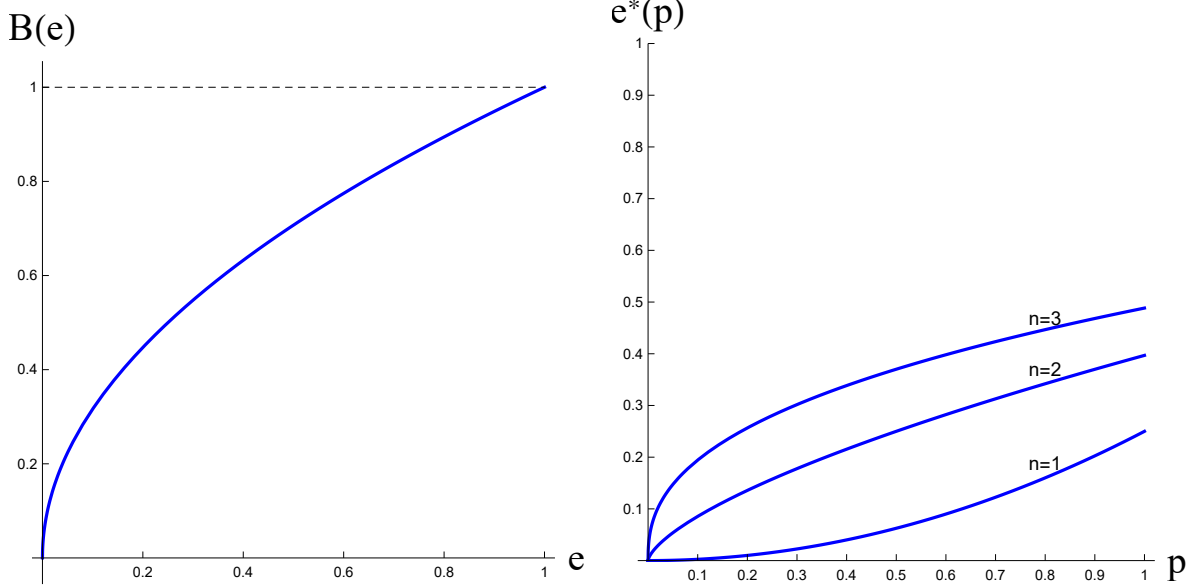


Figure 1: Exogenous pleasure in agency rewards:  $B(e) = \sqrt{e}$  and  $C(e) = e^n$ ,  $n \geq 1$ .

We will use Proposition 1 as an ingredient in our analysis of inspiring regime change in the next section. However, it is worth highlighting its broader applications here. The applications of global games in regime change have been limited to binary actions. Binary action models capture some key aspects of coordination. However, they are restrictive in other aspects, including the study of different repression schemes, violent versus non-violent protest tactics, or the repertoires of contentious performances (Tilly 1978, 2006, 2008; Morris and Shadmehr 2018). Proposition 1 facilitates such studies by providing a general and clean characterization of the equilibrium in continuous actions models of regime change.

## 5 Inspiring Regime Change: The Design of Pleasure in Agency Rewards

We now investigate the optimal design of pleasure in agency rewards,  $B(e)$ , by a leader who aims to maximize the likelihood of regime change. Charismatic leaders can inspire citizen participation by assigning psychological rewards to different levels of anti-regime activities. Through speeches, writings, and meetings, they activate, create, and manipulate intrinsic motivations by creating identities and innovative framing of events, and by “identification, idealization, and elevation of one or more values presumed basic to prospective constituents” (Snow et al. 1986, p. 469). Unlike material benefits, these psychological rewards are non-rival. For example, in a religious context, if a fighter tortured in the righteous struggle to bring down an unjust state or a wicked ruler is to receive one castle in heaven, God can build as many castles as there are deserving fighters. Still, even charismatic leaders can incite only so much intrinsic motivation in their potential followers. Thus, we require that  $B(e) \leq M$  for some exogenous  $M > 0$ . The upper bound on rewards reflects limitations on the leader’s skills and charisma, as well as other exogenous aspects of the environment, such as “cultural idioms” (Skocpol 1997) or “repertoire of common symbolics” (Dabashi 1993), that facilitate the leader’s task of creating, activating, and manipulating intrinsic motivations for revolutionary activities.

The combination of Proposition 1 and Lemma 2 show that, as long as  $C(e)$  is strictly increasing, for any choice of  $B(e)$  there will be a unique equilibrium in which regime change occurs when  $\theta \leq \theta_{e^*} = \int_{p=0}^1 e^*(p) dp$ . Thus the leader’s optimization problem reduces to choosing reward scheme  $B(e)$  to maximize the regime change threshold

$$\int_{p=0}^1 e^*(p) dp \tag{3}$$

subject to the optimality of the effort function

$$e^*(p) \in \arg \max_{e \geq 0} p B(e) - C(e),$$

and the feasibility of the reward scheme

$$B(e) \in [0, M] \text{ for all } e. \tag{4}$$

A reward scheme  $B^*$  that maximizes this problem is an *optimal reward scheme* (there will be some indeterminacy in optimal reward schemes). An effort function  $e^*$  arising in this problem

is an *optimal effort function* (it will be essentially unique). The regime change threshold  $\theta^*$  that results from the optimal effort function is the (essentially unique) *optimal regime change threshold*.

**Proposition 2. (*Optimal Inspiration*)** Suppose costs are strictly increasing and convex.

- The optimal pleasure in agency rewards function  $B^*$  is continuous, strictly increasing and strictly convex on the interval  $[0, \bar{e}]$  with  $B^*(0) = 0$  and  $B^*(\bar{e}) = M$ , where  $\bar{e} > 0$  is the maximum contribution level. Further, the marginal benefit of effort  $dB^*/de$  is strictly greater than the marginal cost of effort  $dC/de$  whenever the marginal benefit is non-zero
- The optimal effort function is continuous and weakly increasing; it is strictly increasing on an interval  $[\underline{p}, 1/2]$ , equal to 0 when  $p \leq \underline{p}$ , and equal to  $\bar{e}$  when  $p \geq 1/2$ .
- The critical level of optimism  $\underline{p}$  is strictly greater than 0 if and only if the marginal cost is strictly positive at zero effort ( $C'(0) > 0$ ).

To illustrate the proposition, we graph the optimal reward scheme and optimal effort function for some examples. As in Figure 1, suppose  $C(e) = e^n$ ,  $n > 1$ , and let  $M = 1$  to facilitate comparisons. Figure 2 illustrates the optimal pleasure in agency reward and optimal effort schemes for different values of  $n$ . The derivation is reported in the appendix. Comparing Figures 1 and 2 reveals the effect of optimal inspiration. The leader could choose  $B(e) = \sqrt{e}$  to generate the outcomes in Figure 1, but she does not. Instead, she chooses pleasure in agency reward functions that are *convex* in their (relevant) strictly increasing segments. This, in turn, induces corresponding contribution functions that look very different from those of Figures 1. One key difference is the emergence of a distinct group of citizens who exert the maximum effort  $\bar{e}$  (corresponding to the flat segment of  $e^*(p)$  in Figure 2). Moreover, the leader's optimal inspiration increases the likelihood regime change: for a given  $n$ , the area under  $e^*(p)$  is larger in Figure 2 than in Figure 1. Alternatively, if the leader knew and could condition rewards on the citizens' degrees of optimism, she would require an effort level of at least  $e^*(p) = p^{1/n}$  for the maximum reward  $M = 1$ . Then, again, optimal effort functions would be very different, and no vanguard would emerge. Our optimal inspiration is the middle ground between these two implausible scenarios, where the leader either has no influence on pleasure in agency rewards, or she knows and can condition rewards on the citizens' degrees of optimism, and hence can extract all the surplus. The previous class of examples has  $C'(0) = 0$ . Figure 3 illustrates an example where  $C(e) = e^2 + 0.1e$ , so that the marginal cost at 0 is strictly positive,  $C'(0) = 0.1$ .

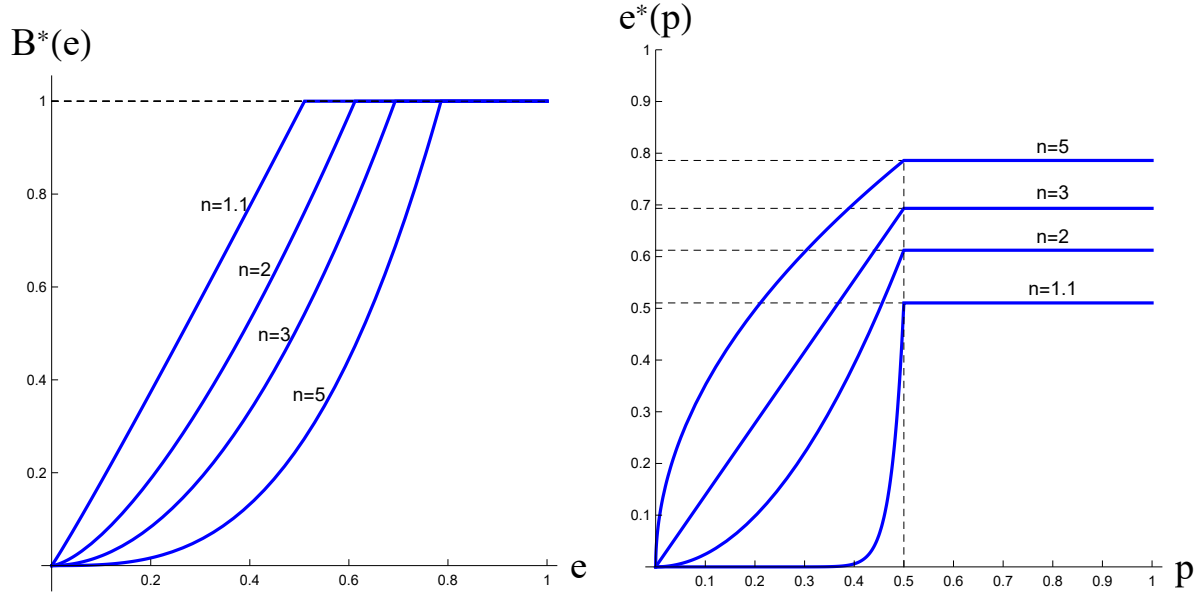


Figure 2: Optimal pleasure in agency rewards,  $B^*(e)$ , and its induced effort function,  $e^*(p)$ , for  $M = 1$  and cost function  $C(e) = e^n$ ,  $n > 1$ .

Further details about the optimal reward scheme and the optimal effort function, as well as an algorithm for calculating them, are reported in the proof of the proposition. In particular,<sup>5</sup>

$$\theta^* = \bar{e} - \frac{M}{4} \frac{1}{C'(\bar{e})}, \quad (5)$$

and  $\bar{e}$  is the unique solution to

$$\sqrt{C'(\bar{e})} \int_0^{\bar{e}} \sqrt{C'(x)} dx = M/2. \quad (6)$$

Moreover, the strictly increasing segment of the optimal effort function has a simple characterization

$$e^*(p) = C'^{-1}((2p)^2 C'(\bar{e})). \quad (7)$$

The basic features of the optimal pleasure in agency rewards and optimal effort function are intuitive: both are weakly increasing and continuous, with optimal rewards strictly increasing when ranging from 0 to the maximum. First, consider the shape of the optimal reward scheme when it is strictly increasing. In this case, marginal benefit exceeds the marginal cost. If the marginal cost did exceed the marginal benefit at some effort level, then any effort in

<sup>5</sup>Equation (6) follows from equations (42), (44) and (46). Equation (5) is calculated through integration by parts:  $\theta^*$  is  $\bar{e}$  minus the area above  $e^*(p)$  and below  $\bar{e}$ . From equation (44), this area is  $\lambda M$ . From (42) and (46),  $\lambda = 1/4C'(\bar{e})$ .

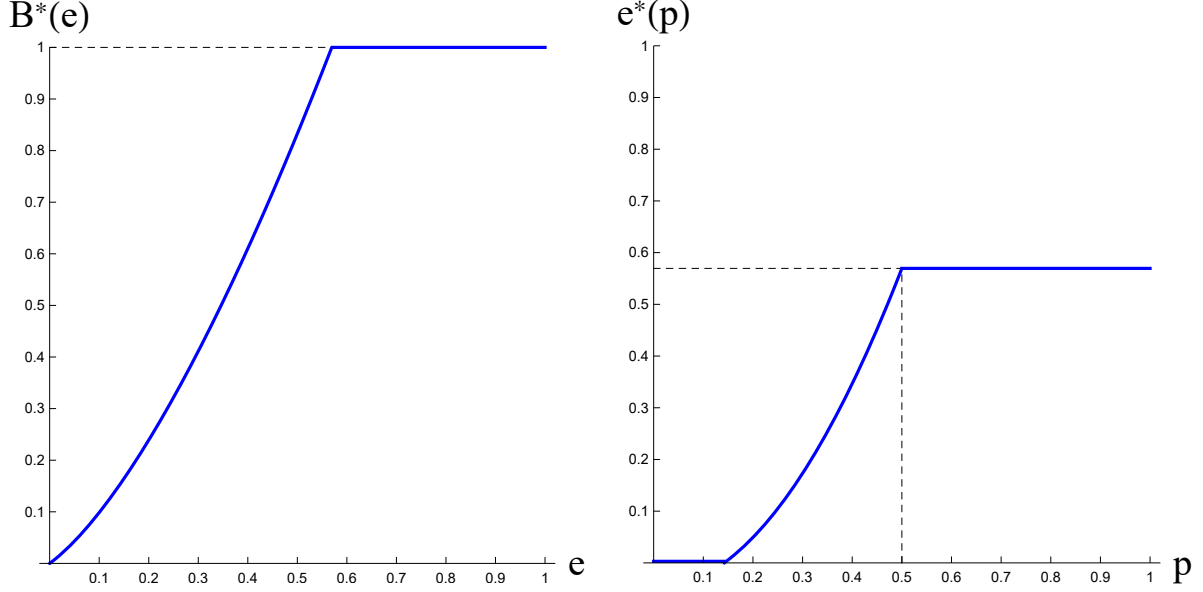


Figure 3: Optimal pleasure in agency rewards,  $B^*(e)$ , and its induced effort function,  $e^*(p)$ , for  $M = 1$  and cost function  $C(e) = e^2 + 0.1e$ . When  $C'(0) > 0$ , citizens with (endogenously) pessimistic beliefs about the likelihood of regime change do not exert any effort ( $\underline{p} > 0$ ).

the neighborhood of that level could not arise in equilibrium. But then it would be possible to replace the reward scheme with one that was constant in that neighborhood and increasing faster elsewhere, in a way that would increase the overall effort. Formally, the first order condition that  $pB'(e^*(p)) = C'(e^*(p))$  implies that  $B'(e^*(p)) > C'(e^*(p))$  when  $p < 1$ . To show strict convexity, observe that effort depends on the ratio of marginal costs to marginal benefits. Increasing marginal benefits will thus have more impact on effort when marginal cost is higher. But marginal cost is itself increasing in effort so marginal benefit will also be increasing.

More formally, we first use optimal control techniques (Arrow 1966; Seierstad and Sydsæter 1993; Kamien and Schwartz 2012) to show that  $e^*(p)$  is continuous, weakly increasing, and single-valued with  $e^*(0) = 0$ . Then, the equilibrium regime change threshold can be written as (Lemma 3 in the Online Appendix):

$$\theta_{e^*} = \int_{e=0}^{e_{max}} \left( 1 - \frac{C'(e)}{B'(e)} \right) de.$$

The leader seeks to maximize this subject to the constraint that  $B(e) \leq M$ . Differentiating the integrand in  $\theta_{e^*}$  with respect to  $B'(e)$  yields  $C'(e)/(B'(e))^2$ . Because  $C'(e)$  is increasing by assumption, the marginal gains of raising  $B'(e)$  are higher for higher effort levels. In contrast, the marginal costs of raising  $B'(e)$  to the leader are constant, because the leader has a fixed



“budget of slopes” ( $\int_{e=0}^{\bar{e}} B'(e)de \leq M$  for some  $\bar{e}$ ). This implies that an optimal  $B'(e)$  is increasing, and hence an optimal  $B(e)$  is convex.

Now consider the “exclusion” regions, where effort is constant. First, consider exclusion at the bottom. Recall that the optimal effort  $e^*(p)$  is continuous and increasing, and consider the smallest level of optimism  $\underline{p}$  after which  $e^*(p)$  becomes strictly positive. Then,  $\underline{p} B'(0) = C'(0)$ , and hence  $C'(0) > 0$  implies  $\underline{p} > 0$ . In fact, because the marginal costs of raising  $B'(e)$  are constant, while its marginal gains are  $C'(e)/(B'(e))^2$  as we discussed above,  $C'(e)/B'(e)$  is proportional to  $\sqrt{C'(e)}$ ,<sup>6</sup> and hence  $\underline{p} > 0$  exactly when  $C'(0) > 0$ . Next, consider high optimism, and suppose that optimal effort was strictly increasing in a neighborhood around 1. The optimal reward scheme would then have to be strictly increasing in the neighborhood of  $\bar{e}$ . Now suppose that we considered benefit functions that were equal to  $M$  in an open neighborhood of  $p = 1$ . This would decrease effort in the open neighborhood, but would allow optimal effort to be increased at other levels of  $p$ . For very small neighborhoods, the latter effect would be of higher order than the former. More formally, the convexity of an optimal  $B(e)$  (and that  $B(e) \geq C(e)$  in the relevant range) implies that an optimal  $B(e)$  hits its constraint  $M$  before  $C(e)$  does. In turn, this implies that an optimal effort scheme is constant for high values of  $p$ .

Finally, we examine why effort reaches its maximum exactly at  $p = 1/2$ . This follows from the endogenous property that the citizens’ optimism  $p$  is uniformly distributed on the interval  $[0, 1]$ . One can show that with a general distribution  $f$  of  $p$  and under some regularity conditions, the screening problem of Proposition 2 yields the critical  $p^*$  that solves

$$\frac{1 - F(p^*)}{f(p^*)} = -\frac{f(p^*)/p^*}{(f(p^*)/p^*)'}.$$

In the uniform case, this expression gives  $1 - p^* = -\frac{1/p^*}{-1/(p^*)^2} = p^*$ , and hence  $p^* = 1/2$ . In the next Section 5.1, we discuss an extension of the model where there is exogenous heterogeneity among citizens; the analysis of this problem formally reduces to allowing for more general distributions on  $p$ .

We end this section by highlighting the methodological contribution of our analysis. As an ingredient to our analysis, we solved a screening problem in a setting without transfers where a principal chooses a reward scheme with no cost, constrained only by the requirement that rewards belong to a closed interval. Our analysis of the screening problem exploits classic arguments from the screening literature. Guesnerie and Laffont (1984) is a key early reference on

---

<sup>6</sup>Formally,  $C'(e)/(B'(e))^2 = \lambda \Rightarrow B'(e) = \sqrt{C'(e)/\lambda} \Rightarrow C'(e)/B'(e) = \sqrt{\lambda} \sqrt{C'(e)}$ .

a rich class of screening problems that embeds monopoly problems of choosing quality (Mussa and Rosen 1978) or quantity (Maskin and Riley 1984) and the government’s regulation of a monopolist (Baron and Myerson 1982). In our problem, a principal gives “benefits” to an agent in exchange for “effort”, but agent utility is not linear in effort (non-transferable utilities) and the principal’s “budget” of rewards is not “smooth”: rewards up to a level are free to the principal, but higher rewards have infinite costs. The bunching-at-the-top feature of our screening solution also appears in Laffont and Robert’s (1996) and Hartline’s (2016, Ch. 8) analysis of optimal auction with financial constraints; because no buyer can bid above a known budget (e.g., due to an exogenous liquidity constraint), the seller cannot separate eager buyers. While non-standard, this screening problem is essentially that of Mookherjee and Png (1994) who study the problem of choosing the optimal likelihood of detection and optimal punishment schedule by an authority. Their key result is that the marginal punishment should remain lower than marginal harms of crime for low crime levels; in fact, if monitoring is sufficiently costly, a range of less harmful acts should be legalized. Although we use different techniques in our analysis, one can show that our screening problems can be transformed to each other. This implies that our total analysis with coordination also applies to a regime designing a repression scheme (expected costs of revolutionary efforts) to minimize the likelihood of regime change.

As we discussed before, our key methodological contribution is to analyze a setting where coordination and screening are intertwined. We show that this complex problem can be cleanly disentangled into solving a continuous action coordination game and a screening problem. The interaction of screening and coordination also arises in the simultaneous work of Shen and Zou (2018), but takes a very different form. They consider a benchmark binary action game but allow screening by the introduction of a third action where agents receive an intervention that only marginal agents will accept. This allows screening to have a larger impact relative to our case where the action set is held fixed.

## 5.1 Citizen Heterogeneity Beyond Beliefs

So far, we have assumed that one citizen is distinguished from another only by his belief about the likelihood of regime change, which depends on his private information. Thus, citizens are easier or harder to motivate solely based on these private beliefs about the likelihood of regime change: more optimistic citizens are easier to motivate. However, in addition to their beliefs about the likelihood of success, differences in the citizens’ past interactions with the state,

their social and economic status, and their inherent psychological dispositions, such as religious convictions, can influence how they react to a leader who aims to inspire anti-regime actions. For example, richer citizens may be harder to motivate (low  $\alpha$  below) (White 1989), or those who were treated worse by the state may be easier to motivate (high  $\alpha$  below) (Wood 2003). In this section, we study the effect of increases in heterogeneity among anti-regime forces on the likelihood of regime change.

We extend citizen payoffs, so that a citizen  $i$  is characterized by both his (endogenous) private belief  $p_i$  that the regime change occurs and his exogenous propensity to be influenced by the leader  $\alpha_i$ . We assume that  $\alpha_i$ s are independent of  $\theta$  and citizens' signals, and  $\alpha_i \sim iid F_\alpha$ . Thus, if a citizen  $i$  exerts effort  $e_i$ , his expected payoff is  $p_i \alpha_i B(e_i) - C(e_i)$ . Because  $\alpha$  and  $p$  are independent, the Uniform Threshold Belief property still holds and the distribution of beliefs is uniform. Thus, we can think of  $p\alpha$  as the augmented type  $t$ , where  $p \sim U[0, 1]$ ,  $\alpha \sim F_\alpha$ , and  $t \sim G$ . Then, equation (2) becomes

$$\theta^* = \int e^*(t) dG(t). \quad (8)$$

If the leader knew and could condition on the citizens' types, he could get a citizen with type  $t = p\alpha$  to contribute  $tM = e(t)$ , and hence  $\theta^* = M \cdot E[t]$ . But because these types are private, the leader has a screening problem. In this extension, we focus on the case where the cost of effort is linear  $C(e) = ce$ , and  $F_\alpha = U[\mu - \delta, \mu + \delta]$ , with  $\mu > 0$  and  $\delta \in (0, \mu)$ . It can be shown that linear costs imply that the leader's optimal reward scheme is to give the maximum benefit  $M$  to those whose efforts exceed some optimally set threshold  $\hat{e}$ . In turn, those whose augmented type exceeds  $c\hat{e}/M$  exert effort  $\hat{e}$ , and the rest do not participate. Thus, the leader's problem becomes

$$\max_{\hat{e}} (1 - G_\delta(c\hat{e}/M)) \hat{e}, \quad (9)$$

where we have made explicit the dependence of this distribution on the degree of dispersion in exogenous heterogeneity,  $\delta$ . The term  $(1 - G_\delta(c\hat{e}/M))$  captures the extensive margin of participation at the equilibrium regime change threshold, and the term  $\hat{e}$  captures the intensive margin.

**Proposition 3.** *Suppose costs are linear, and exogenous heterogeneity is distributed uniformly,  $F_\alpha = U[\mu - \delta, \mu + \delta]$ , with  $\mu > 0$  and  $\delta \in (0, \mu)$ . The equilibrium regime change threshold  $\theta^*(\delta)$  is strictly decreasing in the dispersion of exogenous heterogeneity:  $d\theta^*(\delta)/d\delta < 0$ . Moreover, the extensive margin of participation in revolution at the equilibrium regime change threshold  $\theta^*(\delta)$  is decreasing in the dispersion of exogenous heterogeneity  $\delta$ .*

To see the intuition, suppose  $p = 1$ , so that there is no coordination consideration, and the distribution of augmented types is  $U[\mu - \delta, \mu + \delta]$ . Now, increasing  $\delta$  is a mean-preserving spread, which rotates the “demand curve”  $(1 - G_\delta(t))$  counter clockwise around the median  $t = \mu$ . Thus, as Johnson and Myatt (2006) analyze in detail, when the leader’s optimal strategy is to seek the participation of a majority of potential participants (i.e., when the marginal participating type is below  $\mu$ ), a mean-preserving spread lowers  $\theta^*(\delta)$ . Our results show that the coordination aspects of the game, which cause the distribution of augmented types to be more complex, do not change these properties.

The literature typically focuses on inequality in the whole population, showing that higher inequality in the whole population *increases* the likelihood of regime change. For example, in Acemoglu and Robinson (2001, 2006) and Boix (2003), the society is divided into two groups: the rich who want to maintain the status quo, and the poor who seek to change it. They find that as inequality between the rich and the poor increases, the likelihoods of instability and regime change increase. In contrast, our heterogeneity corresponds to inequality or other forms of heterogeneity among potential revolutionaries, e.g., higher dispersion of income or religious convictions within the poor. Our result shows that higher heterogeneity among potential revolutionaries *reduces* the likelihood of regime change. These findings support “divide and conquer” tactics as a useful counterrevolutionary strategy: the more one can instill heterogeneity among potential revolutionaries, the more one can maintain the status quo.

Proposition 3 also shows that, at the equilibrium regime change threshold, the fraction of citizens who contribute to the revolution is decreasing in the dispersion of exogenous heterogeneity. This reflects the combined effects of the change in the distribution of heterogeneity and the leader’s adjustment. The size of the revolutionary coalition can have important consequences for the nature of post-revolution regime. As we discussed in the Introduction, when regime change happens mainly due to the efforts of a relatively small group who contributed to the cause far more than other polity members, the post-revolution regime is more likely to reflect their preferences and ideologies. In sum, our analysis suggests that higher dispersion of exogenous heterogeneity not only reduces the likelihood of regime change, but also reduces participation in the neighborhood of the equilibrium regime change threshold, with adverse consequences for the post-revolution regime.

## 6 Conclusion

What is the nature of inspiring regime change? In our framework, inspiring regime change refers to processes that activate, create, or manipulate psychological rewards that one receives from participating in a successful movement. Through speeches, meetings, and writings, leaders influence how different participation levels map into these pleasure in agency rewards. We formalized inspiring regime change in a context where citizens face coordination and information frictions. In essence, inspiring regime change is like an exchange in which a leader gives pleasure in agency rewards to citizens in turn for their anti-regime actions. The leader inspires to maximize the likelihood of regime change, but optimal inspiration involves a complex screening problem: leaders must design mechanisms to screen citizens with heterogeneous valuations of pleasure in agency rewards; valuations that are determined endogenously in the citizens' coordination problem. Using techniques from mechanism design and global games literatures, we show that this seemingly complex problem can be disentangled into simpler problems, which we characterize. To optimally inspire, a leader should aim to assign higher pleasure in agency rewards to higher participation levels in an accelerating manner up to its maximum—in contrast to standard concavity assumptions, the relevant part of an optimal pleasure in agency mapping is convex.

A consequence of optimal inspiration is the emergence of a group of citizens distinct from others in that they all engage in the endogenous maximum level of anti-regime activities. That is, when leaders act strategically and optimally, a vanguard emerges naturally, regardless of the leader's charisma and skills or the details of a society's culture. The leader's skills and cultural elements do influence the maximum participation level that the vanguard will engage in, but not their emergence. Other citizens participate at varying degrees depending on their characteristics, and some citizens do not participate. This feature departs from the common binary participation, and generates a distribution of participation levels that captures the plethora of observed anti-regime actions. Extending our framework to allow for heterogeneity among citizens beyond their information asymmetries, we show that higher heterogeneity among potential revolutionaries (e.g., higher inequality) can reduce the likelihood of regime change. More heterogeneity can also reduce the depth of the movement, defined as the fraction of participating citizens. This can affect the nature of the post-revolution regime, as revolutions that succeed based on high efforts of a narrow subset of the population tend to be less democratic.

Two directions for future research stand out. We focused on psychological rewards because

evidence suggests their primacy in various movements and because they have received scant attention in formal studies of regime change. However, even when material rewards are secondary considerations, they often appear in some form in movements, e.g., financial help to families of strikers, or covering the medical expenses of those injured in demonstrations. One could contemplate settings in which leaders have access to both psychological and material rewards, and investigate how material rewards influence the leaders' optimal strategy and the outcomes. A second direction is to endogenize the regime's repression schemes that determine the costs of anti-regime actions.<sup>7</sup> This will generate a rich strategic field in which opposition leaders design (pleasure in agency) rewards to maximize the likelihood of regime change, state officials design repression schemes to minimize that likelihood, and citizens decide their participation levels.

## 7 References

- Acemoglu, Daron, and Matthew O. Jackson. 2015. "History, Expectations, and Leadership in the Evolution of Social Norms." *Review of Economic Studies* 82: 1-34.
- Acemoglu, Daron, and James A. Robinson. 2001. "A Theory of Political Transition." *American Economic Review* 91: 938-63.
- Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Democracy and Dictatorship*. New York: Cambridge University Press.
- Ahlquist, John, and Margaret Levi. 2011. "Leadership: What It Means, What It Does, and What We Want to Know About It." *Annual Review of Political Science* 14: 1-24.
- Arrow, Kenneth. 1966. "Optimal Capital Policy with Irreversible Investment." Institute for Mathematical Studies in Social Science, Stanford University.
- Aytaç, Erdem, and Susan Stokes. 2019. *Why Bother?: Rethinking Participation in Elections and Protests*. New York, NY: Cambridge University Press.
- Barbera, Salvador, and Matthew Jackson. 2020. "A Model of Protests, Revolution, and Information." *Quarterly Journal of Political Science*. *Forthcoming*.
- Baron, David, and Roger Myerson. 1982. "Regulating a Monopolist with Unknown Costs." *Econometrica* 50 (4): 911-30.
- Bass, Bernard. 1985. *Leadership and Performance Beyond Expectations*. NY: Free Press.

---

<sup>7</sup>The large empirical literature on repression includes Davenport (1995, 2007) and Ritter and Conrad (2016).

- Blattman, Christopher, and Edward Miguel. 2010. "Civil War." *Journal of Economic Literature* 48: 3-57.
- Breen, Timothy. 2010. *American Insurgents, American Patriots: The Revolution of the People*. New York: Hill and Wang.
- Boix, Carles. 2003. *Democracy and Redistribution*. New York: Cambridge University Press.
- Boix, Carles, and Milan Svolik. 2013. "The Foundations of Limited Authoritarian Government: Institutions and Power-Sharing in Dictatorships." *Journal of Politics* 75: 300-16.
- Bueno de Mesquita, Ethan. 2010. "Regime Change and Revolutionary Entrepreneurs." *American Political Science Review* 104 (3): 446-66.
- Bueno de Mesquita, Ethan. 2013. "Rebel Tactics." *Journal of Political Economy* 121: 323-57.
- Burns, James MacGregor. 1978. *Leadership*. New York: Harper & Row Publishers.
- Burns, James MacGregor. 2003. *Transforming Leadership: A New Pursuit of Happiness*. New York: Grove Press.
- Carlsson, Hans, and Eric van Damme. 1993. "Global Games and Equilibrium Selection." *Econometrica* 61: 989-1018.
- Casper, Brett, and Scott Tyson. 2014. "Popular Protest and Elite Coordination in a Coup d'etat." *Journal of Politics* 76: 548-64.
- Chen, Heng, and Wing Suen. 2016. "Falling Dominoes: A Theory of Rare Events and Crisis Contagion." *American Economic Journal: Microeconomics* 8: 1-29.
- Chen, Heng, and Wing Suen. 2019. "Radicalism in Mass Movements: Asymmetric Information and Agenda Escalation." Mimeo.
- Dabashi, Hamid. 1993. *Theology of Discontent: The Ideological Foundation of the Islamic Revolution in Iran*. New York: New York University Press.
- Davenport, Christian. 1995. "Multi-Dimensional Threat Perception and State Repression: An Inquiry into Why States Apply Negative Sanctions." *American Journal of Political Science* 38: 683-713.
- Davenport, Christian. 2007. "State Repression and Political Order." *Annual Review of Political Science* 10: 1-23.
- Davies, James C. 1962. "Toward a Theory of Revolution." *American Sociological Review* 27: 5-19.

- DeNardo, James. 1985. *Power in Numbers: The Political Strategy of Protest and Rebellion*. Princeton, NJ: Princeton University Press.
- Dewan, Torun, and David P. Myatt. 2007. "Leading the Party: Coordination, Direction, and Communication." *American Political Science Review* 101: 827-45.
- Dewan T, Myatt David P. 2008. "The Qualities of Leadership: Direction, Communication, and Obfuscation." *American Political Science Review* 102: 351-68.
- Dewan, Torun, and Francesco Squintani. 2018 "Leadership with Trustworthy Associates." *American Political Science Review* 112: 844-59.
- Edmond, Chris. 2013. "Information Manipulation, Coordination and Regime Change." *Review of Economic Studies* 80: 1422-58.
- Frankel, David, Ady Pauzner and Stephen Morris. 2003. "Equilibrium Selection in Global Games with Strategic Complementarities." *Journal of Economic Theory* 108: 1-44.
- Gecas, Viktor. 2000. "Value Identities, Self-motives, and Social Movements." In *Social Movements, Protest, and Contention, v. 13*. Eds. S. Stryker, T. J. Owens, & R. W. White, p. 93-109. University of Minnesota Press.
- Geschwender, James. 1967. "Continuities in Theories of Status Consistency and Cognitive Dissonance." *Social Forces* 46: 165-7.
- Goldstone, Jack A. 2001. "Toward a Forth Generation of Revolutionary Theory." *Annual Review of Political Science* 4: 139-87.
- Guesnerie, Roger, and Jean-Jacques Laffont. 1984. "A Complete Solution to a Class of Principal-Agent Problems with an Application in the Control of a Self-Managed Firm." *Journal of Public Economics* 25: 329-69.
- Guimaraes, Bernardo, and Stephen Morris. 2007. "Risk and Wealth in a Model of Self-fulfilling Currency Attack." *Journal of Monetary Economics* 54: 2205-30.
- Gurr, Ted. 1970. *Why Men Rebel?* Princeton, NJ: Princeton University Press
- Hardin, Russell. 1982. *Collective Action*. Baltimore, MD: Johns Hopkins University Press.
- Hartline, Jason D. 2016. *Mechanism Design and Approximation*. Online Manuscript.
- Humphreys, Macartan, and Jeremy Weinstein. 2008. "Who Fights? The Determinants of Participation in Civil War." *American Journal of Political Science* 52(2): 436-55.
- Johnson, Justin, and David Myatt. 2006. "On the Simple Economics of Advertising, Marketing,



and Product Design.” *American Economic Review* 96: 756-84.

Kamien, Morton, and Nancy Schwartz. 2012. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management (2nd Edition)*. Dover.

Laffont, Jean-Jacques, and Jacques Robert. 1996. “Optimal Auction with Financially Constrained Buyers.” *Economics Letters* 52: 181-86.

Lawrence, Adria. 2017. “Repression and Activism among the Arab Spring’s First Movers: Evidence from Morocco’s February 20th Movement.” *British Journal of Political Science* 47(3): 699-718.

Loeper, Antoine, Jakub Steiner, and Colin Stewart. 2014. “Influential Opinion Leaders.” *Economic Journal* 124: 1147-67.

Maier, Pauline. 1972. *From Resistance to Revolution: Colonial Radicals and the Development of American Opposition to Britain, 1765-1776*. New York: Vintage Books.

Maskin, Eric, and John Riley. 1984. “Monopoly with Incomplete Information.” *RAND Journal of Economics* 15 (2): 171-96.

McAdam, Doug. 1999. *Political Process and the Development of Black Insurgency, 1930-1970. 2nd Ed.* Chicago: University of Chicago Press.

McCullough, David. 2005. *1776*. New York: Simon & Schuster.

Middlekauff, Robert. 2005. *The Glorious Cause: The American Revolution, 1763-1789*. New York, NY: Oxford University Press.

Mookherjee, Dilip, and I. P. L. Png. 1994. “Marginal Deterrence in Enforcement of Law.” *Journal of Political Economy* 102 (5): 1039-66.

Morris, Stephen, and Mehdi Shadmehr. 2018. “Repression and Repertoires.” Mimeo.

Morris, Stephen, and Hyun Song Shin. 1998. “Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks.” *American Economic Review* 88 (3): 587-597.

Morris, Stephen, and Hyun Song Shin. 2003. “Global Games: Theory and Application.” In *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress, Volume I*, Ed. M. Dewatripont, L. Hansen, and S. Turnovsky. NY: Cambridge University Press.

Mussa, Michael, and Sherwin Rosen. 1978. “Monopoly and Product Quality.” *Journal of Economic Theory* 18: 301-17.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*.

Cambridge, MA: Harvard University Press.

Pearlman, Wendy. 2018. "Moral Identity and Protest Cascades in Syria." *British Journal of Political Science* 48: 877-901.

Petersen, Roger D. 2001. *Resistance and Rebellion: Lessons from Eastern Europe*. New York: Cambridge University Press.

Popkin, Samuel. 1979. *The Rational Peasant: The Political Economy of Rural Society in Vietnam*. Berkeley, CA: University of California Press.

Rhodehamel, John. 2001. *The American Revolution: Writings from the War of Independence 1775-1783*. New York, NY: The Library of America.

Ritter, Emily H., and Conrad, Courtenay. 2016. "Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression." *American Political Science Review* 110: 85-99.

Ross, Michael. 2006. "A Closer Look at Oil, Diamonds, and Civil War." *Annual Review of Political Science* 9: 265-300.

Sambanis, Nicholas, and Moses Shayo. 2013. "Social Identification and Ethnic Conflict." *American Political Science Review* 107: 294-325.

Sandoz, Ellis. 1998. *Political Sermons of the American Founding Era: 1730-1805. 2nd Ed. Vol 1*. Indianapolis, IN: Liberty Fund.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior*. New York: Norton.

Seierstad, Atle, and Knut Sydsæter. 1993. *Optimal Control Theory with Economic Applications*. Amsterdam, The Netherlands: Elsevier.

Shadmehr, Mehdi. 2015. "Extremism in Revolutionary Movements." *Games and Economic Behavior* 94: 97-121.

Shadmehr, Mehdi. 2017. "Khomeini's Theory of Islamic State and the Making of the Iranian Revolution." Mimeo.

Shadmehr, Mehdi. 2019. Investment in the Shadow of Conflict: Globalization, Capital Control, and State Repression. *American Political Science Review* 113: 997-1011.

Shadmehr, Mehdi, and Dan Bernhardt. 2019. "Vanguards in Revolution." *Games and Economic Behavior* 115: 146-66.

Shayo, Moses. 2009. "A Model of Social Identity with an Application to Political Economy:

- Nation, Class and Redistribution.” *American Political Science Review* 103: 147-74.
- Shen, Lin, and Junyuan Zou. 2016. “Intervention with Voluntary Participation in Global Games.” Mimeo.
- Skocpol, Theda. 1997. “Cultural Idioms and Political Ideologies in the Revolutionary Reconstruction of State Power: A Rejoinder to Sewell.” In *Social Revolutions in the Modern World*. Ed. by Theda Skocpol. New York: Cambridge University Press.
- Smelser, Neil. 1962. *Theory of Collective Behavior*. New York: The Free Press.
- Snow, David, Burke Rochford, Jr., Steven Worden, and Robert Benford. 1986. “Belief Alignment Processes, Micromobilization, and Movement Participation.” *American Sociological Review* 51: 464-81.
- Siegel, David. 2011. “When Does Repression Work? Collective Action in Social Networks.” *Journal of Politics* 73: 993-1010.
- Tarrow, Sidney G. 2011. *Power in Movement: Social Movements and Contentious Politics*. 3rd Ed. New York: Cambridge University Press.
- Tilly, Charles. 1978. *From Mobilization to Revolution*. Reading, MA: Addison-Wesley.
- Tilly, Charles. 2006. *Regimes and Repertoires*. Chicago, IL: University of Chicago Press.
- Tilly, Charles. 2008. *Contentious Performances*. New York: Cambridge University Press.
- Tullock, Gordon. 1971. “The Paradox of Revolution.” *Public Choice* 11: 89-99.
- Tyson, Scott, and Alastair Smith. 2018. “Dual-Layered Coordination and Political Instability.” *Journal of Politics* 80: 44-58.
- Wantchékon, Léonard, and Omar García-Ponce. 2017. “Critical Junctures: Independence Movements and Democracy in Africa.” Mimeo.
- Washington, George. 1890. *The Writings of George Washington, (1777-1778)*. Collected and edited by Worthington Chauncey Ford. New York: G. P. Putnam’s Sons.
- Weinstein, Jeremy. 2007. *Inside Rebellion: The Politics of Insurgent Violence*. New York, NY: Cambridge University Press.
- White, Robert. 1989. “From Peaceful Protest to Guerrilla War: Micromobilization of the Provisional Irish Republican Army.” *American Journal of Sociology* 94: 1277-1302.
- Wood, Elisabeth J. 2003. *Insurgent Collective Action and Civil War in El Salvador*. New York, NY: Cambridge University Press.

## Online Appendix: Examples and Proofs

### Example

Suppose  $C(e) = e^n$ ,  $n > 1$ , and  $M = 1$ . What is the optimal  $B(e)$ ? From (43),

$$B(e) = \frac{1}{\sqrt{\lambda}} \int_0^e \sqrt{C'(x)} dx = \frac{1}{\sqrt{\lambda}} \int_0^e \sqrt{n} x^{\frac{n-1}{2}} dx = \frac{1}{\sqrt{\lambda}} \frac{2\sqrt{n}}{n+1} e^{\frac{n+1}{2}}. \quad (10)$$

From (44),  $B(\bar{e}) = 1$ , which implies

$$\frac{1}{\sqrt{\lambda}} \frac{2\sqrt{n}}{n+1} \bar{e}^{\frac{n+1}{2}} = 1. \quad (11)$$

From (46),  $\frac{C'(\bar{e})}{B'(\bar{e})} = \frac{1}{2}$ , which implies

$$\frac{n\bar{e}^{n-1}}{\sqrt{\frac{n}{\lambda}} \bar{e}^{\frac{n-1}{2}}} = \sqrt{n\lambda} \bar{e}^{\frac{n-1}{2}} = \frac{1}{2}. \quad (12)$$

From (11),  $\sqrt{\lambda} = \frac{2\sqrt{n}}{n+1} \bar{e}^{\frac{n+1}{2}}$ . Substituting this  $\sqrt{\lambda}$  into equation (12) yields  $\frac{2n}{n+1} \bar{e}^n = \frac{1}{2}$ , and hence

$$\bar{e} = \left( \frac{n+1}{4n} \right)^{\frac{1}{n}} \quad \text{and} \quad \sqrt{\lambda} = \frac{2\sqrt{n}}{n+1} \left( \frac{n+1}{4n} \right)^{\frac{n+1}{2n}}.$$

Substituting these into equation (10) yields the optimal  $B(e) = \left( \frac{4n}{n+1} \right)^{\frac{n+1}{2n}} e^{\frac{n+1}{2}}$  for  $e \leq \bar{e}$ .

Moreover,  $B(e) \leq M = 1$  for  $e > \bar{e}$ . For the purposes of this example, we choose  $B(e) = M = 1$  for  $e > \bar{e}$ . Thus,

$$B^*(e) = \begin{cases} \left( \frac{4n}{n+1} \right)^{\frac{n+1}{2n}} e^{\frac{n+1}{2}} & ; e \leq \bar{e} = \left( \frac{n+1}{4n} \right)^{\frac{1}{n}} \\ M = 1 & ; e > \bar{e}. \end{cases}$$

As expected from our earlier discussion,  $B^*(e)$  is *convex* for  $e \in [0, \bar{e}]$ . Moreover,

$$B^{*'}(e) = \begin{cases} \frac{n+1}{2} \left( \frac{4n}{n+1} \right)^{\frac{n+1}{2n}} e^{\frac{n-1}{2}} & ; e < \bar{e} = \left( \frac{n+1}{4n} \right)^{\frac{1}{n}} \\ 0 & ; e > \bar{e}. \end{cases}$$

How does  $e^*(p)$  look like? Recall that  $e^*(p) = \arg \max_{e \geq 0} p B^*(e) - e^n$ , and hence

$$e^*(p) = \begin{cases} 4^{\frac{1}{n(n-1)}} \left( \frac{n+1}{n} \right)^{\frac{1}{n}} p^{\frac{2}{n-1}} & ; p \in [0, \frac{1}{2}] \\ \bar{e} & ; p \in [\frac{1}{2}, 1]. \end{cases}$$

Figure 2 illustrated this solution.

## Proofs

**Proof of Lemma 2:** Let  $p_2 > p_1 > 0$ ,  $e_i \in e^*(p_i) = \arg \max_{e \geq 0} p_i B(e) - C(e)$ ,  $i \in \{1, 2\}$ . We establish that  $e_2 \geq e_1$  by way of contradiction. Suppose not, so that  $e_2 < e_1$ . First, from the optimality of  $e_1$  and  $e_2$ , we have:

$$p_2 B(e_2) - C(e_2) \geq p_2 B(e_1) - C(e_1) \quad \Leftrightarrow \quad p_2 [B(e_2) - B(e_1)] \geq C(e_2) - C(e_1) \quad (13)$$

$$p_1 B(e_1) - C(e_1) \geq p_1 B(e_2) - C(e_2) \quad \Leftrightarrow \quad C(e_2) - C(e_1) \geq p_1 [B(e_2) - B(e_1)] \quad (14)$$

Because  $C(e)$  is strictly increasing,  $e_2 < e_1$  implies  $C(e_2) < C(e_1)$ . Thus, from (14):

$$0 > C(e_2) - C(e_1) \geq p_1 [B(e_2) - B(e_1)] \quad \Rightarrow \quad B(e_2) - B(e_1) < 0,$$

and hence:

$$p_1 [B(e_2) - B(e_1)] > p_2 [B(e_2) - B(e_1)]. \quad (15)$$

However, combining (13) and (14), we have:

$$p_1 [B(e_2) - B(e_1)] \leq C(e_2) - C(e_1) \leq p_2 [B(e_2) - B(e_1)]. \quad (16)$$

Hence, (15) and (16) contradict each other, and hence our assumption that  $e_2 < e_1$  must be false. The proposition is then implied by Proposition 1.  $\square$

**Lemma 3.** *If the optimal effort correspondence  $e^*$  is single-valued, continuous and weakly increasing, and  $e_{min} = 0$ , then the regime threshold in the unique monotonic equilibrium is given by*

$$\theta_{e^*} = \int_{e=0}^{e_{max}} \left( 1 - \frac{C'(e)}{B'(e)} \right) de. \quad (17)$$

**Proof of Lemma 3:** By assumption,  $e^*(p)$  is a weakly increasing function. Whenever  $e^*(p)$  is strictly increasing, we have  $pB'(e^*(p)) = C'(e^*(p))$ . Let  $[p_1, p_2]$  be an interval on which  $e^*$  is

strictly increasing and continuous. In this case,

$$\begin{aligned}
\int_{p=p_1}^{p_2} e^*(p) dp &= [p e^*(p)]_{p=p_1}^{p_2} - \int_{p=p_1}^{p_2} p [e^*]'(p) dp \\
&= p_2 e^*(p_2) - p_1 e^*(p_1) - \int_{e=e^*(p_1)}^{e^*(p_2)} \frac{C'(e)}{B'(e)} de,
\end{aligned}$$

where the last inequality uses the change of variables  $e = e^*(p)$ , and the derivative of  $B'(e)$  at  $e^*(p_2)$  is the left derivative and at  $e^*(p_1)$  is the right derivative.

Now let  $[p_2, p_3]$  be an interval on which  $e^*$  is constant. In this case,

$$\begin{aligned}
\int_{p=p_2}^{p_3} e^*(p) dp &= (p_3 - p_2) e^*(p_3) \\
&= p_3 e^*(p_3) - p_2 e^*(p_2).
\end{aligned}$$

We conclude that

$$\int_{p=p_1}^{p_3} e^*(p) dp = p_3 e^*(p_3) - p_1 e^*(p_1) - \int_{e=e^*(p_1)}^{e^*(p_2)} \frac{C'(e)}{B'(e)} de.$$

Now, consider a partition of  $[0, 1]$ , and suppose that  $e^*(p)$  is strictly increasing on the intervals  $[p_1, p_2], [p_3, p_4], \dots, [p_{2n-1}, p_{2n}]$ , where  $p_1 < p_2 < \dots < p_{2n}$ . Then,

$$\begin{aligned}
\int_{p=0}^1 e^*(p) dp &= e_{max} - \sum_{m=1}^n \left( \int_{e=e^*(p_{2m-1})}^{e^*(p_{2m})} \frac{C'(e)}{B'(e)} de \right) \\
&= e_{max} - \int_{e=0}^{e_{max}} \frac{C'(e)}{B'(e)} de \\
&= \int_{e=0}^{e_{max}} \left( 1 - \frac{C'(e)}{B'(e)} \right) de.
\end{aligned}$$

□

**Proof of Proposition 2:** As a first step in analyzing this problem, we apply the revelation principle to transform the problem into one where the leader chooses effort levels and benefit levels

$\{(e(p), B(p))\}$  depending on the threshold probability  $p$  that a citizen assigns to regime change. This mechanism design approach amounts to treating the citizens' endogenous beliefs *as if* they were exogenous and distributed uniformly on  $[0, 1]$ . However, determining  $\{(e(p), B(p))\}$  generates a set of "recommended" revolutionary efforts  $E \equiv \{e(p) \text{ s.t. } p \in [0, 1]\}$ , and the corresponding rewards  $B(e)$  for those  $e \in E$ . It remains to characterize  $B(e)$  for  $e \notin E$ . The only requirement for such  $B(e)$  is that players do not choose it. For example, one could set  $B(e) = 0$  for all  $e \notin E$ . Of course, this choice is not unique. We choose  $B(e)$  for  $e \notin E$  such that  $B(e)$  is constant for  $e \notin E$ , and  $B(e)$  is continuous for all  $e \geq 0$ . Therefore, we can write the leader's problem as:

$$\begin{aligned} & \max_{\{(e(p), B(p))\}} \int_{p=0}^1 e(p) dp \\ \text{s.t. } & pB(p) - C(e(p)) \geq 0, & \forall p \in [0, 1] \\ & pB(p) - C(e(p)) \geq p B(p') - C(e(p')), & \forall p, p' \in [0, 1] \\ & B(p) \in [0, M], & \forall p \in [0, 1]. \end{aligned}$$

Observe that under this program, the designer can assign any cost to any citizen through their choice of  $e$ . To simplify notation, write  $h(p) = C(e(p))$ .

We can use this new representation of the problem to use standard arguments from screening models. We first establish that  $B(p)$  is weakly increasing, and hence  $B(p)$  is piecewise continuously differentiable. The incentive compatibility constraints imply  $pB(p) - h(p) \geq p B(p') - h(p')$  and  $p' B(p') - h(p') \geq p' B(p) - h(p)$ . Adding these inequalities implies:  $(p - p')[B(p) - B(p')] \geq 0$ . Hence,  $B(p)$  is weakly increasing, and hence  $B(p)$  is piecewise continuously differentiable. Thus, a necessary first order condition is  $pB'(p) - h'(p) = 0$  almost everywhere, with the corresponding second order condition  $pB''(p) - h''(p) \leq 0$ . Differentiating the FOC w.r.t.  $p$  yields  $B'(p) + pB''(p) - h''(p) = 0$ . Thus, the SOC simplifies to  $B'(p) \geq 0$ . Moreover, because  $p B(p) - h(p)$  is increasing in  $p$ , the condition  $pB(p) - C(e(p)) \geq 0, \forall p \in [0, 1]$ , simplifies to  $C(e(0)) = h(0) = 0$ . Further, because  $B'(p) \geq 0$ , the constraint  $B(p) \in [0, M], \forall p \in [0, 1]$ , can be replaced by  $B(0) \geq 0$  and  $B(1) \leq M$ . Combining these results, the leader's problem becomes:

$$\max_{\{(e(p), B(p))\}} \int_{p=0}^1 e(p) dp \tag{18}$$

$$pB'(p) - h'(p) = 0, \quad h(0) = 0 \tag{19}$$

$$B'(p) \geq 0 \tag{20}$$

$$B(0) \geq 0, \quad B(1) \leq M. \tag{21}$$

We can re-write this problem, letting  $\Pi(\cdot) = C^{-1}(\cdot)$ , so that  $\Pi(h(p)) = e(p)$ . Then, the leader's problem (18)-(21) becomes:

$$\max_{\{(B(p), h(p))\}} \int_{p=0}^1 \Pi(h(p)) dp \quad (22)$$

$$h'(p) = pB'(p), \quad h(0) = 0 \quad (23)$$

$$B'(p) \geq 0, \quad B(0) \geq 0, \quad \text{and} \quad B(1) \leq M. \quad (24)$$

Because incentives are created by the slope of benefits, it is clear that  $B(0) = 0$  and  $B(1) = M$ . To proceed, we use the optimal control techniques by defining two state variables. Taking a similar approach to Kamien and Schwartz (2012, p. 244-6), let  $(h, B)$  be the state and  $B'$  be the control, so that  $h' = pB'$ . The Hamiltonian and Lagrangian are:

$$H = \Pi(h) + \lambda_h pB' + \lambda_B B'.$$

$$L = H + \mu B'.$$

Then, by the maximum principle,

$$\frac{\partial L}{\partial B'} = \lambda_B + \lambda_h p + \mu = 0, \quad \mu \geq 0, \quad \mu B' = 0. \quad (25)$$

$$\lambda'_B(p) = -\frac{\partial L}{\partial B} = 0. \quad (26)$$

$$\lambda'_h(p) = -\frac{\partial L}{\partial h} = -\Pi'(h(p)), \quad \lambda_h(1) = 0. \quad (27)$$

$$B' \quad \text{must maximize } H \text{ given optimal } (B, h) \text{ and } (\lambda_B, \lambda_h). \quad (28)$$

Moreover, we recognize that  $B(1) = M$  is captured by having a free  $\lambda_B(1)$  as opposed to  $\lambda_h(1) = 0$  that stems from a free  $h(1)$ . Now, because  $C(e)$  is strictly convex, its inverse  $\Pi(\cdot)$  is strictly concave. This allows us to apply Arrow's (1966) approach together with a theorem from Seierstad and Sydsæter (1993) to show that  $h(p)$  and  $B(p)$  have no jumps.

**Lemma 4.** *Optimal  $h(p)$  and  $B(p)$  have no jump.*

**Proof of Lemma 4:** We show that an optimal  $h(p)$  has no jump, which then implies that an optimal  $B(p)$  has no jump. First, we prove that there is no interior jumps at any  $p \in (0, 1)$ . Our proof is based on Arrow (1966, p. 11-3). Suppose  $h(p)$  has a jump at  $\tau \in (0, 1)$ . Let  $h(\tau^+) \equiv \lim_{p \rightarrow \tau^+} h(p)$  and  $h(\tau^-) \equiv \lim_{p \rightarrow \tau^-} h(p)$ . Because  $h(p)$  is increasing,  $h(\tau^-) < h(\tau^+)$ . From (26) and (27), recall that  $\lambda'_B(p) + p\lambda'_h(p) = -p\Pi'(h(p))$  in  $(\tau - \epsilon, \tau) \cup (\tau, \tau + \epsilon)$  for some  $\epsilon > 0$ . Because  $\Pi(x)$  is strictly concave,  $\Pi'(h(\tau^+)) < \Pi'(h(\tau^-))$ , and hence:

$$\lambda'_B(\tau^-) + \tau \lambda'_h(\tau^-) < \lambda'_B(\tau^+) + \tau \lambda'_h(\tau^+). \quad (29)$$



Moreover, from (25),  $\lambda_B(p) + p\lambda_h(p) \leq 0$  and  $\lambda_B(\tau) + \tau\lambda_h(\tau) = 0$ .<sup>8</sup> Further, from condition (74) of Theorem 7 in Seierstad and Sydsæter (1993, p. 197),  $\lambda_h(p)$  and  $\lambda_B(p)$  are continuous at  $\tau$ . Thus, in a right neighborhood of  $\tau$ , we have  $\lambda'_B(p) + p\lambda'_h(p) + \lambda_h(p) \leq 0$ , and in a left neighborhood of  $\tau$ , we have  $\lambda'_B(p) + p\lambda'_h(p) + \lambda_h(p) \geq 0$ . But this contradicts (29).

Next, we prove that there is no jump at  $p = 0$  or  $p = 1$ . From (26),  $\lambda_B(p) = \text{constant} \equiv \bar{\lambda}$  on  $p \in (0, 1)$ . Moreover,  $\lambda_B(p)$  is continuous (Seierstad and Sydsæter 1993, p. 197). Thus,  $\lambda_B(p) = \bar{\lambda}$ , for  $p \in [0, 1]$ . From conditions (74) and (75) in Theorem 7 of Seierstad and Sydsæter (1993, p. 197),  $\lambda_B(\tau) + \tau\lambda_h(\tau) = 0$ . Combining this with  $\lambda_B(p) = \bar{\lambda}$  and (27),  $\bar{\lambda} + \tau \int_{x=\tau}^1 \Pi'(h(x))dx = 0$ . In particular, if  $\tau = 0$  or  $\tau = 1$ , then  $\bar{\lambda} = 0$ . But from (25),  $\bar{\lambda} + \lambda_h p \leq 0$  and  $\lambda_h(p) = \int_{x=p}^1 \Pi'(h(x))dx$  is positive for some  $p \in (0, 1)$ . Thus,  $\bar{\lambda} < 0$ . A contradiction.  $\square$

From (26) and (27),

$$\lambda_B(p) = \text{constant} = \bar{\lambda} \quad \text{and} \quad \lambda_h(p) = - \int_{x=p}^1 \lambda'_h(x)dx = \int_{x=p}^1 \Pi'(h(x))dx. \quad (30)$$

Moreover, from (25),  $\bar{\lambda} + p \lambda_h(p) \leq 0$ . Because at  $p \in (0, 1)$ ,  $p \lambda_h(p) = p \int_{x=p}^1 \Pi'(h(x))dx > 0$ , we must have  $\bar{\lambda} < 0$ .

**Lemma 5.** *There is no interior bunching. That is, there is no  $0 < p_1 < p_2 < 1$  such that  $h'(p) = 0$  for  $p \in (p_1, p_2)$ , with  $h'(p) > 0$  in a left neighborhood of  $p_1$  and in a right neighborhood of  $p_2$ .*

**Proof of Lemma 5:** Suppose not. Let  $h(p) = \bar{h}$  for  $p \in [p_1, p_2]$ . Because  $h(p)$  and hence  $B(p)$  must be strictly increasing in a left neighborhood of  $p_1$  and in a right neighborhood of  $p_2$ ,  $\mu = 0$  in those neighborhoods. Moreover,  $\lambda_B(p)$  and  $\lambda_h(p)$  are continuous. Therefore, (25) implies that  $\lambda_B(p_1) + p_1\lambda_h(p_1) = 0 = \lambda_B(p_2) + p_2\lambda_h(p_2)$ . Because  $\lambda_B(p) = \bar{\lambda}$  from (30), we have,

$$p_1\lambda_h(p_1) = p_2\lambda_h(p_2). \quad (31)$$

Further, from (30),

$$\lambda_h(p_1) = \int_{p_1}^{p_2} \Pi'(h(x))dx + \int_{p_2}^1 \Pi'(h(x))dx = (p_2 - p_1) \Pi'(\bar{h}) + \lambda_h(p_2). \quad (32)$$

Combining equations (31) and (32) yields:

$$p_1 \Pi'(\bar{h}) = \lambda_h(p_2). \quad (33)$$

---

<sup>8</sup>More formally,  $\lambda_B(\tau) + \tau\lambda_h(\tau) = 0$  obtains from conditions (74) and (75) in Theorem 7 of Seierstad and Sydsæter (1993, p. 197).

Next, consider  $p_3 \in (p_1, p_2)$ . From (25),  $\lambda_B(p_3) + p_3 \lambda_h(p_3) \leq 0 = \lambda_B(p_2) + p_2 \lambda_h(p_2)$ . Hence,

$$p_3 \lambda_h(p_3) \leq p_2 \lambda_h(p_2). \quad (34)$$

Mirroring the calculations of equation (32), we have:

$$\lambda_h(p_3) = (p_2 - p_3) \Pi'(\bar{h}) + \lambda_h(p_2). \quad (35)$$

Combining equations (34) and (35) yields:

$$p_3 \Pi'(\bar{h}) \leq \lambda_h(p_2). \quad (36)$$

From equations (33) and (36),  $p_3 \Pi'(\bar{h}) \leq p_1 \Pi'(\bar{h})$ . However, because  $p_3 > p_1$  and  $\Pi'(\bar{h}) > 0$ , we must have  $p_3 \Pi'(\bar{h}) > p_1 \Pi'(\bar{h})$ , a contradiction.  $\square$

Next, we solve the problem ignoring the constraint  $B'(p) \geq 0$ , and subsequently check whether and when this constraint binds. Without  $B'(p) \geq 0$ ,  $\mu = 0$ , and hence from (25) and (30),

$$p \int_{x=p}^1 \Pi'(h(x)) dx + \bar{\lambda} = 0, \text{ and hence } \frac{d}{dp} \left\{ p \int_{x=p}^1 \Pi'(h(x)) dx + \bar{\lambda} \right\} = 0 \quad (37)$$

Differentiating (37) with respect to  $p$  yields  $\Pi'(h(p)) = -\frac{\bar{\lambda}}{p^2}$ , and hence

$$h(p) = \Pi'^{-1} \left( -\frac{\bar{\lambda}}{p^2} \right). \quad (38)$$

Moreover, Because  $\Pi(h)$  is strictly concave,  $h(p)$  is strictly increasing, and hence  $B(p)$  is strictly increasing as far as the constraints  $B(0) = 0$  and  $B(1) = M$  are satisfied. Therefore, an optimal  $B(p)$  takes the following form:

$$B(p) = \begin{cases} 0 & ; p \in [0, p_1] \\ \text{strictly increasing function} & ; p \in [p_1, p_2] \\ M & ; p \in [p_2, 1], \end{cases}$$

for  $0 \leq p_1 < p_2 \leq 1$ . This, in turn, implies a similar form for  $h(p)$ , and hence for optimal effort schedule:

$$e^*(p) = \begin{cases} 0 & ; p \in [0, \underline{p}] \\ \text{strictly increasing and satisfies the first order condition} & ; p \in [\underline{p}, \bar{p}] \\ \bar{e} & ; p \in [\bar{p}, 1], \end{cases} \quad (39)$$

for some  $0 \leq \underline{p} < \bar{p} \leq 1$  and  $\bar{e} > 0$ .

Equation (39) together with Lemma 3 allows us to write the leader's objective function as:

$$\theta^* = \int_{e=0}^{\bar{e}} \left( 1 - \frac{C'(e)}{B'(e)} \right) de,$$

where we recognize that  $e^*(p)$  satisfying the first order condition implies  $B'(e) \geq C'(e)$ , and that  $\underline{p} = \frac{C'(0)}{B'(0)}$  and  $\bar{p} = \frac{C'(\bar{e})}{B'(\bar{e})}$ . Thus, we can formulate the leader's problem as:

$$\max_{B', \bar{e}} \int_0^{\bar{e}} \left( 1 - \frac{C'(e)}{B'(e)} \right) de \quad (40)$$

$$s.t. \ B(\bar{e}) = M, \ B(0) = 0, \ B'(e) \geq C'(e), \ \bar{e} \geq 0,$$

Writing  $B(\bar{e})$  as  $B(\bar{e}) = \int_{e=0}^{\bar{e}} B'(e)de$ , (40) can be written as:

$$\begin{aligned} \max_{B', \bar{e}} \quad & \int_0^{\bar{e}} \left( 1 - \frac{C'(e)}{B'(e)} \right) de \\ s.t. \quad & \int_{e=0}^{\bar{e}} B'(e)de = M, \ B(0) = 0, \ B'(e) \geq C'(e), \ \bar{e} \geq 0. \end{aligned}$$

This shows that one can think of the leader's problem as him deciding a maximum level of effort  $\bar{e} \geq 0$  to induce, and then deciding how to allocate a fixed supply of marginal benefit to different effort levels  $B' : [0, \bar{e}] \rightarrow \mathbb{R}$ . There are two other constraints:  $\bar{e} \geq 0$  and  $B'(e) \geq C'(e)$ . However, as it will be clear from the solution, these two constraints are automatically satisfied.

Thus, the leader's problem becomes a constrained point-wise optimization. The Lagrangian is

$$\begin{aligned} L &= \int_0^{\bar{e}} \left( 1 - \frac{C'(e)}{B'(e)} \right) de + \lambda \left( M - \int_{e=0}^{\bar{e}} B'(e)de \right) \\ &= \int_0^{\bar{e}} \left( 1 - \frac{C'(e)}{B'(e)} - \lambda B'(e) \right) de + \lambda M. \end{aligned}$$

Optimal  $B'(e)$  simplifies to a point-wise maximization

$$\frac{\partial L}{\partial B'(e)} \left( 1 - \frac{C'(e)}{B'(e)} - \lambda B'(e) \right) = \frac{C'(e)}{[B'(e)]^2} - \lambda = 0. \quad (41)$$

This shows that at the optimum, the marginal gain of raising  $B'(e)$ , i.e.,  $C'(e)/[B'(e)]^2$ , equals its marginal cost  $\lambda$ . Thus,

$$B'(x) = \frac{1}{\sqrt{\lambda}} \sqrt{C'(x)}, \quad (42)$$

and

$$B(e) = \int_0^e B'(x) dx = \frac{1}{\sqrt{\lambda}} \int_0^e \sqrt{C'(x)} dx. \quad (43)$$

Combining this with the constraint yields

$$B(\bar{e}) = \frac{1}{\sqrt{\lambda}} \int_0^{\bar{e}} \sqrt{C'(x)} dx = M. \quad (44)$$

Moreover, optimal  $\bar{e}$  also satisfies the first order condition

$$\frac{\partial L}{\partial \bar{e}} = 1 - \frac{C'(\bar{e})}{B'(\bar{e})} - \lambda B'(\bar{e}) = 0 \Rightarrow 1 - \frac{C'(\bar{e})}{B'(\bar{e})} = \lambda B'(\bar{e}). \quad (45)$$

Substituting for  $\lambda$  from (41) into (45) yields

$$1 - \frac{C'(\bar{e})}{B'(\bar{e})} = \frac{C'(\bar{e})}{B'(\bar{e})} \Rightarrow \frac{C'(\bar{e})}{B'(\bar{e})} = \frac{1}{2}. \quad (46)$$

Thus, if  $\bar{e}$  and  $\lambda$  are the solution to equations (44) and (46), we have solved for  $B$ .

From equation (42), we have  $\frac{C'(e)}{B'(e)} = \sqrt{\lambda C'(e)} > 0$ . This has two implications: (1) because  $C(e)$  is strictly convex,  $B(e)$  is strictly convex for  $0 < e < \bar{e}$ ; and  $\frac{C'(e)}{B'(e)}$  is strictly increasing. Thus,  $\frac{C'(\bar{e})}{B'(\bar{e})} = \frac{1}{2}$  implies that  $B'(e) \geq C'(e)$ . (2)  $e^*(p)$  is strictly increasing between  $\underline{p} = \sqrt{\lambda} \sqrt{C'(0)}$  and  $p = 1/2$ ; it is equal to 0 for  $p \leq \underline{p}$  and equal to  $\bar{e}$  for  $p \geq 1/2$ .  $\square$

**Proof of Proposition 3:** First, we prove a lemma.

**Lemma 6.** *Suppose  $X$  and  $Y$  are two independent random variables, where  $X \sim U[0, 1]$  and  $Y \sim U[\mu - \delta, \mu + \delta]$ , with  $0 < \delta \leq \mu$ . Then,  $X \cdot Y \sim G(\cdot)$ , where*

$$G(t) = \begin{cases} \max\{0, t \log(\frac{\mu+\delta}{\mu-\delta})/2\delta\} & ; t \leq \mu - \delta \\ (t \log(\frac{\mu+\delta}{t}) + t - (\mu - \delta)) / 2\delta & ; \mu - \delta \leq t \leq \mu + \delta \\ 1 & ; \mu + \delta \leq t. \end{cases}$$

Moreover,  $G(t)$  is differentiable except at 0.

**Poof of Lemma 6:**  $G(t)$  is

$$Pr(X \cdot Y \leq t) = \int_{\mu-\delta}^{\mu+\delta} Pr(X \leq t/Y) \frac{dY}{2\delta} = \begin{cases} 0 & ; t \leq 0 \\ \int_{\mu-\delta}^{\mu+\delta} \frac{t}{Y} \frac{dY}{2\delta} & ; 0 \leq t \leq \mu - \delta \\ \int_{\mu-\delta}^t \frac{dY}{2\delta} + \int_t^{\mu+\delta} \frac{t}{Y} \frac{dY}{2\delta} & ; \mu - \delta \leq t \leq \mu + \delta. \\ 1 & ; t \geq \mu + \delta. \end{cases}$$

The rest follows from calculating the integrals, differentiating  $G(t)$ , and checking its differentiability at  $t = \mu \pm \delta$ .  $\square$

From (9) and Lemma 6, the leader maximizes

$$(1 - G(t)) t = \begin{cases} \left(1 - \frac{t}{2\delta} \log\left(\frac{\mu+\delta}{\mu-\delta}\right)\right) t & ; t \in [0, \mu - \delta] \\ \left(1 - \frac{t}{2\delta} \log\left(\frac{\mu+\delta}{t}\right) - \frac{t-(\mu-\delta)}{2\delta}\right) t & ; t \in [\mu - \delta, \mu + \delta]. \end{cases} \quad (47)$$

Differentiating with respect to  $t$  yields

$$\frac{d[(1 - G(t)) t]}{dt} = \begin{cases} 1 - \frac{t}{\delta} \log\left(\frac{\mu+\delta}{\mu-\delta}\right) & ; t \in (0, \mu - \delta] \\ 1 - \frac{t}{\delta} \log\left(\frac{\mu+\delta}{t}\right) - \frac{t-(\mu-\delta)}{2\delta} & ; t \in [\mu - \delta, \mu + \delta]. \end{cases}$$

When  $t \approx 0$ , the derivative is clearly positive, and when  $t \approx \mu + \delta$ , using Taylor series expansion, one can show that the derivative is negative. Next, observe that

$$\left. \frac{d[(1 - G(t)) t]}{dt} \right|_{t=\mu-\delta} = 1 - \frac{\mu - \delta}{\delta} \log\left(\frac{\mu + \delta}{\mu - \delta}\right).$$

One can show that  $\frac{\mu-\delta}{\delta} \log\left(\frac{\mu+\delta}{\mu-\delta}\right)$  is strictly decreasing in  $\delta$ , and using L'Hospital's rule, one can show that it approaches 2 when  $\delta \rightarrow 0$ , and it approaches 0 when  $\delta \rightarrow \mu$ . Thus, there is a unique  $\delta \in (0, \mu)$  that solves  $1 = \frac{\mu-\delta}{\delta} \log\left(\frac{\mu+\delta}{\mu-\delta}\right)$ . Let  $\hat{\delta}$  be that solution.

$$1 = \frac{\mu - \hat{\delta}}{\hat{\delta}} \log\left(\frac{\mu + \hat{\delta}}{\mu - \hat{\delta}}\right). \quad (48)$$

Then,

$$\left. \frac{d[(1 - G(t)) t]}{dt} \right|_{t=\mu-\delta} > 0 \Leftrightarrow \delta > \hat{\delta}. \quad (49)$$

Further, observe that

$$\frac{d^2[(1 - G(t)) t]}{dt^2} = -\frac{1}{\delta} \log\left(\frac{\mu + \delta}{\mu - \delta}\right) < 0, \quad t \in (0, \mu - \delta). \quad (50)$$

Thus, if  $\delta < \hat{\delta}$ , then there is a local maximum  $t^* \in (0, \mu - \delta)$ , satisfying the first order condition:

$$1 - \frac{t^*}{\delta} \log\left(\frac{\mu + \delta}{\mu - \delta}\right) = 0 \Leftrightarrow t^* = \delta / \log\left(\frac{\mu + \delta}{\mu - \delta}\right). \quad (51)$$

This yields the result if it is also the global maximum, which we will show below. Moreover, it can be shown that  $t^*(\delta) = \delta / \log\left(\frac{\mu + \delta}{\mu - \delta}\right)$  is strictly decreasing in  $\delta \in (0, \mu)$ .

If  $\delta > \hat{\delta}$ , then  $t^*$  must be in  $(\mu - \delta, \mu + \delta)$  and satisfy the first order condition:

$$1 - \frac{t^*}{\delta} \log\left(\frac{\mu + \delta}{t^*}\right) - \frac{t^* - (\mu - \delta)}{2\delta} = 0. \quad (52)$$

First, one can confirm that  $t^* = \frac{\mu - \hat{\delta}}{\mu + \hat{\delta}} (\mu + \delta)$  satisfies this equation, where we recall the definition of  $\hat{\delta}$  from equation (48). To see this solution is unique, rewrite equation (52) as:

$$\frac{\mu + \delta}{2t^*} - \frac{1}{2} = \log\left(\frac{\mu + \delta}{t^*}\right) \Leftrightarrow z - 1 = 2 \log(z), \text{ where } z \equiv \frac{\mu + \delta}{t^*}.$$

Because  $t^* \in (\mu - \delta, \mu + \delta)$  implies  $z \in (1, \infty)$ , the solution to  $z - 1 = 2 \log(z)$  is unique. This solution corresponds to a maximum because we showed that the derivative is negative when  $t \approx \mu + \delta$ , and it is positive when  $t \approx \mu - \delta$  and  $\delta > \hat{\delta}$  (see (49)).

It remains to show that, if  $\delta < \hat{\delta}$ , then the local maximum in (51) is the global maximum. Because we showed in (50) that the objective function is concave for  $t \in (0, \mu - \delta)$ , it suffices to show that the global maximum is not in  $(\mu - \delta, \mu + \delta)$ . Suppose it is. This implies that there are more than one extremum in  $(\mu - \delta, \mu + \delta)$ . But we showed that the first order condition (52) has a unique solution, which is a contradiction. Thus, we have shown

$$t^*(\delta) = \begin{cases} \frac{\delta}{\log\left(\frac{\mu + \delta}{\mu - \delta}\right)} & ; \delta \leq \hat{\delta} \\ \frac{\mu - \hat{\delta}}{\mu + \hat{\delta}} (\mu + \delta) & ; \delta \geq \hat{\delta}. \end{cases}$$

Substituting this into (47) yields  $\theta^*(\delta)$ .

$$\theta^*(\delta) = [1 - G(t^*(\delta), \delta)] t^*(\delta) = \begin{cases} \frac{1}{2} t^*(\delta) & ; \delta \leq \hat{\delta} \\ \frac{(t^*)^2}{2\delta} \log\left(\frac{\mu + \delta}{t^*}\right) = \frac{\mu - \hat{\delta}}{\mu + \hat{\delta}} \log\left(\frac{\mu + \hat{\delta}}{\mu - \hat{\delta}}\right) \frac{(\mu + \delta)^2}{2\delta} & ; \delta \geq \hat{\delta}. \end{cases}$$

For  $\delta \geq \hat{\delta}$ ,  $t^*(\delta)$  is strictly increasing, and it is easy to confirm that  $\theta^*(\delta)$  is strictly decreasing in  $\delta \in (0, \mu)$ . Finally, we show that  $t^*(\delta)$  is strictly decreasing in  $\delta \in (0, \hat{\delta})$ . Differentiating  $\frac{\delta}{\log\left(\frac{\mu + \delta}{\mu - \delta}\right)}$  with respect to  $\delta$ , and substituting for  $y = \frac{\mu + \delta}{\mu - \delta}$  yields

$$\frac{d}{d\delta} \left( \delta / \log\left(\frac{\mu + \delta}{\mu - \delta}\right) \right) < 0, \text{ for } \delta \in (0, \mu) \Leftrightarrow 2 \log(y) - y + \frac{1}{y} < 0, \text{ for } y > 1.$$

The latter inequality is true because its derivative is  $-(1 - 1/y)^2 < 0$ .  $\square$