Protest tactics and organizational structure

Elaine Yao*

September 30, 2024

Work in progress. For the latest draft, click here.

Abstract

How do protest movements collectively choose between peaceful or violent tactics over time? I study a dynamic regime change game in which individuals decide whether to participate in a protest, given the tactic chosen by previous protesters. I assume that the power of peaceful tactics is more reliant on turnout than violent tactics. I demonstrate that tactical choice involves both prospective and retrospective considerations. On the one hand, tactical choice responds to information learned from past protests, reflect a process of learning and experimentation. On the other hand, tactical choice must manage public optimism to preserve future opportunities for collective action. Failed protests can undercut public optimism, reducing the efficacy of future tactics and curtailing the possibilities for further contention. The repeated failure of protests, moreover, presents movements with a tactical dilemma. While violence may seem attractive since it is less dependent on turnout, these failures may indicate that violence is futile, leading movements to gamble on peace in hopes of securing high turnout by chance. The game also offers a solution to the problem of equilibrium multiplicity in repeated global games. Individuals' per-period participation decisions are cutoff in idiosyncratic participation costs, while beliefs about the regime are commonly held across periods.

*Ph.D candidate, Princeton University. E-mail: eyao@princeton.edu

I thank Germán Gieczewski and Matias Iaryczower for support and guidance throughout this project. This project has benefited from feedback and suggestions from seminar participants at Princeton, MPSA, and the 2024 EITM Summer Institute.

1 Introduction

What drives protests to choose and change tactics? Existing research has documented the structural, behavioral, and external factors that explain how different methods of resistance weaken and overcome regimes, when these methods are more or less likely to succeed, and how they respond to regime repression (Ackerman and DuVall 2000; Chenoweth and Stephan 2011; Cunningham 2013; Dornschneider-Elkink and Henderson 2024; Schock 2005; Sharp 1973; Sutton, Butcher and Svensson 2014). What has been comparatively sidelined is how tactical choice influences, and influenced by the ongoing interplay of past protest outcomes, public opinion, and future prospects of contestation. This dynamic is usually treated problematic, as it means that endogeneity and path dependence are inevitable concerns for research seeking external explanations for the success and outcomes of tactical choice. However, the fact that tactical choices, public opinion, and protest outcomes are tightly interwoven is more than an inconvenience: The churn of these forces is at the heart of any social movement, and it is essential to understand how they shape behavior and decisionmaking.

This paper systematically unravels how this interplay generates *dynamic* and *organizational* incentives behind protests' choices of tactics. I develop a framework in which public opinion, observed mobilization, and tactical decisions are tightly coupled, and influence one another over time. I specifically focus on the choice between peaceful tactics, which are most powerful when there is massive turnout, and violent or unconventional tactics, which are less dependent on mass participation for power and visibility.

Two sets of incentives emerge from the model. The first set is *organizational*, and reflects incumbent protesters' desire to retain control of the movement and to protest using tactics in line with their ideology and personal preferences. The second set of incentives is *dynamic and success-oriented,* and reflects protesters' incentives to use tactics which are not only likely to achieve success today, but also facilitate chances of future success in the case of a failure or setback today. Success-oriented incentives therefore involve both backwards- and forwards-looking considerations. On one hand, protesters are able to learn from their past attempts and failures, and can pick tactical choices based on the knowledge they have gained from contesting the regime in the past. On the other hand, failures today inevitably impact public opinion tomorrow – and hence the efficacy of tactics tomorrow. I show that protesters sometimes sacrifice having a more powerful protest today so as to not undercut the possibility of a strong protest tomorrow: in other words, that tactics are sometimes chosen to manage public optimism.

The incentives and dynamics which emerge from the model mirror the concerns, conflicts, and choices experienced by social movements in the real world. Importantly, they reflect the fact that

protests which do not quickly and easily accomplish their goals often wrestle with the decision of switching to tactics at odds with the convictions of existing membership. The African National Congress (ANC), for instance, initially contested apartheid using peaceful and passive tactics, but its leadership made the intentional turn to armed violence when nonviolent demonstrations met with severe regime repression and failed to result in change. The next three decades of contention were dominated by violent attacks, vandalism, and bombings, until the South African government ultimately agreed to enter into negotiations which ultimately ended apartheid. Instances where organizations considered but ultimately resisted a transition to violence are also interesting, especially as they help shed light on how organizational incentives may conflict with success-oriented incentives: Internecine conflict over the primacy of nonviolence and organizational structure of the movement ultimately resulted in the dissolution of the Student Nonviolent Coordinating Committee (SNCC), a major organization in the American civil rights movement. These examples suggest that the ability of organizations to shift and experiment with tactics is constrained by incumbents' desire to succeed on their own terms – and to retain control of the movement in the future. However, as the model also predicts, violence is not always a terminal or permanent state. Movements which begin with violent tactics are sometimes is superceded by peaceful civil resistance, as was the case with several episodes of Latin American resistance in the 1980s (Carter 2009).

My model spotlights a logic behind these changes which originates from how protests learn from their past experiences and seek to manage their future prospects. A key implication is that *order in which tactics are used matters*: A protest that begins with violent tactics faces a very different set of prospects than a protest that begins with peaceful tactics. Peaceful "people power" tactics are more dependent on turnout for success, and are therefore also more sensitive to changes in public opinion. A protest which uses violence (which is less sensitive to public opinion) early on essentially squanders its best chance at a powerful peaceful protest. Protesters therefore may eschew violent tactics in early phases of the movement even if violence is expected to be very powerful. It is only when public opinion is eroded by repeated failures and regime repression that violence becomes an optimal choice, not only for the protest today but for its chances in the future. This logic offers alternative explanation for why several empirical studies find that nonviolent tactics seem to more frequently result in successful protest outcomes (Ackerman, Karatnycky et al. 2005; Celestino and Gleditsch 2013; Chenoweth and Stephan 2011). If, as the model suggests, violence tends to be chosen when beliefs are already quite optimistic, then it will be used situations where a successful outcome is already very unlikely.

However, the model emphasizes that violence is not the inevitable choice after failures accumulate and pessimism builds. The reason relates to the fact that protesters learn from what has previously been tried. When protesters learn that the regime is strong, they may also conclude

that the violent or unconventional tactics available to them are simply not strong enough to present any real challenge: It is only a massive peaceful protest which has any change of successfully triggering regime change. Movements may thus choose peaceful tactics in hopes that a stroke of luck – a politically charged national holiday, international intervention, or otherwise – will give them the massive turnout that they need. In other words, the outcome of learning and experimentation may be a return to nonviolent tactics. The conclusions of the model echo scholarship that asserts that violence, rather than being the product of a "terrorist personality," religious fanaticism, or heightened emotional states, is a strategic choice employed by actors motivated by largely the same concerns as actors who choose nonviolence resistance (Dornschneider 2016).

On a technical level, this paper contributes to a growing literature which uses global games in order to model protests and other collective action problems. The key innovation of this paper is that citizens do not receive a noisy signal of regime strength; rather, their per-period cost of participation is idiosyncratically drawn. As a result, individual participation decisions are cutoff in their period-specific participation cost, while beliefs about regime strength are commonly held. Hence, given a realization of cost shocks over the course of the game, there is a unique equilibrium. This is an appealing feature for this paper because it allows me to derive clean equilibrium predictions about a movement that can engage in repeated protests.

The draft proceeds as follows: Section 2 discusses related literature. Section 3 presents the model. Section 4 discusses equilibrium individual participation decisions. Section 5 presents a toy example which clarifies prospective and retrospective success-maximizing considerations. In Section 6, I offer preliminary comments on full model results, extensions, and conclusions.

2 Related literature

This paper connects two major bodies of substantive literature. These ask, respectively, how protests affect public opinion and citizens' information about the regime, and how protests' tactical choices translate into success or failure. In the first category are papers, beginning with Lohmann (1994), which examine how protests *reveal information* about the regimes they contest – in particular, revealing that regimes are more fragile or vulnerable than protesters previously thought. More recent evidence is mixed about the consequences of such "informational protests." Tertytchnaya and Lankina (2020) study contemporaneous public opinion over the course of electoral protests in Russia. They find that while public opinion initially shifted in favor of public demands, media coverage and regime repression ultimately decreased public support for the movement's demands. By contrast, Pop-Eleches, Robertson and Rosenfeld (2022) find that participation in Euromaidan protests induced participants to be *more* sympathetic to protest demands

and frames, shifting public opinion in favor of the protest. These and other studies use snapshots of public opinion gathered at a few discrete points before, during, or after protests – while they offer suggestive evidence that the success of protest movements influences public opinion after the fact, they cannot capture the continuous evolution of public opinion during protests, and how that in turn influenced how protests chose to contest regimes.

Related work on how repression suppresses or inflames dissent offers similarly mixed conclusions, due in part to the fact that, as Ritter and Conrad (2016) observe, both repression and dissent are chosen strategically: governments and dissidents act not only in response to one another's realized actions, but in anticipation of what one other *may* do. Chiang (2021) finds that repression, particularly physical repression, deters participation in nonviolent movements, but does not find a clear effect of repression on participation in violent movements – results consistent with a model where participation in peaceful protests is more sensitive to beliefs about the strength of the regime. Butcher and Pinckney (2022) points out, as my model shows, that participation is endogenous to participants' beliefs about the regime, so that "large protest sizes are almost certainly indicative of the widely-shared expectation of government concessions." When controlling for this endogenous effect, they find that the positive relationship between protest size and government concessions goes away.

The second body of literature encompasses a decades-long effort to explain and recommend how nonviolence, in particular, can be used to present a serious challenge to the coercive power of regimes (Ackerman, Karatnycky et al. 2005; Ackerman and DuVall 2000; Lyall and Wilson 2009; Pape 2003, 2005; Schock 2005; Sharp 1973). This literature offers a rich and nuanced view of the mechanisms by which specific techniques of violent or nonviolent resistance attract participants and undercut the ways that regimes maintain power. My approach to violence and nonviolence mirrors that of Chenoweth and Stephan (2011): I assume that the key difference between these tactics is their reliance on *mass* participation to generate force. Peaceful tactics are fundamentally "people power" tactics which rely *exclusively* on turnout to produce power which puts pressure on regimes to yield to their demands. By contrast, violent tactics are less reliant on turnout. Similarly to the difference between conventional and irregular warfare, violent tactics depend on a relatively small number of people who can conduct activities which are disproportionately visible and disruptive, drawing wider attention to the excesses of a regime or directly sabotaging key elements that prop up the regime.

How tactics are chosen to strategically in response to evolving beliefs is a question that has been mostly taken up by some scholars of rebel groups, civil wars, and international conflicts (Bueno de Mesquita 2013; Qiu 2022). As these authors observe, rebels' wartime strategies are responsive to their current strength – but are also consequential for the duration of fighting and

the likelihood of reaching a peace settlement. Because these models typically treat governments and rebels as unitary actors, they do not capture the incentives that arise from collective action. In particular, these models do not capture the incentive to manage public opinion, which derives form the fact that collective action is ultimately dependent on individually optimal choices.

This paper is closely related to the literature that models mass protests using global games (De Mesquita 2010; De Mesquita and Shadmehr 2023; Edmond 2013; Gieczewski and Koçak 2024; Morris and Shin 1998, 2003; Morris and Shadmehr 2024). A limitation of this technology, however, is that the uniqueness properties that guarantee uniqueness in the stage game break down in repeated or dynamic extensions (Angeletos, Hellwig and Pavan 2007; Little 2017). In the canonical one-shot global game, agents receive heterogeneous signals about the strength of the regime. As Morris and Shin (1998) show, in the unique equilibrium of this game, each agent's optimal action is a best response to the proportion of agents which they believe have a lower signal than they do. In a repeated global game, however, beliefs informed by private signals interact with beliefs informed by the fact that the regime has survived past protests. This interaction produces a multiplicity of equilibria. This motivates my choice to decouple beliefs about the regime from the beliefs about the quantity which underlies participation decisions. In the structure I present, beliefs in each period are commonly held; global games-style heterogeneity in participation *costs* drive individual participation decisions.

Finally, modeling a protest movement as an organization where incumbent members vote on tactics (thereby influencing the composition of future membership) connects this paper to the literature on experimentation in organizations with endogenous membership (Bai and Lagunoff 2011; Gieczewski 2021; Gieczewski and Kosterina 2024). This literature analyzes the long-term consequences of policy choices made by organizations where membership endows voting rights over future policies A common theme of this literature is that sub-optimal policies can become entrenched when the organization is "captured" by relatively extreme membership. This paper is the first, to my knowledge, to use a dynamic model of an organization with endogenous membership to study protests and collective action.

3 Model

Overview. A protest movement exists in discrete time $t = 0, 1, ..., T \leq \infty$. In each period $t \geq 1$ that a protest did not succeed, incumbent protesters choose one tactic to be implemented in $[t, t+1)$. The choice is observed publicly. A participation stage then takes place where all agents can decide whether or not to join the protest in $[t, t+1)$ (there is no priority given to previous incumbents). If their protest is successful, the game ends. If it is unsuccessful, the game proceeds to the next period, where the set of agents who choose to be members at time t constitutes the

new set of incumbent protesters who can choose tactics for the next period. In the beginning of the game, $t = 0$, the game starts with a participation stage where the movement's initial tactic is exogenously specified.

Types. There exist two types of agents, *peaceful* and *violent*. Let τ denote the type of an individual agent (peaceful or violent). I assume there is a unit mass of each type. An agent's cost of taking their *aligned* action (e.g. a violent agent participating in a protest endowed with the violent tactic) is $c_{it}^{\tau} = c_t^{\tau} + \epsilon_{it}^{\tau}$. The two components of cost, which are both drawn i.i.d. each period, are a population shock $c_t^{\tau} \sim Unif[\underline{c}, \bar{c}]$ which affects all agents of a given type equally, and an individual shock $\epsilon_{it}^{\tau} \sim Unif[-\epsilon, \epsilon]$, where $\epsilon > 0$. An agent's cost of taking their unaligned action is infinite. As such, a peaceful agent will only ever join when the peaceful tactic has been chosen, and likewise for a violent agent.

Tactics and success. Each tactic has a protest production function which transforms turnout, ℓ, into protest power. Turnout $\ell \in [0, 1]$ is measured as the fraction of all aligned agents who show up. I denote the production functions by $A_V(\ell)$ and $A_P(\ell)$, and assume that $A_V(0) > A_P(0), A_V(1) < A_P(1), A_V'(\ell) < A_P'(\ell)$. The strength of the regime is given by a state variable θ. Before the game begins, $\theta \sim Unif[m, M]$, which also constitutes the common prior held by agents. If $A_P(\ell) > \theta$, the protest succeeds with probability ρ_P, and if $A_V(\ell) > \theta$ it succeeds with probability ρ_V. Let $\mathcal{W}_t$ denote whether a t-period protest succeeds. An individual's flow payoff is

$$u_{it}^{\tau} = \underbrace{(-c_{it}^{\tau})\mathbb{1}_{\{a_{it}=1\}}}_{\text{Participation cost}} + \underbrace{\mathbb{1}_{\{a_{it}=1 \cap \mathcal{W}_t=1\}}}_{\text{Private benefits}} + \underbrace{\nu \mathbb{1}_{\{\mathcal{W}_t=1\}}}_{\text{Public benefits}}$$

and their total payoff is the discounted sum of their flow payoffs $u_i = (1-\delta)\sum_{t=0}^{\infty} \delta^t u_{it}$.

The timing of each period proceeds as follows:

1. The protest's tactic is inherited from the last period's choice.
2. Costs are drawn. Individual agents choose whether or not to join the protest.
3. Protest occurs. Agents observe turnout and the outcome. Flow payoffs are realized.
4. If the protest succeeds, the game ends. If not, current organization members vote on next period's tactics. Repeat from step 1.

The solution concept is Perfect Bayesian Equilibrium.

4 Equilibrium participation decisions

I first demonstrate foundational results about per-period individual participation decisions. Taking tactical choices and beliefs as given, I will demonstrate that individuals' participation decisions follow a turnout rule in each period: that is, an agent chooses to protest if their cost c_{it}^{τ} is less than some threshold $\hat{c}_t^{\tau}$. I then show that $\hat{c}_t^{\tau}$ is weakly decreasing over time, the consequence of increasing pessimism about θ brought on by repeated failures. Although each tactic's threshold is decreasing, however, aggregate turnout may not. Because participation costs are re-drawn each period and tactical choices may change, either one or both of these factors could drive an increase in turnout from one period to the next even if the threshold decreases.

These results provide the close link between beliefs and participation behavior. In brief summary, the realization of cost shocks combined with prior beliefs determines turnout. Turnout, produced with the chosen tactic, produces protest power, which provides the information upon which beliefs are updated and induces path dependence in the model. Since this entire process hinges upon the realization of cost shocks, the path of the population-level costs generates path dependence in the model.

Beliefs in any given period about θ are simply the posterior after viewing any number of rounds of failed protests. The variation which produces "Laplacian" beliefs is over private *costs:* players' choice to protest is a best response to the proportion of agents who they expect to have a lower participation *cost* than they do. To see this result, note that an agent aligned with the current tactic possessing cost c_{it}^{τ} protests if

$$\begin{aligned} &\mathbb{P}\left(\mathcal{W}_t = 1\right)(1+\nu) - c_{it}^{\tau} > \mathbb{P}\left(\mathcal{W}_t = 1\right)(\nu) \\ \iff &\mathbb{P}\left(\mathcal{W}_t = 1\right) > c_{it}^{\tau} \end{aligned}$$

Let F_t denote the CDF of the (public) posterior belief about θ at time t. At the point of indifference between participating and not participating,

$$\begin{aligned} \mathbb{P}\left(\mathcal{W}_t = 1\right) &= c_{it}^{\tau} \\ \rho_{\tau} \int_{\tilde{c}} F_t\Big(A_{\tau}(\ell(\tilde{c}))\Big) g(\tilde{c}|c_{it}^{\tau}) d\tilde{c} &= c_{it}^{\tau} \end{aligned}$$

Here, $g(\tilde{c}|c_{it}^{\tau})$ denotes a player's belief about $\tilde{c}$, (the population-level cost) conditional on their own observed c_{it}^{τ}. $\ell(\tilde{c})$ is the proportion of the population that would mobilize if the population cost was indeed $\tilde{c}$. Recall that $A_{\tau}(.)$ indicates the tactic-specific protest production function. Therefore, the condition is the average expected *success* of all possible realizations of protest power induced by possible cost shocks. We can simplify this expression substantially using standard global games properties. We know that the marginal agent who is indifferent between participation and

non-participation believes that the proportion of agents with cost less than hers is distributed $Unif[0,1]$. Therefore, we can abstract away from explicit beliefs about costs and shift instead to beliefs directly about turnout, ℓ:

$$\rho_\tau \int_0^1 F_t(A_\tau(\ell))dl = c_{it}^\tau \quad \equiv \hat{c}_t^\tau \tag{1}$$

This condition yields a per-period cutoff participation rule: an agent protests in time t iff his cost $c_{it}^\tau < \hat{c}_t^\tau$. The key element is F_t, agents' commonly held posterior beliefs about θ, which is what is used to compute the expected success of a protest produced with tactic τ and with mobilization l. Changes in the cost threshold over time therefore directly reflect updates in posterior beliefs. Beliefs are only ever updated with bad news. Since a success ends the game, beliefs are only ever updated in the case of failure. As a result, F_t at each period first-order must (weakly) stochastically dominates F_{t-1}, placing more and more (relative) weight on higher values of θ. The consequence is that the average expected success over all possible realization of protests must fall – that is, $\hat{c}_t^\tau$ must fall. These results are summarized in the following statement:

Lemma 1. *An individual with cost c_{it}^τ protests in period t if*

$$c_{it}^\tau < \rho \int_0^1 F_t\Big(A_\tau(\ell)\Big)d\ell \equiv \hat{c}_t^\tau$$

Furthermore, $\hat{c}_t^\tau \leq \hat{c}_{t-1}^\tau \ \forall t$.

Note that the statement $\hat{c}_t^\tau \leq \hat{c}_{t-1}^\tau \ \forall t$ applies *within* a tactic – it does not make a statement about the relationship *between* the different tactical thresholds, i.e. for a given t, whether $c_{it}^p > c_{it}^v$ or vice versa.

5 Success-maximizing tactical choice

I now demonstrate how the structure of equilibrium participation maps onto tactical decisions. The aim of this section is to clarify the intertemporal incentives to make *success-maximizing* tactical decisions. The game as it is written contains many other possible incentives: for instance, individual incentives to acquire private participation benefits, or conversely to free-ride off the costly participation of others. This sections abstracts away from those incentives to isolate how tactics can be chosen to maximize the movement's chances at success. To this end, I focus on the decision of a social planner who is solely concerned with maximizing success across the total (two-period) lifetime of the protest. The social planner does not discount, and does not care about the payoffs incurred by individual agents. The social planner can choose tactics in both periods, but cannot control turnout.

I show that there are, broadly, two forces that affect the social planner's choice: retrospective and prospective incentives. The retrospective incentives are strongly reminiscent of the experimentation literature: essentially, the more accurate the information that the social planner has about the regime, the better able he is to make an "accurate" tactical choice. The prospective incentive originates from the coordination dynamic. It reflects the fact that tactics' (expected) power in the future depends on beliefs that will be informed by what happens today. Since the social planner is not myopic, he therefore has incentive to "manage public opinion" so as to not sabotage his chances of succeeding in the future if there is a failure today. Note that while an experimentation incentive would be present even without coordination concerns, the latter category of incentives – to preserve future chances at contesting the regime – is purely driven by coordination concerns. The social planner knows that citizens will learn from what happens today, and that that learning will affect their turnout decisions, changing the potential of both tactics. In short, coordination means that the tactics available to use tomorrow are not the same tactics available to use today.

One temporary simplification will make it possible to cleanly isolate these two incentives: assume that the output of violence is *independent of turnout*. The reason this works is because retrospective ("experimental") and prospective ("coordination") concerns map onto uncertainty over θ and uncertainty over turnout, respectively. The tactical assumption made up until now is that the peaceful tactic is *more* sensitive to turnout than the violent tactic. The simplification is simply to make this difference stark: to assume that turnout uncertainty does not apply at all to the violent tactic. As a result, we can simply consider the choice between the violent tactic whose success is subject *only* to regime strength uncertainty, and the peaceful tactic whose success depends on both regime strength uncertainty and turnout uncertainty (peace). The result is starkly illustrated in the following lemma:

Lemma 2. *Suppose that $A_v(\ell) = \bar{v}\ \forall \ell, A_p(\ell)$ is linear, and that there is no backfiring ($\rho_P = \rho_V = 1$). Consider the choice of a success-maximizing social planner who does not discount. The following facts are true:*

(i) The social planner will never use violence in the first period.

(ii) The range of failed first-period protests for which violence is used in the second period is increasing in $\bar{v}$.

These two facts isolate how prospective and retrospective concerns map onto success-maximizing tactical choice. The first fact – that violence is never used in the first period – comes directly from *prospective* concerns. The reason for this is that turnout in the second period is affected by the realization of the first-period protest, and therefore the order in which tactics are used is not neutral. A failed first-period violent protest deterministically reduces the expected power of peace

in the second period. However, a failed first-period peaceful protest has no effect on the power of violence in the second period. This is true for *any* level of $\bar{v}$, even if $\bar{v}$ is very high. Hence, forward-looking concerns can lead the social planner to make choices which are not myopically optimal: even if violence has much higher expected success than peace in the first period, using it in the first period takes away the chance that a large peaceful turnout will be able to overcome a regime too strong to be defeated even by the strong violent technology.

This emphasizes the importance of choosing tactics to "manage public optimism." Since the power of the peaceful tactic is sensitive to turnout, it is also vulnerable to being impacted by public posteriors about the regime formed after the first protest. Using violence first essentially *guarantees* that, in the case of failure, the "peaceful tactic" available to tomorrow will be weaker in expectation than the peaceful tactic available today. By contrast, since the powerful violence is not sensitive to turnout in this simplified example, it offers, essentially, a risk-free guarantee of protest of power $\bar{v}$. Therefore, forward-looking concerns are directly connected to the *collective action* dynamic established in Section 4. Because tactics depend (to varying extents) on collective action to generate power, the versions of tactics available at time $t+1$ are *not* the same versions of tactics available at time t – and that this impacts optimal tactical choice at time t.

The second part of the lemma deals with retrospective considerations, and isolates the effect of period 1 learning on optimal period 2 choices. This is essentially an *experimentation* result: the more information that is revealed about the state variable, θ, the better-informed that the organizer's choice is. Just because the organizer should eschew violence in the first period does not mean that he necessarily will, or will not use violence in the second period – this depends on the information revealed about the state by the power, and outcome of the period 1 protest. Without backfiring, learning about the state is especially stark: the failure of a protest of size A_1 confirms that $\theta > A_1$.

Figure 1 plots the expected second-period success of each tactic as a function of the power of a (failed) first-period protest. The figure makes several dynamics clear: Holding $\bar{v}$ constant, the larger the (peaceful) first-period protest, the more likely it is that peace will be chosen in the second period. Part of this is mechanical: if the first period protest exceeds $\bar{v}$, then violence has zero chance of success, while peace has a positive chances of success – as such, peace will be chosen in the second period. However, it is not entirely mechanical: even when A_1, the size of the first period protest, is smaller than $\bar{v}$, the social planner has enough doubt that $\bar{v} > \theta$ to choose peace. By contrast, when the first period protest is small and the social planner learns relatively little, violence is a more appealing "safe" option. The more powerful that $\bar{v}$ is, however, the wider the range of failed first-period protests for which violence is chosen the second period. In the figure, when violence is relatively weak ($\bar{v} = 0.5$), there is no failed first-period

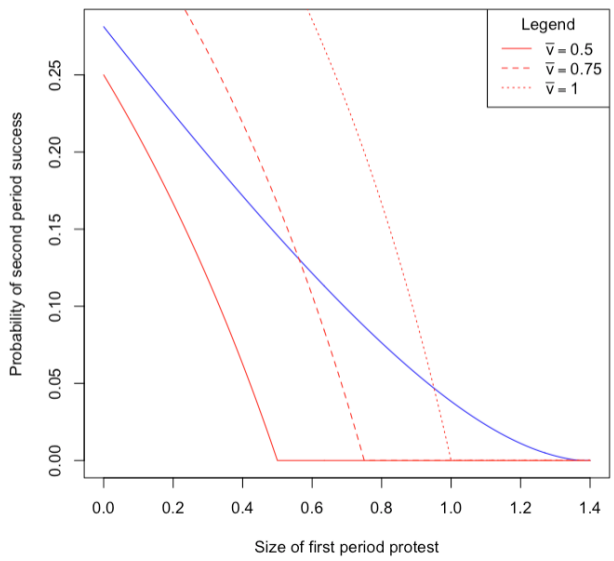

Figure 1: Expected success of peaceful (in blue) and violent (in red) tactics as a function of first period protest power. $A_p(\ell) = 1.5\ell$, different levels of violence displayed in legend. Other starting parameter values are: $\theta \sim Unif[0, 2], \epsilon = 0.1, c_t^p \sim Unif[0, 1]$.

protest for which violence is chosen. As we progressively increase the power of violence to 0.75, then to 1, so too we increase the range of histories for which violence is chosen in the second period.

The existence of a collective action dynamic, however, sets this model apart from existing models of experimentation. Unlike models of experimentation where the technologies or tactics available to experimenter are fixed, a key component of this model is that the potential participants are *also* learning from past protests – not only the experimenter, and adjusting their own participation thresholds in response. Hence, the tactics available for an experimenter to "use" in the future are not identical to the tactics available in the past. The retrospective calculus that may lead organizers to choose peace in the second period is fundamentally different from the prospective calculus that drives organizers to choose peace in the first period.

A dynamic that obvious in this example is the idea that *failures do not necessarily lead towards violence*. The failure of a peaceful protest that is more powerful than $\bar{v}$ teaches protesters that a violent protest is futile. The cleanliness of this result is owed to the starkness of learning in this example. Without backfiring or other considerations that would create more uncertainty in learning, players learn immediately that it is not only unlikely, but indeed *impossible* that violence can succeed. In the fuller version of the game, where learning is less stark and both tactics are responsive to turnout uncertainty, this dynamic is significantly more complex.

This example also shows how path dependence in the model does not have immediately obvious consequences. In particular, the likelihood of success over the lifetime of the movement can vary non-monotonically with the realization of first-period costs. To illustrate this with a toy example, I continue with the parameterization displayed in Figure 1, with $\bar{v} = 0.5$ (so that violence is never chosen in the second period). In this case, the first-period participation threshold is $\int_0^1 \frac{1.5\ell - 0}{2-0} d\ell = 0.375$. Suppose that the first-period aggregate shock is low enough that all citizens turn out. Since this is the maximally powerful protest, its failure indicates with certainty that it is *impossible* to succeed. Nonetheless, since all plausibly defeatable values of θ would have fallen, there is also no "money left on the table."

Conversely suppose that the first-period aggregate shock is high enough that no citizens turn out. In this case, nothing is learned from the protest, and as a result, expectations of success in the second period are identical to those in the first period: in particular, the second-period attendance participation cost is once again 0.375. Finally, suppose that the cost draw is intermediate. Consider what happens if $c_1^p = \hat{c}_1^p$, so exactly half the population turns out. If this protests fails, then the second period cost threshold is 0 – and in the very best scenario, then only half the population turns out again. The first period experience shows that this is futile, but also that violence with $\bar{v} = 0.5$ is also futile. Indeed, it is this intermediate cost draw that is most detrimental for *total* chances of success over the lifetime of the protest.

6 Preliminary conclusions

Organizational incentives. The natural next step is to extend the example provided in the previous section to consider the decision that would actually be made if peaceful incumbents were installed period 1 and had control over period 2 decisions. Under what conditions would they be willing to cede control of the movement, and under what conditions would they be willing to sacrifice chances of success in order to succeed on their own terms? In the two-period example, the condition under which an incumbent would be willing to change tactics in the second period is:

$$\begin{aligned} F_{t+1}\Big(\mathbb{E}\left[A_v(l_{t+1})\right]\Big)(\nu) > & F_{t+1}\Big(\mathbb{E}\left[A_p(l_{t+1})\right]\Big)\left[\mathbb{P}\left(c_{i,t+1}^p < \hat{c}_{t+1}^p\right) + \nu\right] \\ & + \mathbb{P}\left(c_{i,t+1}^p < \hat{c}_{t+1}^p\right)\Big(-\mathbb{E}_{t+1}[c_{it}^p | c_{it}^p < \hat{c}_{t+1}^p]\Big) \end{aligned}$$

For reference, the condition for success-maximizing planner to change tactics to violence was simply $F_{t+1}\Big(\mathbb{E}\left[A_v(l_{t+1})\right]\Big) > F_{t+1}\Big(\mathbb{E}\left[A_p(l_{t+1})\right]\Big)$. This comparison provides us with some clarity. Peaceful incumbents' unwillingness to change tactics originates from the likelihood that they will participate in a successful protest and acquire public benefits – the first term of the right-hand side, but is curbed by cost that necessarily accompanies participation. Since participants never show up if the cost is higher than the expected benefits, it is clear that this calculus tilts in the

favor of recalcitrance – failure to change tactics even in some instances where it would be success-maximizing. Hence, it is a reasonable conjecture that *incumbents respond less to retrospective incentives than a success-maximizing social planner*. The desire to acquire private benefits dilutes incumbents' ability to capitalize off learning and make decisions maximally conducive to a social outcome. Of course, a decrease in success probability is not identical in this model to a decrease in *global welfare* – and a further task of this exercise is to characterize precisely the degree of success probability loss and a comparison in welfare. This portion of the exercise, once proven, offers preliminary and suggestive evidence regarding the difference between movements with decentralized decisionmaking structures and those with explicitly defined leadership. However, a more complete exercise, which looks at leaders with more realistic goals (e.g. those who are biased towards one tactic or the other, or care more about participants' welfare than success) in the context of the full model, is necessary to make definitive statements.

After completing the exercise introduced in Section 5, the next task is to provide a characterization of all three sets of incentives – prospective, retrospective, and organizational – map onto equilibrium behavior in a full version of the model with all the elements described in Section 3. With this characterization complete, it will be possible to look at situations where some of these incentives reinforce one another – such as cases when both retrospective and prospective considerations both push organizers towards initial usage of peaceful tactics – and when they conflict – such as organizational incentives interfering with retrospective incentives. It will then also be possible to isolate effects on likelihood of success, as well as implications for citizens' welfare.

Several comparative statics and extensions will then also be possible. For instance, how does tactical choice change when public benefits are large relative to private benefits, or vice versa? It is also possible to look at systematic differences between the two populations: what is the result when the distribution of possible cost shocks for peaceful and violent engagement are very different, as opposed to relatively similar? What about when one type has a larger population than the other (rather than both being a unit mass)? Finally, a number of interesting and realistic extensions to the model are worth investigating. For instance, the model provides a natural framework for understanding how multiple rebel groups might interact when collectively contesting the same cause. One can imagine, for instance, a simple extension when two groups with different fundamentals organize subsequent protests, learning from their counterparts' choices as well as their own. This extension has the potential to address questions on, for instance, the emergence of violent flanks in protest movements.

References

Ackerman, Peter, Adrian Karatnycky et al. 2005. "How freedom is won: From civic resistance to durable democracy." *New York, Freedom House* .

Ackerman, Peter and Jack DuVall. 2000. *A force more powerful: A century of non-violent conflict*. Palgrave Macmillan.

Angeletos, George-Marios, Christian Hellwig and Alessandro Pavan. 2007. "Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks." *Econometrica* 75(3):711–756.

Bai, Jinhui H and Roger Lagunoff. 2011. "On the faustian dynamics of policy and political power." *The Review of Economic Studies* 78(1):17–48.

Bueno de Mesquita, Ethan. 2013. "Rebel tactics." *Journal of Political Economy* 121(2):323–357.

Butcher, Charles and Jonathan Pinckney. 2022. "Friday on my mind: Re-assessing the impact of protest size on government concessions." *Journal of Conflict Resolution* 66(7-8):1320–1355.

Carter, April. 2009. People Power and Protest: The Literature on Civil Resistance in Historical Context. In *Civil resistance and power politics: the experience of non-violent action from Gandhi to the present*, ed. Adam Roberts and Timothy Garton Ash. Oxford University Press chapter 2, pp. 25–42.

Celestino, Mauricio Rivera and Kristian Skrede Gleditsch. 2013. "Fresh carnations or all thorn, no rose? Nonviolent campaigns and transitions in autocracies." *Journal of Peace Research* 50(3):385–400.

Chenoweth, Erica and Maria J Stephan. 2011. *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press.

Chiang, Amy Yunyu. 2021. "Violence, non-violence and the conditional effect of repression on subsequent dissident mobilization." *Conflict Management and Peace Science* 38(6):627–653.

Cunningham, Kathleen Gallagher. 2013. "Understanding strategic choice: The determinants of civil war and nonviolent campaign in self-determination disputes." *Journal of Peace Research* 50(3):291–304.

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

De Mesquita, Ethan Bueno and Mehdi Shadmehr. 2023. "Rebel motivations and repression." *American Political Science Review* 117(2):734–750.

Dornschneider-Elkink, Stephanie and Nick Henderson. 2024. "Repression and dissent: How tit-for-tat leads to violent and nonviolent resistance." *Journal of Conflict Resolution* 68(4):756–785.

Dornschneider, Stephanie. 2016. *Whether to kill: The cognitive maps of violent and nonviolent individuals.* University of Pennsylvania Press.

Edmond, Chris. 2013. "Information manipulation, coordination, and regime change." *Review of Economic studies* 80(4):1422–1458.

Gieczewski, Germán. 2021. "Policy persistence and drift in organizations." *Econometrica* 89(1):251–279.

Gieczewski, Germán and Korhan Koçak. 2024. "Collective Procrastination and Protest Cycles." *American Journal of Political Science (Forthcoming)* .

Gieczewski, Germán and Svetlana Kosterina. 2024. "Experimentation in Endogenous Organizations." *Review of Economic Studies* 91(3):1711–1745.

Little, Andrew T. 2017. "Coordination, learning, and coups." *Journal of Conflict Resolution* 61(1):204–234.

Lohmann, Susanne. 1994. "The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989–91." *World Politics* 47(1):42–101.

Lyall, Jason and Isaiah Wilson. 2009. "Rage against the machines: Explaining outcomes in counterinsurgency wars." *International Organization* 63(1):67–106.

Morris, Stephen and Hyun Song Shin. 1998. "Unique equilibrium in a model of self-fulfilling currency attacks." *American Economic Review* pp. 587–597.

Morris, Stephen and Hyun Song Shin. 2003. Global games: Theory and applications. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 1.* Cambridge University Press pp. 56–114.

Morris, Stephen and Mehdi Shadmehr. 2024. "Repression and Repertoires." *American Economic Review: Insights (Forthcoming)* .

Pape, Robert. 2003. "The strategic logic of suicide terrorism." *American Political Science Review* 97(3):343–361.

Pape, Robert. 2005. *Dying to win: The strategic logic of suicide terrorism.* Random House.

Pop-Eleches, Grigore, Graeme Robertson and Bryn Rosenfeld. 2022. "Protest participation and attitude change: evidence from Ukraine's Euromaidan revolution." *The Journal of Politics* 84(2):625–638.

Qiu, Xiaoyan. 2022. "Rebel strategies and the prospects for peace." *American Journal of Political Science* 66(1):140–155.

Ritter, Emily Hencken and Courtenay R Conrad. 2016. "Preventing and responding to dissent: The observational challenges of explaining strategic repression." *American Political Science Review* 110(1):85–99.

Schock, Kurt. 2005. *Unarmed insurrections: People power movements in nondemocracies*. Vol. 22 U of Minnesota Press.

Sharp, Gene. 1973. *The Politics of Nonviolent Action*. Porter Sargent.

Sutton, Jonathan, Charles R Butcher and Isak Svensson. 2014. "Explaining political jiu-jitsu: Institution-building and the outcomes of regime violence against unarmed protests." *Journal of Peace Research* 51(5):559–573.

Tertytchnaya, Katerina and Tomila Lankina. 2020. "Electoral protests and political attitudes under electoral authoritarianism." *The Journal of Politics* 82(1):285–299.

Appendix

A Proofs

A.1 Proof of Lemma 1

Proof. Recall that we defined

$$\hat{c}_t^{\tau} = \rho_\tau \int_0^1 F_t(A_\tau(l))dl$$

where A_τ is an increasing function of l and F_t denotes the CDF of the public belief about θ with which agents make their participation decision in time t.

A protest of size l succeeds if $A_\tau(l) > \theta$. Suppose F FOSD F'. Then, for some fixed a, $F(a) < F'(a)$. This must be true of any size of mobilization a. Hence, in order to show that $\hat{c}_{t+1} < \hat{c}_t$, it suffices to show that F_{t+1} FOSD F_t $\forall t$.

I proceed by induction on t.

Base case: $t = 1$. At $t = 0, F_0(\theta) = \frac{\theta - m}{M - m}$. After one failed protest of size $A_1 \in [m, M]$:

$$F_1(\theta) = \begin{cases} \frac{(1-\rho)(\theta-m)}{(1-\rho)(A_1-m)+(M-A_1)} & \text{if } \theta \in (m, A_1) \\ \frac{\theta-A_1}{(1-\rho)(A_1-m)+(M-A_1)} + \frac{(1-\rho)(A_1-m)}{(1-\rho)(A_1-m)+(M-A_1)} & \text{if } \theta \in (A_1, M) \end{cases}$$

We will show that in both cases, $F_0 > F_1$. For the first case,

$$\frac{(1-\rho)(\theta-m)}{(1-\rho)(A_1-m)+(M-A_1)} < \frac{\theta-m}{M-m} \iff \rho > \rho\left(\frac{A_1-m}{M-m}\right)$$

which is true. For the second case,

$$\begin{aligned} &\frac{\theta-A_1}{(1-\rho)(A_1-m)+(M-A_1)} + \frac{(1-\rho)(A_1-m)}{(1-\rho)(A_1-m)+(M-A_1)} \\ =&\frac{(\theta-A_1)+(1-\rho)(A_1-m)}{(1-\rho)(A_1-m)+(M-A_1)} < \frac{\theta-m}{M-m} \\ \iff &\frac{M-m}{\theta-m} > 1 \end{aligned}$$

which is true for all $\theta \in [A_1, M]$.

Inductive step: Prior to the start of period $t+1$ there exists some past sequence of sizes of past (unsuccessful mobilizations): $A_\tau(l_1), A_\tau(l_2), ..., A_\tau(l_t)$. Order these from **smallest to**

largest and denote them $b_1, ..., b_t$. Remove any mobilizations from the sequence that less than m, since these do not affect beliefs. Thus, we have some non-decreasing sequence of mobilizations of length s, where $t \geq s \geq 1$. Add the cut-points of the support of the distribution θ, so the non-decreasing sequence of intervals consists of $m, \{b\}_{r=1}^{s}, M$.

The density of the belief f_t about θ is

$$f_t(\theta) = \begin{cases} \frac{(1-\rho)^s}{(b_1-m)(1-\rho)^s+(b_2-b_1)(1-\rho)^{s-1}+\cdots+(b_s-b_{s-1})(1-\rho)+(M-b_s)} & \text{if } \theta \in [m, b_1] \\ \frac{(1-\rho)^{s-1}}{(b_1-m)(1-\rho)^s+(b_2-b_1)(1-\rho)^{s-1}+\cdots+(b_s-b_{s-1})(1-\rho)+(M-b_s)} & \text{if } \theta \in [b_1, b_2] \\ ... & \\ \frac{1}{(b_1-m)(1-\rho)^s+(b_2-b_1)(1-\rho)^{s-1}+\cdots+(b_s-b_{s-1})(1-\rho)+(M-b_s)} & \text{if } \theta \in [b_s, M] \end{cases} \tag{A1}$$

where the generic interval $b_r - b_{r-1}$ has density

$$\frac{(1-\rho)^{s-(r-1)}}{(b_1-m)(1-\rho)^s+(b_2-b_1)(1-\rho)^{s-1}+\cdots+(b_s-b_{s-1})(1-\rho)+(M-b_s)}$$

The corresponding CDF is

$$F_t(\theta) = \begin{cases} \frac{(1-\rho)^s(\theta-m)}{(b_1-m)(1-\rho)^s+(b_2-b_1)(1-\rho)^{s-1}+\cdots+(b_s-b_{s-1})(1-\rho)+(M-b_s)} & \text{if } \theta \in [m, b_1] \\ \frac{(b_1-m)(1-\rho)^s+(\theta-b_1)(1-\rho)^{s-1}}{(b_1-m)(1-\rho)^s+(b_2-b_1)(1-\rho)^{s-1}+\cdots+(b_s-b_{s-1})(1-\rho)+(M-b_s)} & \text{if } \theta \in [b_1, b_2] \\ \vdots & \\ \frac{(b_1-m)(1-\rho)^s+(b_2-b_1)(1-\rho)^{s-1}+...+(\theta-b_s)}{(b_1-m)(1-\rho)^s+(b_2-b_1)(1-\rho)^{s-1}+\cdots+(b_s-b_{s-1})(1-\rho)+(M-b_s)} & \text{if } \theta \in [b_s, M] \end{cases} \tag{A2}$$

Consider now the mobilization that is initiated at time $t+1$:

$$A_\tau\left(\frac{\hat{c} - (c_t^\tau - \epsilon)}{2\epsilon}\right) \equiv b'$$

If $b' < m$ then agents learn nothing and beliefs do not change from one period to the next. If $b' > M$, replace it with the value of M (the upper bound on the support of θ). If $b' \in (m, M)$,, then it must be that b' is contained within an existing interval. Denote this interval by $[b_k, b_{k+1}]$. Let $\underline{B}$ denote the set of all intervals from $[m, b_1]...[b_{k-1}, b_k]$ and $\bar{B}$ denote the set of all intervals from $[b_{k+1}, b_{k+2}]...[b_s, M]$. Note that the new $(t+1)$ sequence of cutpoints $m, b_1, ..., b_k, b', b_{k+1}, ..., M$ has $s+1$ terms.

I will show that the FOSD relation must hold at least weakly in the interval $[b', b_k + 1]$ and must hold strictly everywhere else.

(1) First consider all intervals in $\underline{B}$. For any $\theta \in [b_j, b_{j+1}] \in \underline{B}$, F_t FOSD F_{t+1} if

$$\frac{(1-\rho)^{j-1}(\theta - b_j) + \cdots + (1-\rho)^s(b_1 - m)}{N_t} > \frac{(1-\rho)^j(\theta - b_j) + \cdots + (1-\rho)^{s+1}(b_1 - m)}{N_{t+1}}$$

where

$$N_t = (b_1 - m)(1-\rho)^s + (b_2 - b_1)(1-\rho)^{s-1} + \cdots + (b_{k+1} - b_k)(1-\rho)^{s-k} + \cdots + (M - b_s)$$
$$N_{t+1} = (b_1 - m)(1-\rho)^{s+1} + (b_2 - b_1)(1-\rho)^s + \cdots$$
$$+ (b' - b_k)(1-\rho)^{s+1-k} + (b_{k+1} - b')(1-\rho)^{s-k} + \cdots + (M - b_s)$$

Note that N_t has one fewer term than N_{t+1}. Rearrange the initial inequality:

$$\iff \frac{N_{t+1}}{N_t} > \frac{(1-\rho)^j(\theta - b_j) + \cdots + (1-\rho)^{s+1}(b_1 - m)}{(1-\rho)^{j-1}(\theta - b_j) + \cdots + (1-\rho)^s(b_1 - m)}$$
$$\iff \frac{N_{t+1}}{N_t} > (1-\rho)\frac{(1-\rho)^{j-1}(\theta - b_j) + \cdots + (1-\rho)^s(b_1 - m)}{(1-\rho)^{j-1}(\theta - b_j) + \cdots + (1-\rho)^s(b_1 - m)}$$
$$\iff \frac{N_{t+1}}{N_t} > (1-\rho)$$

To see why this is true, note that for N_t restricted to terms preceding the "split term", i.e. $(b_1 - m)(1-\rho)^s + \cdots + (b_k - b_{k-1})(1-\rho)^{s-(k-1)}$ it is true that (with some abuse of notation) $(1-\rho)N_t = N_{t+1}$. For N_t restricted to terms following the split, i.e. $(b_{k+2} - b_{k+1})(1-\rho)^{s-(k+1)} + \cdots + (b_1 - m)$, we have (with the same abuse of notation) a strict inequality: $(1-\rho)N_t < N_{t+1}$.

Now I compare the term in N_t that is split into two terms in N_{t+1}. For N_t and N_{t+1} restricted to these terms, first note that

$$\begin{aligned}(1-\rho)N_t &= (b_{k+1} - b_k)(1-\rho)^{s-k}(1-\rho) \\ &= (b_{k+1} - b_k)(1-\rho)^{s-k+1}\end{aligned}$$

By definition, $N_{t+1} = (b' - b_k)(1-\rho)^{s-k+1} + (b_{k+1} = b')(1-\rho)^{s-k}$. This implies that

$$\begin{aligned}N_{t+1} &> (b' - b_k)(1-\rho)^{s-k+1} + (b_{k+1} - b')(1-\rho)^{s-k+1} \\ &= (1-\rho)^{s-k+1}(b' - b_k + b_{k+1} - b') \\ &= (1-\rho)^{s-k+1}(b_{k+1} - b_k) \\ &= (1-\rho)N_{t+1}\end{aligned}$$

We have therefore proved that there must be (weakly greater than zero) terms where $N_{t+1} = (1-\rho)N_t$ and there must be at least one term where $N_{t+1} > (1-\rho)N_t$. This proves the claim that for intervals in $\underline{B}, F_t > F_{t+1}$.

(2) Suppose $\theta \in [b_k, b']$. In this region, F_t FOSD F_{t+1} if

$$\begin{aligned}
\frac{(1-\rho)^{s-k}(\theta - b_k) + \cdots + (1-\rho)^s(b_1 - m)}{N_t} &> \frac{(1-\rho)^{s-k+1}(\theta - b_k) + \cdots + (1-\rho)^{s+1}(b_1 - m)}{N_{t+1}} \\
\iff \frac{N_{t+1}}{N_t} &> \frac{(1-\rho)^{s-k+1}(\theta - b_k) + \cdots + (1-\rho)^{s+1}(b_1 - m)}{(1-\rho)^{s-k}(\theta - b_k) + \cdots + (1-\rho)^s(b_1 - m)} \\
\iff \frac{N_{t+1}}{N_t} &> (1-\rho)
\end{aligned}$$

which was already been proved above.

(3) Suppose $\theta \in [b', b_k + 1]$. I will show that in this region, $F_t \geq F_{t+1}$ if

$$\begin{aligned}
&\frac{(1-\rho)^{s-k}(\theta - b_k) + \cdots + (1-\rho)^s(b_1 - m)}{N_t} \\
&\geq \frac{(1-\rho)^{s-k}(\theta - b') + (1-\rho)^{s-k+1}(b' - b_k) + \cdots + (1-\rho)^{s+1}(b_1 - m)}{N_{t+1}} \\
\iff &\frac{(1-\rho)^{s-k}(\theta - b') + (1-\rho)^{s-k}(b' - b_k) + \cdots + (1-\rho)^s(b_1 - m)}{N_t} \\
&\geq \frac{(1-\rho)^{s-k}(\theta - b') + (1-\rho)^{s-k+1}(b' - b_k) + \cdots + (1-\rho)^{s+1}(b_1 - m)}{N_{t+1}} \\
\iff \frac{N_{t+1}}{N_t} &\geq \frac{(1-\rho)^{s-k}(\theta - b') + (1-\rho)^{s-k}(b' - b_k) + \cdots + (1-\rho)^s(b_1 - m)}{(1-\rho)^{s-k}(\theta - b') + (1-\rho)^{s-k}(b' - b_k) + \cdots + (1-\rho)^s(b_1 - m)} \\
\iff \frac{N_{t+1}}{N_t} &\geq \frac{(1-\rho)^{s-k}(\theta - b') + (1-\rho)\Big[(1-\rho)^{s-k}(b' - b_k) + \cdots + (1-\rho)^s(b_1 - m)\Big]}{(1-\rho)^{s-k}(\theta - b') + \Big[(1-\rho)^{s-k}(b' - b_k) + \cdots + (1-\rho)^s(b_1 - m)\Big]}
\end{aligned}$$

Let $\mathcal{S} \equiv \Big[(1-\rho)^{s-k}(b' - b_k) + \cdots + (1-\rho)^s(b_1 - m)\Big]$. Note that

$$\frac{N_{t+1}}{N_t} = \frac{(M - b_s) + \cdots + (1-\rho)^{s-k}(b_{k+1} - b') + (1-\rho)\mathcal{S}}{(M - b_s) + \cdots + (1-\rho)^{s-k}(b_{k+1} - b') + \mathcal{S}}$$

Restating the problem:

$$\frac{\Big[(M - b_s) + \cdots + (1-\rho)^{s-k}(b_{k+1} - b')\Big] + (1-\rho)\mathcal{S}}{\Big[(M - b_s) + \cdots + (1-\rho)^{s-k}(b_{k+1} - b')\Big] + \mathcal{S}} \geq \frac{\Big[(1-\rho)^{s-k}(\theta - b')\Big] + (1-\rho)\mathcal{S}}{\Big[(1-\rho)^{s-k}(\theta - b')\Big] + \mathcal{S}}$$

Collecting terms, we can restate the problem as:

$$\begin{aligned}
&\frac{\mathcal{M}-(1-\rho)\mathcal{S}}{\mathcal{M}+\mathcal{S}} \geq \iff \frac{\alpha-(1-\rho)\mathcal{S}}{\alpha+\mathcal{S}} \\
\iff &(\mathcal{M}+(1-\rho)\mathcal{S})(\alpha+\mathcal{S}) \geq (\alpha+(1-\rho)\mathcal{S})(\mathcal{M}+\mathcal{S}) \\
\iff &\mathcal{M}\mathcal{S}+\alpha(1-\rho)\mathcal{S} \geq \alpha\mathcal{S}+\mathcal{M}(1-\rho)\mathcal{S} \\
\iff &\mathcal{M}+\alpha(1-\rho) \geq \alpha+\mathcal{M}(1-\rho) \\
\iff &\mathcal{M}(1-1+\rho) \geq \alpha(1-1+\rho) \\
\iff &\mathcal{M} \geq \alpha \\
\iff &(M-b_s)+\cdots+(1-\rho)^{s-k}(b_{k+1}-b') \geq (1-\rho)^{s-k}(\theta-b')
\end{aligned}$$

Since all terms are positive and $\theta \in [b', b_{k+1}]$, we must have

$$(1-\rho)^{s-k}(b_{k+1}-b') \geq (1-\rho)^{s-k}(\theta-b')$$

This holds with "=" if $\theta = b_{k+1}$ and with ">" if $\theta \in [b', b_{k+1})$.

(4) Last, consider all intervals in $\bar{B}$. For any $\theta \in [b_j, b_{j+1}] \in \bar{B}$, F_t FOSD F_{t+1} if

$$\begin{aligned}
&\frac{(1-\rho)^{s-j}(\theta-b_j)+...+(1-\rho)^{s-k}(b_{k+1}-b_k)+...+(1-\rho)^{s}(b_1-m)}{N_t} \\
>&\frac{(1-\rho)^{s-j}(\theta-b_j)+...+(1-\rho)^{s-k+1}(b_{k+1}-b')+(1-\rho)^{s-k+1}(b'-b_k)+...+(1-\rho)^{s+1}(b_1-m)}{N_{t+1}}
\end{aligned}$$

$$\iff \frac{N_{t+1}}{N_t} > \frac{(1-\rho)^{s-j}(\theta-b_j)+...+(1-\rho)^{s-k}(b_{k+1}-b')+(1-\rho)\mathcal{S}}{(1-\rho)^{s-j}(\theta-b_j)+...+(1-\rho)^{s-k}(b_1-b')+\mathcal{S}}$$

Using the same definition of $\mathcal{S}$ that we employed previously, we can restate the problem:

$$\begin{aligned}
&\frac{\Big[(M-b_s)+\cdots+(1-\rho)^{s-k}(b_{k+1}-b')\Big]+(1-\rho)\mathcal{S}}{(M-b_s)+\cdots+(1-\rho)^{s-k}(b_{k+1}-b')\Big]+\mathcal{S}} \\
&> \frac{\Big[(1-\rho)^{s-j}(\theta-b_j)+...+(1-\rho)^{s-k}(b_{k+1}-b')\Big]+(1-\rho)\mathcal{S}}{\Big[(1-\rho)^{s-j}(\theta-b_j)+...+(1-\rho)^{s-k}(b_1-b')\Big]+\mathcal{S}}
\end{aligned}$$

Collecting terms, we can restate this expression as

$$\frac{\mathcal{M}-(1-\rho)\mathcal{S}}{\mathcal{M}+\mathcal{S}} > \frac{\alpha'-(1-\rho)\mathcal{S}}{\alpha'+\mathcal{S}}$$

Which holds whenever $\mathcal{M} > \alpha'$, i.e.

$$(M-b_s)+\cdots+(1-\rho)^{s-k}(b_{k+1}-b') > (1-\rho)^{s-j}(\theta-b_j)+...+(1-\rho)^{s-k}(b_{k+1}-b')$$

Note that terms $(M - b_s) + \cdots + (1-\rho)^{s-j+1}(b_{j+2} - b_{j+1})$ on the LHS have no analogue on the RHS, so we can cancel terms and simplify the inequality to

$$(M - b_s) + \cdots + (1-\rho)^{s-j+1}(b_{j+2} - b_{j+1}) > (1-\rho)^{s-j}(\theta - b_j)$$

since $\theta \in [b_j, b_{j+1}]$, the inequality must hold strictly.

□

A.2 Proof of Lemma 2

(i) Claim: *The social planner will never use violence in the first period.*

Proof. Recall the prior is that $\theta \sim Unif[m, M]$. I will show that there does not exist any $\theta' \in [m, M]$ such that there is a strictly higher probability of defeating θ' if violence is used in the first period and peace in the second period rather than using peace in the first period.

Suppose $\theta' \in [m, \bar{v}]$. If violence is used initially, θ is defeated with certainty. If peace is used in the first period and generates power less than θ, then the social planner uses violence and defeats θ' with certainty. Hence there is no difference in the probability of success.

Suppose $\theta \in [\bar{v}, M]$. If violence is used initially, then the likelihood of defeating θ is the probability that the peaceful protest in the second period is greater than θ', which is

$$\mathcal{P}_2 \equiv \frac{\hat{c}_2^p(\bar{v}) + \epsilon - (2\epsilon)A_p^{-1}(\theta') - \underline{c}}{\bar{c} - \underline{c}}$$
$$\text{where } \hat{c}_2^p(\bar{v}) = \int_0^1 \frac{A_p(l) - \bar{v}}{M - \bar{v}} dl$$

If peace is used initially, then the likelihood of defeating θ is the probability that the peaceful protest in the first period is greater than θ', which is

$$\mathcal{P}_1 \equiv \frac{\hat{c}_1^p + \epsilon - (2\epsilon)A_p^{-1}(\theta') - \underline{c}}{\bar{c} - \underline{c}}$$
$$\text{where } \hat{c}_1^p = \int_0^1 \frac{A_p(l) - m}{M - m} dl$$

Note that $\hat{c}_2^p(\bar{v}) < \hat{c}_1^p$, and besides that all the other terms are identical between $\mathcal{P}_1$ and $\mathcal{P}_2$. Hence, $\mathcal{P}_1 < \mathcal{P}_2$, so there is a strictly smaller probability of defeating θ if peace is used first.

Now note that there is no benefit to using violence more than once. Any θ which cannot be defeated using violence in one period cannot be defeated using violence in another period,

and any θ which can be defeated using violence in one period can also be defeated using violence in another period. By contrast, using peace in the first period affords positive probability of defeating values of θ greater than $\bar{v}$, so it is strictly preferable to use peace first and violence second than using violence twice.

Thus, it is never optimal to use violence in the first period. □

(ii) Claim: *The range of failed first-period protests for which violence is used in the second period is increasing in $\bar{v}$.*

Proof. Let the power of the first-period protest be denoted by A_1. The game only continues if the first-period protest failed. Then, the expected success of peaceful protest is given by:

$$\int_{\underline{c}}^{\bar{c}} \frac{1}{\bar{c}-\underline{c}} \frac{\max\left\{A_p\Big(\min(\max(\frac{\hat{c}_2^p - c+\epsilon}{2\epsilon}, 0), 1)\Big) - A_1, 0\right\}}{M - A_1} dc \qquad (\star)$$

where

$$\hat{c}_2^p = \int_0^1 \frac{\max(A_p(l) - A_1, 0)}{M - A_1} dl$$

Note that $\underline{c} < \bar{c}$, $0 \leq A_1 \leq A_p(1) \leq M$. I further assume that A_p is a linear function with slope p and intercept 0.

The expected success of a violent protest is given by

$$\frac{\max\{\bar{v} - A_1, 0\}}{M - A_1}$$

The social planner chooses peace for the second period if $(\star) > \frac{\max\{\bar{v}-A_1,0\}}{M-A_1}$. If $\bar{v} < A_1$ then this is trivially true. Consider the case where $\bar{v} \geq A_1$:

$$\int_{\underline{c}}^{\bar{c}} \frac{1}{\bar{c}-\underline{c}} \frac{\max\left\{A_p\Big(\min(\max(\frac{\hat{c}_2^p - c+\epsilon}{2\epsilon}, 0), 1)\Big) - A_1, 0\right\}}{M - A_1} dc > \frac{\max\{\bar{v} - A_1, 0\}}{M - A_1}$$

$$\iff \frac{1}{\bar{c}-\underline{c}} \int_{\underline{c}}^{\bar{c}} \max\left\{A_p\Big(\min(\max(\frac{\hat{c}_2^p - c+\epsilon}{2\epsilon}, 0), 1)\Big) - A_1, 0\right\} dc > \bar{v} - A_1$$

$$H(A_1) \equiv \frac{1}{\bar{c}-\underline{c}} \int_{\underline{c}}^{\bar{c}} \max\left\{A_p\Big(\min(\max(\frac{\hat{c}_2^p - c+\epsilon}{2\epsilon}, 0), 1)\Big) - A_1, 0\right\} dc + A_1 > \bar{v}$$

The task now reduces to proving that $H(A_1)$ is convex in A_1. Since the RHS ($\bar{v}$) is constant in A_1, this suffices to show that the set of A_1 for which the inequality is true is decreasing in $\bar{v}$ (equivalently, the set of A_1 for which violence is the success-maximizing choice in the second period is increasing in $\bar{v}$.

To differentiate A_1, it is helpful to adjust the bounds of integration.

$$A_p\Big(\min(\max(\frac{\hat{c}_2^p - c + \epsilon}{2\epsilon}, 0), 1)\Big) < A_1$$
$$\iff c > \hat{c}_2^p + \epsilon - 2\epsilon\Big(\frac{A_1}{p}\Big)$$

Note that $\min\{\max\{\frac{\hat{c}_2^p - c + \epsilon}{2\epsilon}, 0\}, 1\}$ implies that

$$\frac{\hat{c}_2^p - c + \epsilon}{2\epsilon} > 1 \quad \iff \quad c < \hat{c} - \epsilon$$
$$\frac{\hat{c}_2^p - c + \epsilon}{2\epsilon} < 0 \quad \iff \quad c > \hat{c} + \epsilon$$

Taken together, this yields the following constraints:

$$\hat{c} - \epsilon < c < \hat{c} + \epsilon \quad \text{and} \quad \hat{c} + \epsilon - 2\epsilon\frac{A_1}{p} < c$$

Note that the latter constraint is the stricter lower bound on c. Hence, we can simplify to:

$$\hat{c} + \epsilon - 2\epsilon\frac{A_1}{p} < c < \hat{c} + \epsilon$$

Hence, adjusting the bounds of integration yields

$$H(A_1) = \int_{\max\{\underline{c}, \hat{c}+\epsilon-2\epsilon(A_1/p)\}}^{\min\{\overline{c}, \hat{c}+\epsilon\}} \frac{1}{\overline{c} - \underline{c}}\left[A_p\Big(\frac{\hat{c}_2^p - c + \epsilon}{2\epsilon}\Big) - A_1\right]dc + \int_{\underline{c}}^{\hat{c}+\epsilon-2\epsilon(A_1/p)} \frac{1}{\overline{c} - \underline{c}}\left[A_p(1) - A_1\right]dc + A_1$$

Assuming that A_p is linear,

$$= \int_{\max\{\underline{c}, \hat{c}+\epsilon-2\epsilon(A_1/p)\}}^{\min\{\overline{c}, \hat{c}+\epsilon\}} \frac{1}{\overline{c} - \underline{c}}\left[p\Big(\frac{\hat{c}_2^p - c + \epsilon}{2\epsilon}\Big) - A_1\right]dc + \int_{\underline{c}}^{\hat{c}+\epsilon-2\epsilon(A_1/p)} \frac{1}{\overline{c} - \underline{c}}\left[p - A_1\right]dc + A_1$$

Case (1): Both of the $\hat{c}$ constraints bind. Then, the first integral is simply a *linear* function of c which takes value ranging from 0 to $A_p(1) - A_1$:

$$\int_{\hat{c}+\epsilon-2\epsilon(A_1/p)}^{\hat{c}+\epsilon}\left[p\Big(\frac{\hat{c}_2^p - c + \epsilon}{2\epsilon}\Big) - A_1\right]dc$$

Note that the integrand is a linear function of x. Therefore, this integral evaluates to:

$$\Big(\hat{c} + \epsilon - (\hat{c} + \epsilon - 2\epsilon(A_1/p))\Big)\frac{1}{2}\left[p\Big(\frac{\hat{c} - (\hat{c} + \epsilon) + \epsilon}{2\epsilon}\Big) - A_1 + p\Big(\frac{\hat{c} - (\hat{c} + \epsilon - 2\epsilon(A_1/p)) + \epsilon}{2\epsilon}\Big) - A_1\right]$$
$$= 2\epsilon\Big(\frac{A_1}{p}\Big)\frac{1}{2}\left[0 - A_1 + p(\frac{A_1}{p}) - A_1\right]$$
$$= \epsilon\Big(\frac{A_1}{p}\Big)(-A_1)$$
$$= -\frac{\epsilon A_1^2}{p}$$

The second integral is easy to evaluate (as the argument does not appear). Thus, $H(A_1)$ is

$$H(A_1) = \frac{1}{\bar{c}-\underline{c}}\left(-\frac{\epsilon A_1^2}{p}\right) + \frac{1}{\bar{c}-\underline{c}}\left((p-A_1)(\hat{c}+\epsilon-2\epsilon(A_1/p)-\underline{c})\right) + A_1$$

$$\begin{aligned}\text{Note that:} \quad \hat{c} &= \int_{A_p^{-1}(A_1)}^{1} \frac{A_p(x)-A_1}{M-A_1}dx \\ &= \int_{A_1/p}^{1} \frac{px-A_1}{M-A_1}dx \text{ (with linear assumption)} \\ &= \frac{(p-A_1)^2}{2p(A_1-M)}\end{aligned}$$

Hence, $H(A_1)$ is

$$\frac{1}{\bar{c}-\underline{c}}\left(\frac{-\epsilon(A_1)^2}{p} + (p-A_1)\left(\frac{(p-A_1)^2}{2p(A_1-M)} - 2\epsilon\frac{A_1}{p} + \epsilon - \underline{c}\right)\right) + A_1$$

Differentiate with respect to A_1:

$$\begin{aligned}\frac{dH(A_1)}{dA_1} &= \frac{1}{\bar{c}-\underline{c}}\left[\frac{-2\epsilon A_1}{p} + \frac{(p-A_1)^2(p-3M+2A_1)}{2p(A_1-M)^2} + \frac{2\epsilon(2A_1-p)}{p} - \epsilon + \underline{c}\right] + 1 \\ &= \frac{1}{\bar{c}-\underline{c}}\left[\frac{2\epsilon(A_1-p)}{p} + \frac{(p-A_1)^2(p-3M+2A_1)}{2p(A_1-M)^2} - \epsilon + \underline{c}\right] + 1\end{aligned}$$

Differentiate once again to obtain the second derivative:

$$\begin{aligned}&\frac{1}{\bar{c}-\underline{c}}\left[\frac{2\epsilon}{p} + \frac{(p-A_1)\left(3M^2+p^2+pA_1+A_1^2-3M(p+A_1)\right)}{p(M-A_1)^3}\right] \\ =&\frac{1}{\bar{c}-\underline{c}}\left[\frac{2\epsilon}{p} + \frac{(p-A_1)\left(3M(M-p-A_1)+p^2+pA_1+A_1^2\right)}{p(M-A_1)^3}\right]\end{aligned}$$

Every term is positive (since $M \geq p + A_1$). Hence, the entire expression is positive and $H(A_1)$ is convex.

Case (2): $\underline{c}$ and $\bar{c}$ constraints bind. In this case, full turnout is never obtained, so we drop the second term (which describes instances where the cost draw is such that full turnout is obtained). Then,

$$\begin{aligned}H(A_1) &= \int_{\underline{c}}^{\bar{c}} \frac{1}{\bar{c}-\underline{c}}\left(\frac{p(\hat{c}-c+\epsilon)}{2\epsilon} - A_1\right)dc + A_1 \\ &= \frac{1}{\bar{c}-\underline{c}}\frac{1}{2}\left(\frac{p(\hat{c}-\bar{c}+\epsilon)}{2\epsilon} - A_1 + \frac{p(\hat{c}-\underline{c}+\epsilon)}{2\epsilon} - A_1\right) + A_1 \\ &= \frac{1}{2(\bar{c}-\underline{c})}\left(\frac{2p(\hat{c}+\epsilon)}{2\epsilon} + \frac{p(\bar{c}-\underline{c})}{2\epsilon} - 2A_1\right) + A_1 \\ &= \frac{1}{2(\bar{c}-\underline{c})}\left(\frac{p(\hat{c}+\epsilon)}{\epsilon} + \frac{p(\bar{c}-\underline{c})}{2\epsilon}\right) - \frac{A_1}{\bar{c}-\underline{c}} + A_1\end{aligned}$$

Differentiate wrt A_1:

$$\begin{aligned}\frac{dH(A_1)}{dA_1} &= \frac{1}{2(\overline{c}-\underline{c})}\left(\frac{p}{2\epsilon}\right)\left(\frac{d\hat{c}(A_1)}{dA_1}\right)+\frac{1}{\overline{c}-\underline{c}}+1\\ &= \frac{1}{2(\overline{c}-\underline{c})}\left(\frac{p}{2\epsilon}\right)\left(\frac{1}{2p}\ \frac{p^2-2Mp-A_1^2+2MA_1}{(M-A_1)^2}\right)-\frac{1}{\overline{c}-\underline{c}}+1\end{aligned}$$

Differentiate again:

$$= \frac{1}{2(\overline{c}-\underline{c})}\left(\frac{p}{2\epsilon}\right)\left(\frac{1}{p}\ \frac{(M-p)^2}{(M-A_1)^2}\right) \qquad > 0$$

Case (3): $\underline{c}, \hat{c}+\epsilon$ constraints bind. In this case, full turnout is again never obtained, so we drop the second term. Then,

$$\begin{aligned}H(A_1) &= \int_{\underline{c}}^{\hat{c}+\epsilon}\frac{1}{\overline{c}-\underline{c}}\left(\frac{p(\hat{c}-c+\epsilon)}{2\epsilon}-A_1\right)dc+A_1\\ &= \frac{1}{\overline{c}-\underline{c}}\frac{1}{2}\left(\frac{p(\hat{c}-(\hat{c}+\epsilon)+\epsilon)}{2\epsilon}-A_1+\frac{p(\hat{c}-\underline{c}+\epsilon)}{2\epsilon}-A_1\right)+A_1\\ &= \frac{1}{2(\overline{c}-\underline{c})}\left(\frac{p(\hat{c}-\underline{c}+\epsilon)}{2\epsilon}-2A_1\right)+A_1\\ &= \frac{1}{2(\overline{c}-\underline{c})}\left(\frac{p(\hat{c}-\underline{c}+\epsilon)}{2\epsilon}\right)-\frac{A_1}{\overline{c}-\underline{c}}+A_1\end{aligned}$$

Differentiate wrt A_1:

$$\frac{dH(A_1)}{dA_1} = \frac{1}{2(\overline{c}-\underline{c})}\left(\frac{p}{2\epsilon}\right)\left(\frac{1}{2p}\ \frac{p^2-2Mp-A_1^2+2MA_1}{(M-A_1)^2}\right)-\frac{1}{\overline{c}-\underline{c}}+1$$

Differentiate again:

$$= \frac{1}{2(\overline{c}-\underline{c})}\left(\frac{p}{2\epsilon}\right)\left(\frac{1}{p}\ \frac{(M-p)^2}{(M-A_1)^2}\right) \qquad > 0$$

Case (4): $\hat{c}-\epsilon+2\epsilon(A_1/p), \overline{c}$ constraints bind. Then,

$$\begin{aligned}H(A_1) &= \int_{\hat{c}+\epsilon-(2\epsilon)(A_1/p)}^{\overline{c}}\frac{1}{\overline{c}-\underline{c}}\left(\frac{p(\hat{c}-c+\epsilon}{2\epsilon}-A_1\right)dc+\int_{\underline{c}}^{\hat{c}+\epsilon-2\epsilon(A_1/p)}\frac{p-A_1}{\overline{c}-\underline{c}}dc+A_1\\ &= \frac{1}{2(\overline{c}-\underline{c})}\left(\frac{p(\hat{c}-\overline{c}-\epsilon+\epsilon(A_1/p)}{2\epsilon}-2A_1\right)+\frac{1}{\overline{c}-\underline{c}}\left((p-A_1)(\hat{c}+\epsilon-2\epsilon(A_1/p)-\underline{c})\right)+A_1\end{aligned}$$

Differentiating wrt A_1:

$$\begin{aligned}\frac{dH(A_1)}{dA_1} &= \frac{1}{2(\overline{c}-\underline{c})}\left[\frac{p}{2\epsilon}\left(\frac{d\hat{c}}{dA_1}+\frac{2\epsilon}{p}\right)-2\right]\\ &\quad+\frac{1}{\overline{c}-\underline{c}}\left[+\frac{(p-A_1)^2(p-3M+2A_1)}{2p(A_1-M)^2}+\frac{2\epsilon(2A_1-p)}{p}-\epsilon+\underline{c}\right]+1\\ &= \frac{1}{2(\overline{c}-\underline{c})}\left[\frac{p}{2\epsilon}\left(\frac{d\hat{c}}{dA_1}\right)-1\right]+\frac{1}{\overline{c}-\underline{c}}\left[+\frac{(p-A_1)^2(p-3M+2A_1)}{2p(A_1-M)^2}+\frac{2\epsilon(2A_1-p)}{p}-\epsilon+\underline{c}\right]+1\end{aligned}$$

Differentiate once again to obtain the second derivative:

$$= \frac{1}{\overline{c} - \underline{c}} \left[\frac{p}{2\epsilon} \frac{d^2 \hat{c}}{dA_1^2} + \frac{(p - A_1)\big(3M(M - p - A_1) + p^2 + pA_1 + A_1^2\big)}{p(M - A_1)^3} \right]$$

By previous arguments we know all these terms are positive.

Hence in all four cases, $H(A_1)$ is convex.

□