Stronger Challengers can Cause More (or Less) Conflict and Institutional Reform

Andrew Little*		Jack Paine†

June 3, 2022

Abstract

Prominent theories propose that phenomena such as war and democratization occur when rulers cannot commit to future promises. However, existing work disagrees about a basic question: how does the coercive strength of a challenger affect prospects for conflict and/or institutional reform? We analyze a formal model with a general distribution of the probability that the challenger would win a conflict in a given period ("threat"). Whether a stronger challenger means more conflict or reform depends on exactly how this affects the distribution of threats; in particular the average and maximum threat. If the maximum threat is fixed and stronger challengers pose a higher average threat, then weaker challengers are prone to rebel during rare periods when they pose a high threat. However, if stronger challengers pose a higher maximum threat, then they are harder to buy off. We apply these results to advance theoretical and debates about democratization.

*Associate Professor, Department of Political Science, University of California Berkeley.
†Associate Professor, Department of Political Science, University of Rochester.

1 Introduction

Why do countries vary in their incidence of civil or international conflict? Why do some countries democratize? Under what conditions do dictators share power? Much existing research points to *dynamic commitment problems* to explain these varied phenomena.[1] The core premise is that an actor who controls a flow of rents, such as a government, cannot commit to promises about how they will distribute spoils in the future. Limited commitments make any challenger—for example, a domestic opposition group or foreign adversary—anxious about its future interactions with the government.

Limited commitment ability matters because, in most foreseeable real-world scenarios, the challenger's threat fluctuates over time. Sometimes the domestic masses have favorable opportunities to mobilize anti-government demonstrations, and sometimes they do not. Sometimes foreign states enjoy economic booms that bolster their military strength, and sometimes they do not. If the challenger poses a threat today but not tomorrow, they will lose their ability to compel the government for concessions. Consequently, limited commitment can cause a temporarily strong challenger to fight the government. Alternatively, a credible threat may compel the government to reform institutions, which would bolster their commitment to share spoils with the challenger in the future.

This style of argument is pervasive because the core intuition is straightforward, compelling, and broadly applicable. However, a basic, substantively important question remains ambiguous: do coercively stronger challengers make conflict and/or institutional reform more or less likely? How does a bigger non-elite class, a better-organized civil society, or a more advanced neighboring state affect these outcomes? These are widely debated questions in broader literatures within

[1] For democratization, see Acemoglu and Robinson (2006); Ansell and Samuels (2014); Leventoğlu (2014); Dower et al. (2018). For authoritarian power sharing and democratic separation of powers, see Helmke (2017); Christensen and Gibilisco (2020); Meng (2019); Powell (2020); Paine (2022). For civil conflict, see Fearon (2004); Chassang and Padro-i Miquel (2009); Walter (2009); Powell (2012); Gibilisco (2021). For international war, see Fearon (1995); Powell (2006); Debs and Monteiro (2014); Krainin (2017). For the general mechanism, see Powell (2004).

comparative politics and international relations, yet we lack an appropriate formal model to study them.

The core takeaway from this paper is that different ways of conceptualizing the challenger's coercive strength yield divergent conclusions. Theoretically, we examine a more general distribution of threats—measured by the probability that the challenger would win a conflict in the current period—than existing models. This allows us to understand the varied consequences of stronger challengers, which we show boils down to how their strength affects the maximum and average threat. To our knowledge, we are the first to consider any discrete or continuous distribution of threat levels, or, more generally, whatever random variable fluctuates over time. We show that seemingly technical model assumptions about this distribution in fact matter greatly for substantively important questions. We also highlight the consequences of these assumptions for empirical testing.

Most existing work makes a simplifying assumption about the distribution of threats: the challenger fluctuates between two threat levels, minimal and maximal. Minimal threats usually represent periods in which the challenger would lose for sure, or in which conflict would create such high costs that it would never be the optimal choice. By contrast, maximum threats arise when the challenger would win with certainty, which creates a dire threat. In this setup with binary threat levels and fixed values for each, a natural way to capture the coercive strength of the challenger is the probability in any period that the challenger will pose their maximum threat, with stronger challengers posing the maximal threat more frequently. Intuitively, one might expect that stronger challengers are harder to buy off. However, instead, conflict occurs along the equilibrium path only if the challenger rarely faces opportunities to coerce the regime. This corresponds with an coercively weak challenger. Their rare moments in the sun are too tempting to pass up and forgo revolting, given their poor prospects to gain concessions in the future if the status quo regime remains intact. For the same reason, the ruler faces greater incentives to extend the franchise or share power with weak challengers.

We indeed recover this scenario as a special case in our model. More generally, in this case, we can think of the underlying commitment problem as less pressing because the challenger is stronger on *average*. This source of strength bolsters their opportunities to compel concessions from the ruler across periods, which lessens motives to fight.

However, as we demonstrate, this is not a general result about the consequences of stronger challengers. The problem with the seemingly innocuous assumption to construct the distribution of threats in this manner is that an important margin is held fixed—the maximum threat. If stronger challengers pose a higher maximum threat, they will be harder to buy off peacefully when this threat is realized and can force more institutional change.

Generally, coercively strong challengers should pose greater threats both on average and when maximally strong. This means that the overall effect of coercive strength is theoretically ambiguous. Stronger challengers are more likely to fight or gain institutional concessions whenever a shift in the strength parameter raises the maximum threat by at least as much as the average, and can occur even if the average threat increases at a somewhat higher rate. For example, a uniform rightward shift in the distribution of threats makes the challenger harder to buy off peacefully. Such a shift also makes institutional reform more likely. Even in a simple binary threats model, we can recover the core intuitions as long as we allow both the maximum and average threats to vary. Overall, we cannot understand the consequences of challenger strength without taking into account how coercive strength affects both the maximum and average threats.

Beyond "challenger strength" specifically, our theoretical results provides a new lens to study the effects of many possible stimuli. For example, exercising repression may either increase or decrease prospects for conflict, depending on how it changes the distribution of the challenger's probability of winning. If repression creates a uniform downward shift in these probabilities, then the probability of conflict and the need to offer institutional reform will decrease. By contrast, if repression usually prevents people from mobilizing but creates rare instances where they are able to forge cross-class coalitions, such regimes might be subject to revolutionary outbursts because the maximum threat is high whereas the average threat is low—hence leaving challengers "no other way out" than revolution (Goodwin, 2001).

We conclude by demonstrating how our model yields new insights for debates about democratization and authoritarian power sharing. In models such as Acemoglu and Robinson (2006) and Dower et al. (2018), coercively weak challengers trigger institutional reform. A low average threat makes the shadow of the future unfavorable. This, combined with a high maximum threat, bolsters the challenger's bargaining leverage in a rare maximum-threat period. However, other seemingly similar models yield the opposite implication about challenger strength (Ansell and Samuels, 2014; Meng, 2019; Paine, 2022). By disaggregating maximum and average threats, our model explains the conditions under which we recover each implication. These findings also offer guidance for empirical research designs that test these models. Recent studies propose innovative ways to measure key parameters, but do not consider the countervailing effects of higher maximum and average threats. Future work must push on this frontier to extend our understanding of how the coercive strength of societal challengers affects prospects for conflict and institutional reform.

2 Formal Model

2.1 Setup

A ruler and challenger bargain over spoils in periods $t = 1, 2, ...$ with a common discount factor $\delta \in (0, 1)$. We normalize total spoils in each period to 1. In each period, the ruler makes a take-it-or-leave it offer $x_t \leq 1$. That is, we impose the common assumption in this literature that the ruler cannot transfer more than the entire contemporaneous budget in any period, and hence cannot borrow across periods. It is less standard to not impose a lower bound on the ruler's offer. Briefly, the case without a lower bound leads quickly to clear results, though the analysis without this assumption is qualitatively similar. We elaborate upon this assumption following the model analysis, and we extend the model to include a lower bound in Appendix A.2.

If the challenger accepts an offer in some period t, then the ruler and challenger respectively consume $(1-x_t, x_t)$ and engage in a strategically identical interaction in period $t+1$. If instead the challenger rejects in period t, then conflict occurs. Fighting is a game-ending move that eliminates all consumption in the period of the fight and permanently destroys a fraction $\phi \in (0,1)$ of total spoils in each future period, with the winner consuming all the remaining spoils.

The challenger's probability of winning a conflict varies by period. The parameter is p_t, which depends on an independently and identically distributed choice by Nature revealed to both players at the outset of each period. Thus, at the bargaining stage, both actors are perfectly informed about p_t. We call p_t the *threat* posed by the challenger in period t. The distribution function of p_t is $F(p;s)$, where s is a parameter that captures the challenger's latent coercive capabilities, or strength. The distribution has mean $\bar{p}(s) \equiv \mathbb{E}[p;s]$ and support on $\left[p^{\min}(s), p^{\max}(s)\right]$, for $0 \leq p^{\min} < p^{\max} \leq 1$. To capture the general notion that stronger challengers tend to pose a higher threat, we assume that $\bar{p}(s)$, $p^{\min}(s)$, and $p^{\max}(s)$ each weakly increase in s. To streamline the exposition, we suppress s when doing so does not cause confusion.

Later we summarize various model extensions, none of which alter the core insights: allowing for a path-dependent distribution of threats, modeling fluctuations in the cost of conflict rather than the probability of winning, and allowing the ruler to engage in institutional reform.

2.2 The Distribution of Threats and Conflict

We examine the conditions under which a Markov Perfect Equilibrium (MPE) exists in which conflict occurs with probability 0 along the equilibrium path. We refer to this as a peaceful equilibrium. In Appendix A.1, we discuss properties of conflictual paths of play.

Along a peaceful equilibrium path, in every period t, the ruler makes an offer $x_t \leq 1$ that the challenger accepts. In any equilibrium, the challenger accepts only offers for which its lifetime expected stream of consumption along a peaceful path weakly exceeds the value of its fighting outside option. Thus, if we write the challenger's future continuation value along a peaceful path as V^C, a necessary condition for peaceful bargaining in any period t is:

$$\underbrace{x_t + \delta V^C}_{\text{Accept}} \geq \underbrace{\frac{\delta}{1-\delta} p_t (1-\phi)}_{\text{Fight}} . \tag{1}$$

Given our present assumption that x_t is not bounded from below, the ruler never makes offers that the challenger strictly prefers to accept. Otherwise, the ruler could profitably deviate by making a slightly lower offer that the challenger would accept. Consequently, Equation (1) must hold with equality for every period t, and thus the optimal transfer in every period satisfies:

$$x^*(p_t) = \delta\Big(p_t \frac{1-\phi}{1-\delta} - V^C\Big). \tag{2}$$

The next step is to solve for the continuation value V^C. In a peaceful MPE in which the ruler uses this offer function in every period, we can write the continuation value as equal to the average transfer divided by $1-\delta$. An analytically convenient aspect of the optimal offer is that it is linear in the current-period threat p_t, and hence the average value of p_t is the only aspect of the distribution that affects the continuation value. As demonstrated in Appendix A.1, this property holds in any equilibrium with conflict as well.

Formally, we can write the continuation value as:

$$V^C = \frac{1}{1-\delta} \underbrace{\int_{p^{\text{min}}}^{p^{\text{max}}} \delta\Big(p_t \frac{1-\phi}{1-\delta} - V^C\Big) dF(p)}_{\text{Average per-period transfer}}$$

$$\implies V^C = \frac{\delta}{1-\delta} \bar{p}(1-\phi). \tag{3}$$

Combining Equations (2) and (3) enables us to explicitly solve for the equilibrium per-period offer:

$$x^*(p_t) = \frac{\delta}{1-\delta}\big(p_t - \delta\bar{p}\big)(1-\phi). \tag{4}$$

A peaceful equilibrium requires that the challenger can be bought off in every period. Equation (4) makes clear that this condition is most difficult to satisfy when the challenger poses their maximum threat, which we formalize in Proposition 1.

Proposition 1 (Existence of a peaceful equilibrium)**.** *The following inequality is a necessary and sufficient condition for a peaceful equilibrium to exist:*

$$\frac{\delta}{1-\delta}\big(p^{max} - \delta\bar{p}\big)(1-\phi) \leq 1.$$

The condition in Proposition 1 enables us to take comparative statics on the challenger's strength, s. If we move the threat parameters to one side of the inequality and write them explicitly as a function of s, we have:

$$\frac{1-\delta}{\delta(1-\phi)} \geq p^{\max}(s) - \delta\bar{p}(s) \equiv \tau(s). \tag{5}$$

The overall effect of increasing the challenger's strength on the prospects for peace can be summarized by how s affects the $\tau(s)$ term. Increases in $\tau(s)$ raise prospects for conflict, whereas decreases in $\tau(s)$ have the opposite effect. The relevant derivative is:

$$\frac{\partial\tau(s)}{\partial s} \equiv \underbrace{\frac{\partial p^{\max}(s)}{\partial s}}_{\uparrow\text{ max threat}} - \delta \cdot \underbrace{\frac{\partial\bar{p}(s)}{\partial s}}_{\uparrow\text{ average threat}}. \tag{6}$$

This expression expounds our main point about the need to compare the maximum and average probabilities of winning. These parameters exert countervailing effects on prospects for conflict. On the one hand, higher s raises prospects for conflict through its effect on raising $p^{\max}$. This effect raises the challenger's opportunity cost to not fighting in a maximum-threat period. When we raise the maximum threat while leaving constant other elements of the distribution, we increase the discrepancy between the challenger's threat in the current period and their threat in future periods. This creates the temptation to fight now to "lock in" their temporary advantage. Consequently, the inequality from Proposition 1 holds for a smaller range of parameter values.

On the other hand, higher s diminishes prospects for conflict through its effect on raising $\bar{p}$. When the challenger contemplates fighting in a maximum-threat period, it considers the magnitude of the adverse shift in the future distribution of power. High $\bar{p}$ lowers the opportunity cost of not fighting. The challenger expects favorable draws of p_t in the future along a peaceful path, which diminishes their incentives to fight now. Consequently, the inequality from Proposition 1 is easier to meet.

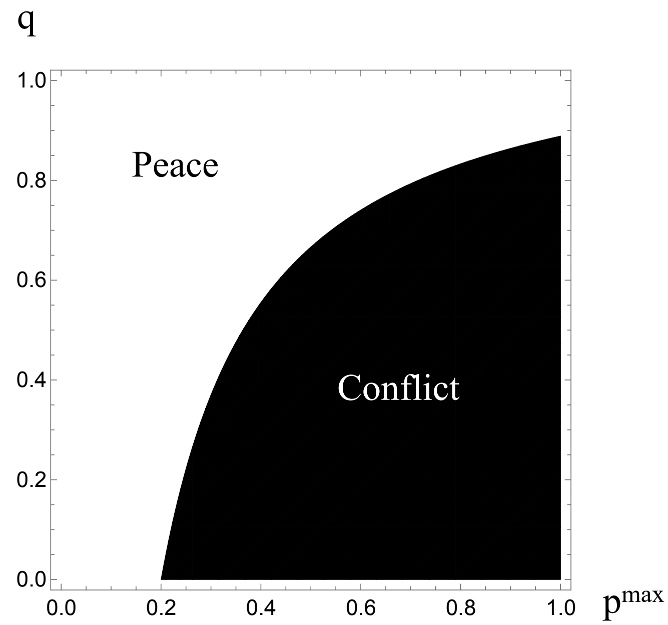
General binary distribution To connect this result more directly to past work, suppose the per-period threat takes one of two values, which we write as $p_t \in \{p^{\min}, p^{\max}\}$, $0 \leq p^{\min} < p^{\max} \leq 1$, with $q = Pr(p_t = p^{\max})$. In this case, the average threat is $(1-q)p^{\min} + qp^{\max}$. Substituting this term into Equation (6) and taking comparative statics yields:

$$\frac{\partial \tau(s)}{\partial s} = (1-\delta q) \cdot \underbrace{\frac{\partial p^{\max}}{\partial s}}_{\uparrow \text{ max threat}} - \delta(1-q) \cdot \underbrace{\frac{\partial p^{\min}}{\partial s}}_{\uparrow \text{ min threat}} - \delta(p^{\max} - p^{\min}) \cdot \underbrace{\frac{\partial q}{\partial s}}_{\uparrow \text{ max-threat periods}} . \tag{7}$$

In Figure 1, we graphically summarize some key comparative statics predictions. It is a region plot with $p^{\max}$ on the x-axis and q on the y-axis; all other parameters are fixed at values stated in the accompanying note. The white region corresponds with parameter values in which the equilibrium path of play is peaceful (that is, the inequality in Proposition 1 holds), whereas conflict occurs in the dark region.

Equation (7) and Figure 1 clarify the intuition for the result from Acemoglu and Robinson (2006) and other models in which an increase in challenger strength makes it *easier to buy them*

Figure 1: Peace and Conflict in the Binary Threats Model



Parameter values: $\delta = 0.9$, $\phi = 0.5$, $p^{\text{min}} = 0$.

off; or, equivalently in Fearon (2004), that an decrease in the government's strength makes civil war less likely to occur. In a distribution in which the values of the minimum and maximum threats are fixed, the first two terms in Equation (7) are 0. Hence, higher s improves the shadow of the future for the challenger along a peaceful path (q) without altering the opportunity cost of fighting in the maximum-threat state (p^{max}). This corresponds with an upward shift in Figure 1, which can move parameter values from conflict to peace.

Our analysis also suggests a sense in which we can generalize this finding. For any distribution shift such that the upper bound is fixed but the average increases, it will be easier to buy off the challenger peacefully. With a binary distribution, this implies fixing p^{max} and raising either p^{min} or q.

However, even with a binary distribution of threats, raising the challenger's strength can instead produce the opposite effect. The simplest case is one in which greater coercive strength raises p^{max} while the other parameter values are fixed. This corresponds with a rightward shift in Figure 1, which can move parameter values from peace to conflict.

Another, perhaps more substantively natural case, is when increasing s shifts F uniformly to the right. This corresponds to increasing the probability of the challenger winning by a fixed amount in each period. Hence, the minimum and maximum threat increase at the same rate, $\frac{\partial p^{\max}}{\partial s} = \frac{\partial p^{\min}}{\partial s} = d > 0$; but the per-period probability of each threat realization does not, $\frac{\partial q}{\partial s} = 0$. In this case, facing a stronger challenger makes *peace harder to sustain*. This can be seen by substituting this case into Equation (7), which yields $\frac{\partial \tau(s)}{\partial s} = (1-\delta)d > 0$.

These examples highlight a useful fact for future theorizing: a binary distribution in of itself does not discernibly limit the generality of insights from models with dynamic commitment problems. Even with a simple distribution, increasing the challenger's strength can either increase or decrease prospects for conflict. Instead, the important takeaway is that how the researcher conceptualizes strength and structures the parameters in the distribution determines the direction of the comparative statics prediction. A binary distribution of threats contains three key parameters, and different changes carry divergent implications for the prospect of peace.

2.3 Discussion of Assumptions and Extensions

In the baseline model, we do not impose a lower bound on the offers. This makes it possible for the ruler to offer $x_t < 0$ and hence to demand a net transfer from the challenger. The case without a lower bound is analytically simpler because the ruler can hold down the challenger to their reservation value in every period. This implies the offer is linear in the challenger's strength, and hence the average threat is the only part of the distribution that matters for the continuation value. With a lower bound, the continuation value depends on other aspects of the distribution. However, the core insights are sometimes identical and otherwise qualitatively similar when we assume that offers must be weakly positive or above some other bound $\underline{x}$, which we demonstrate in Appendix A.2. To preview the intuition, suppose $\underline{x} = 0$. From Equation (4), it is immediately apparent that all interior-optimal offers strictly exceed zero if $p^{\min} \geq \delta\bar{p}$. Thus, the lower bound never binds if there is a small range of feasible values of p and the actors are not too patient. If instead $p^{\min} < \delta\bar{p}$, then the zero-lower bound binds. This case adds additional terms, but does not qualitatively alter the main insight that we need to compare the maximum and average threats.

Another simplifying assumption is that threats are drawn independently and identically across periods. In Appendix A.3, we relax this assumption and demonstrate that our key findings hold when we allow for a specific type of path dependence: there is a probability that Nature does not change the state of the world in the next period (and, with complementary probability, Nature draws from the same underlying distribution of threats as in the baseline model). Although our model does not nest all forms of path dependence or deterministic shifts (Krainin, 2017; Gibilisco, 2021), this extension demonstrates that our core findings are not knife-edge implications of assuming iid shocks.

Finally, we have assumed that p_t varies over time but all other parameters remain constant. This is the most natural way to capture our core substantive interest in understanding the effects of coercively stronger challengers. In Appendix A.4, we show that the intuition is identical when we instead allow the permanent costs of fighting, ϕ_t, to vary across periods. This alteration more closely resembles how Acemoglu and Robinson (2006) conceptualize shifts in power over time.

2.4 Prospects for Institutional Reform

Our final extension is more substantively oriented and addresses endogenous institutional reform. We have shown that the challenger's strength parameter, s, exhibits ambiguous consequences for conflict. The intuition is identical when we allow the ruler to strategically reform institutions. In Appendix A.5, we assume that the ruler in each period can choose to permanently increase the basement level of spoils the challenger consumes in all periods (that is, choose the value of $\underline{x}$, introduced in Appendix A.2). We interpret a higher basement level of spoils as capturing a power-sharing agreement, democratization, or any other institutional reform that constrains the ruler's ability to dictate the division of spoils.

The parameter region in which institutional reform occurs in this extension is identical to that in which conflict occurs in the baseline game. Along the equilibrium path, in the first maximum-threat period, the ruler offers a sufficient level of institutional reforms to enable buying off the challenger then and in all future periods. The continuous choice of institutional reform enables the ruler to hold the challenger down to indifference, and the ruler would immediately incur the costs of conflict if she did not reform institutions. Consequently, the ruler never lets conflict occur along the equilibrium path.

The equivalence of the institutional reform region with the conflict region implies that all comparative statics from the baseline model carry over to explain institutional reform: a greater average threat diminishes incentives for institutional reform, and a greater maximum threat increases incentives for institutional reform. Higher $p^{\max}$ also increases the extent of institutional reform (that is, raises the optimal choice of $\underline{x}$), conditional on any occurring. A challenger with high $p^{\max}$ requires greater assurances to compensate for the higher opportunity cost of not fighting in a maximum-threat period.

3 Application to Democratization and Power Sharing

To illustrate the substantive importance of our findings, we engage with debates about causes of democratization and authoritarian power sharing. We adjudicate divergent theoretical implications and provide implications for empirical research designs.

Adjudicating divergent theoretical implications In Acemoglu and Robinson's baseline model of authoritarian politics,[2] economic elites (the equivalent to our generic reference to a "ruler") control the political regime. Elites interact with the masses (equivalently, "challenger"), whose threat alternates over time according to a binary distribution with $p^{\min} = 0$ and $p^{\max} = 1$.[3] Thus, in

[2]See Acemoglu and Robinson (2006), Chapter 5.

[3]Again, this is a slightly different interpretation of their parameters, but is conceptually equivalent (see Appendix A.4).

maximum-threat periods, the masses can threaten to stage a revolution, which succeeds with probability 1 and removes elites from power forever. In every maximum-threat period, elites would like to buy off the masses by setting a high tax rate and redistributing wealth. However, elites cannot credibly commit to make concessions in any future period in which the masses pose the minimum threat, in the sense that a revolutionary attempt succeeds with probability 0. If maximum-threat periods arise rarely, then in any such period, the masses stage a revolution to establish a new regime—given their unfavorable shadow of the future engendered by a high frequency of minimum-threat periods. Consequently, costly fighting occurs in equilibrium because of the confluence of two factors: the distribution of threats fluctuates over time, and elites cannot commit to compensate the masses in weak periods.

Acemoglu and Robinson then extend their framework to explain institutional reform.[4] If revolution would otherwise occur along the equilibrium path, then elites will extend the franchise. The drawback for elites is that democratization enables the masses to set the tax rate in all future periods. However, elites benefit by preventing the catastrophic destruction that a revolution would unleash. In our model, increasing the lower-bound offer $\underline{x}$ corresponds with franchise expansion.

In the Acemoglu and Robinson model, a stronger challenger is synonymous with more frequent maximum-threat periods. Thus, strength affects the average but not the maximum threat, which is fixed at $p^{\max} = 1$. As we highlighted in our analysis of the general binary distribution, this implies that weaker challengers have a more credible threat to revolt. This, in turn, compels the ruler to offer institutional concessions.

Ansell and Samuels (2014) confront a core assumption underlying these results (see especially pp. 70-71). They contend that the material resources of a group should influence their probability of winning. Industrialization should create a stronger capitalist class that is better-positioned to challenge landed elites who monopolize power. Rather than fix $p^{\max} = 1$, they parameterize the

[4]See Acemoglu and Robinson (2006), Chapter 6. They also introduce a strategic option for elites to repress the masses, which lies outside the scope of our discussion here and hence we ignore it.

challenger's probability of winning in a similar fashion to our term $p^{\max}(s)$. They conclude that stronger challengers have better bargaining leverage, which enables them to compel institutional reform—thus producing the opposite result as in Acemoglu and Robinson. However, Ansell and Samuels' model is a one-shot game, which means that threats do not fluctuate over time. As we demonstrate with our more general model, this is a special case in which challenger strength affects the maximum threat and its effect on the average threat is perfectly autocorrelated.

A parallel, although previously unrecognized, debate exists about motives for authoritarian power sharing. Dower et al. (2018) extend the Acemoglu and Robinson framework to incorporate the possibility of partial institutional reform within an authoritarian regime, as opposed to the all-or-nothing choice of full democratization. Once again, challenger strength affects the average but not the maximum threat, and thus weaker challengers compel power sharing.

By contrast, in Meng's (2019) two-period game, the challenger grows weaker over time as the dictator consolidates power between periods 1 and 2. Consequently, challengers that begin strong (or, in her phrasing, dictators who begin their tenure in a weak position) anticipate a larger adverse shift in the future distribution of power. This makes stronger challengers more prone to fight (in her substantive setting, stage a coup) if the ruler does not share power with them at the outset, which induces power sharing. Here, greater strength affects the maximum threat more than the average threat.

In Paine (2022), the relationship between challenger strength and prospects for fighting (and power-sharing deals) are inverted U-shaped. Very weak challengers have a low chance of ever prevailing (low maximum threat), and very strong challengers frequently enjoy maximum-threat periods (high average threat). Only intermediate-strong challengers have a credible threat to fight, which induces the ruler to share power. In this range, the maximum threat is large relative to the average threat.

In sum, we can recover implications from several seemingly inconsistent arguments about democratization and authoritarian power sharing as special cases of our more general model. Existing models yield divergent comparative statics for challenger strength because of varying assumptions that affect the relationship between the maximum and average threat. Understanding that these are the key theoretical quantities in these models should help to advance future theoretical work. Seemingly technical model assumptions carry important substantive implications.

Implications for empirical research designs Our analysis also helps to clarify impediments to empirically testing the relationship between challenger strength and either conflict or institutional reform. Recent research proposes innovative measures to assess this relationship. However, these studies do not engage with our core point: the hypothesized direction of the effect of challenger strength is ambiguous, and depends on the relationship between the maximum and the average threat. We encourage researchers to address this key point in future studies, although we caution that the theoretical maximum threat may be a parameter that is fundamentally impossible to pin down precisely.

Exemplifying the importance of our main point, leading empirical evaluations of democratization models with commitment problems assess opposing hypotheses.[5] Dower et al. (2018) study endogenous representation for peasants in Imperial Russia. Reforms created district-level assemblies, *zemstvo*, which varied in the extent of representation for peasants. The authors use the frequency of past protests in a district to proxy for the ability to protest in the future, that is, the q parameter. They find that high levels of past unrest engendered less representation for peasants, which is consistent with the model under scope conditions in which coercive strength raises the average more than the maximum threat. However, these scope conditions (and the ensuing theoretical implication) require a key additional assumption: historical threat levels minimally impacted the magnitude of the threat posed at the time institutional reforms were offered. That is, we must

[5] Many other studies empirically assess predictions from Acemoglu and Robinson (2006) about the relationship between economic inequality and democratization. Because these theoretical implications follow directly from underlying assumptions about the effects of challenger strength, the considerations raised here apply to these empirical tests as well.

assume the maximum threat $p^{\max}$ was constant.[6]

By contrast, Aidt and Franck (2015) focus on the present threat posed by the masses. Specifically, they leverage incidence in the so-called Swing Riots to measure how British MPs perceived the threat level in their districts and how this perception affected their votes on the bill that became known as the Great Reform Act of 1832. Drawing explicitly from Acemoglu and Robinson's theory, they interpret widespread protests and rioting as a credible signal to autocratic elites that the generic hurdles to mobilizing and coordinating popular support have been temporarily overcome, that is, the masses pose their maximum threat and this threat is ominous (i.e., high $p^{\max}$). Hence, they anticipate that MPs are *more* likely to vote for reform when more riots and protests occur in their district. However, comparing this hypothesis to the opposing one tested in Dower et al. (2018) again highlights the additional steps need to link the theory to empirics. Aidt and Franck assume that strong challengers pose purely transitory threats and hence their average threat is low. However, suppose instead that riots and protests proxy for regions in which the masses posed *persistently* strong threats, even if not activated at all points in time. Then the model would anticipate that MPs in high-protest districts should be able to pacify the recalcitrant masses with temporary transfers rather than permanent reforms. In this case, we should expect them to vote against the Reform Act.

Similarly, Ansell and Samuels (2014) anticipate that higher levels of industrialization and a stronger capitalist class improve prospects for democratization. The problem, though, is that if the capitalist class is persistently strong, then the dynamic model implies that institutional reform is unnecessary. A high average threat enables capitalists to constantly pressure landowning elites for temporary concessions. Of course, in the real world, bargaining through such non-institutional channels may be prohibitively difficult to sustain over time because of transaction costs or costs of mobilizing. However, these are precisely the elements of these models that need to be developed

[6]Since the reforms were enacted across a range of districts at the same time, it may be more precise to say that the current period threat within districts was less impacted.

in future research, and measured empirically.

Our analysis highlights some fundamental impediments to empirically measuring key parameters from models of dynamic commitment problems. However, we conclude by suggesting some theoretical and empirical paths forward. On the theoretical end, we show that modeling a general distribution of challenger threats can be quite tractable, while also highlighting when restricting to a binary distribution entails minimal loss in generality. Future work can build on these insights to answer questions about how factors like repression, technology for mobilization, and economic factors affect prospects for conflict and institutional reform. Future empirical work should seek to tease apart average versus maximum threats, or perhaps more realistically, how the volatility of threats relates to conflict and reform.

References

Acemoglu, Daron and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

Aidt, Toke S and Raphaël Franck. 2015. “Democratization under the Threat of Revolution: Evidence from the Great Reform Act of 1832.” *Econometrica* 83(2):505–547.

Ansell, Ben W. and David J. Samuels. 2014. *Inequality and Democratization: An Elite Competition Approach*. Cambridge University Press.

Chassang, Sylvain and Gerard Padro-i Miquel. 2009. “Economic Shocks and Civil War.” *Quarterly Journal of Political Science* 4(3):211–228.

Christensen, Darin and Michael Gibilisco. 2020. “How Budgets Shape Power Sharing in Autocracies.” Working paper.

Debs, Alexandre and Nuno P. Monteiro. 2014. “Known Unknowns: Power Shifts, Uncertainty, and War.” *International Organization* 68(1):1–31.

Dower, Paul Castañeda, Evgeny Finkel, Scott Gehlbach and Steven Nafziger. 2018. "Collective Action and Representation in Autocracies: Evidence from Russia's Great Reforms." *American Political Science Review* 112(1):125–147.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3):379–414.

Fearon, James D. 2004. "Why Do Some Civil Wars Last So Much Longer Than Others?" *Journal of Peace Research* 41(3):275–301.

Gibilisco, Michael. 2021. "Decentralization, Repression, and Gambling for Unity." *Journal of Politics* 83(4):1353–1368.

Goodwin, Jeff. 2001. *No Other Way Out: States and Revolutionary Movements, 1945–1991*. Cambridge University Press.

Helmke, Gretchen. 2017. *Institutions on the Edge: The Origins and Consequences of Inter-branch Crises in Latin America*. Cambridge University Press.

Krainin, Colin. 2017. "Preventive War as a Result of Long Term Shifts in Power." *Political Science Research and Methods* 5(1):103–121.

Leventoğlu, Bahar. 2014. "Social Mobility, Middle Class, and Political Transitions." *Journal of Conflict Resolution* 58(5):825–864.

Meng, Anne. 2019. "Accessing the State: Executive Constraints and Credible Commitment in Dictatorships." *Journal of Theoretical Politics* 33(4):568–599.

Paine, Jack. 2022. "Strategic Power Sharing: Commitment, Capability, and Authoritarian Survival." *Journal of Politics* 84(2):1226–1232.

Powell, Robert. 2004. “The Inefficient Use of Power: Costly Conflict with Complete Information.” *American Political Science Review* 98(2):231–241.

Powell, Robert. 2006. “War as a Commitment Problem.” *International Organization* 60(1):169–203.

Powell, Robert. 2012. “Persistent Fighting and Shifting Power.” *American Journal of Political Science* 56(3):620–637.

Powell, Robert. 2020. “Power Sharing with Weak Institutions.”. Working paper.

Walter, Barbara F. 2009. “Bargaining Failures and Civil War.” *Annual Review of Political Science* 12:243–261.

Appendix

We present various additional results and extensions in the appendix. In Appendix A.1, we characterize conflictual paths of play. In Appendix A.2, we impose a lower bound on the per-period offer. In Appendix A.3, we model path-dependent shocks. In Appendix A.4, we model shocks in the cost of fighting (rather than the probability of winning). In Appendix A.5, we model endogenous institutional reform.

A.1 Properties of Conflictual Paths of Play

In this section, we characterize the properties of an equilibrium in which conflict occurs along the path of play.

What happens along the path of play When the condition in Proposition 1 fails, then the challenger will reject any offer in periods where p_t is sufficiently high; but bargaining will be peaceful in periods with a lower threat. The threshold is unique, has bounds $p^* \in (\bar{p}, p^{\max})$, and satisfies:

$$\frac{\delta}{1-\delta}\big(p^* - \delta\bar{p}\big)(1-\phi) = 1 \implies p^* = \frac{1-\delta}{\delta(1-\phi)} + \delta\bar{p}.$$

The following proves the aforementioned properties of p^*. We apply the intermediate value theorem to demonstrate the existence of one such p^*.

- We are currently assuming the upper bound condition holds: $\frac{\delta}{1-\delta}\big(p^{\max} - \delta\bar{p}\big)(1-\phi) > 1$.
- The lower-bound condition is $\frac{\delta}{1-\delta}\big(\bar{p} - \delta\bar{p}\big)(1-\phi) < 1$, which easily rearranges to a statement we know is true (because each of the constituent terms on the left-hand side are strictly less than 1): $\delta\bar{p}(1-\phi) < 1$.
- Continuity is trivial to establish.

Finally, the unique threshold claim follows because $\frac{\delta}{1-\delta}\big(p_t - \delta\bar{p}\big)(1-\phi)$ strictly increases in p_t.

How challenger strength affects the per-period probability of conflict Throughout the analysis in the paper, when we assess prospects for conflict, we mean prospects for an equilibrium in which conflict occurs along the path of play. Here we extend the analysis by considering how challenger strength affects outcomes within the set of parameter values in which conflict occurs along the equilibrium path. Along a conflictual equilibrium path, the per-period probability of conflict (assuming none has occurred previously) is the probability of drawing $p_t > p^*$, which equals $1 - F\left(\frac{1-\delta}{\delta(1-\phi)} + \delta\bar{p}\right)$.

Increasing the challenger strength changes two terms in this expression: $\bar{p}$ and the F function. Suppose we define increasing the challenger strength as a uniform rightward shift in the probability of winning of conflict, such that this probability is $p_t + d$ for some constant $d > 0$. In this formulation, p_t is the "baseline" probability of winning, which still follows distribution F, and d is the change in strength from this baseline. Thus, we can use the expressions from above while replacing p_t with $p_t + d$, and $\bar{p}$ with $\bar{p} + d$. Consequently, the per-period probability of conflict is $Pr\Big(p_t + d > \frac{1-\delta}{\delta(1-\phi)} + \delta(\bar{p} + d)\Big) = 1 - F\left(\frac{1-\delta}{\delta(1-\phi)} + \delta\bar{p} - d(1-\delta)\right)$. This term strictly increases in d. Therefore, conditional on conflict occurring along the equilibrium path, a stronger challenger makes conflict occur sooner (on average). Intuitively, a uniform rightward shift in threats improves the challenger's continuation value from accepting (because they gain higher average offers in the future) and from fighting (because they win with higher probability). The latter term dominates the former term because it is not discounted by a period.

The binary threat case permits us to explore the effects of a shift in the distribution function itself. One notion of a stronger challenger is a higher frequency of maximum-threat periods, expressed by q. In the text, we demonstrated that higher q increases the range of parameter values in which the equilibrium is peaceful. However, conditional on the equilibrium path featuring conflict, higher q in fact *raises* the per-period probability of conflict. The rationale is straightforward. A large value of q raises the average threat by enough to guarantee peace (it is straightforward to verify that the condition in Proposition 1 always holds in the binary case if $q = 1$). However, the cause of the higher average threat is that maximum-threat periods arise more frequently—which means that conflict is expected to occur sooner if that event ever occurs along the equilibrium path. Overall, the effect of q on the per-period probability of conflict is non-monotonic: positive and strictly increasing until it drops to 0.

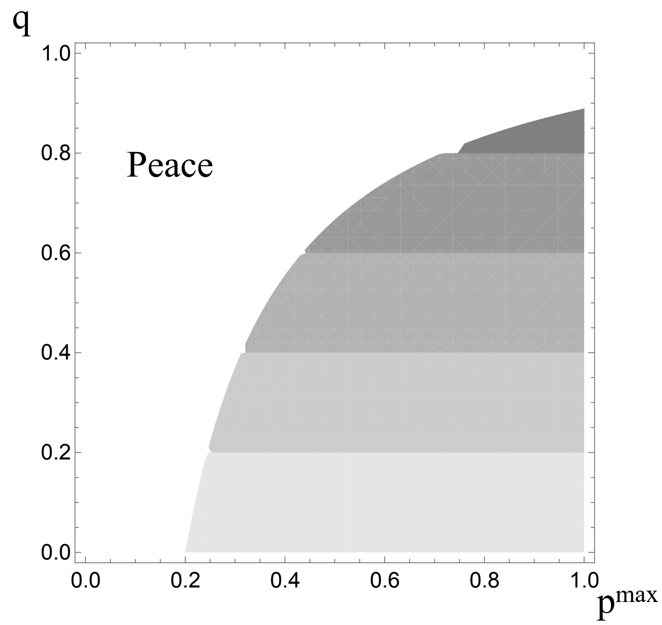
We can see this visually in Figure A.1. It has the same parameter values and general setup as in Figure 1, except now we provide information on what happens in a conflictual path of play. The per-period probability of conflict is 0 in the white area (i.e., a peaceful path of play), and is positive in the gray areas (i.e., a conflict path of play); and darker colors indicate a higher per-period probability of conflict. The non-monotonic effect of q is readily apparent: the total size of the conflict region is smaller for higher values of q, but conditional on conflict occurring along the equilibrium path, it is expected to occur sooner.

This finding highlights another twist in understanding the overall relationship between challenger strength and conflict. Depending on parameter values, a medium-sized increase in q can actually make conflict *more imminent*, whereas a large increase in q will eliminate conflict entirely.

A.2 Lower Bound on Offers

Here we extend the model to assume that the per-period offer must satisfy $x_t \in [\underline{x}, 1]$, for an exogenously specified $\underline{x} < 1$. A natural value to consider is $\underline{x} = 0$, that is, the ruler cannot demand net transfers away from the challenger, although the following results hold for more general values of $\underline{x}$. We derive these results under the specific case of binary challenger strength, while allowing strength to affect the minimum and maximum threats in addition to the probability of a maximum-threat period. Specifically, $p_t \in \{p^{\text{min}}, p^{\text{max}}\}$, with $q = Pr(p_t = p^{\text{max}})$. Let $x(p_t, \underline{x})$ be the offer made when the current-period threat is p_t and the lower bound on offers is $\underline{x}$. For the unbounded case we analyze in the text, we write $x(p_t, -\infty)$. At the end of this section, we comment on modeling a lower bound for the more general distribution of threats.

Figure A.1: Expected Time Until Conflict in Binary Threats Model



Parameter values: $\delta = 0.9$, $\phi = 0.5$, $p^{\min} = 0$.

By the analysis in the text, in any peaceful MPE, the offers in each period satisfy:

$$x^*(p^{\min}, -\infty) = \frac{\delta}{1-\delta}\Big((1-\delta(1-q))p^{\min} - \delta q p^{\max}\Big)(1-\phi)$$
$$x^*(p^{\max}, -\infty) = \frac{\delta}{1-\delta}\Big(p^{\max}(1-\delta q) - \delta(1-q)p^{\min}\Big)(1-\phi).$$

If $\underline{x} \leq x^*(p^{\min}, -\infty)$, then the lower bound never binds and the analysis is equivalent to the unbounded case. At the other extreme, if $\underline{x} > \delta p^{\max}(1-\phi)$, then the challenger accepts the basement offer even in a maximum-threat period.

If $\underline{x}$ is in-between these extremes, then along a peaceful equilibrium path, the ruler will offer $\underline{x}$ in a minimum-threat period, and make a higher offer in a maximum-threat period. In such an equilibrium, the offer made in a maximum-threat period must make the challenger indifferent between accepting and not:

$$x^*(p^{\max}, \underline{x}) + \frac{\delta}{1-\delta}\Big(qx^*(p^{\max}, \underline{x}) + (1-q)\underline{x}\Big) = \frac{\delta}{1-\delta}p^{\max}(1-\phi)$$

$$\implies x^*(p^{\max}, \underline{x}) = \frac{\delta p^{\max}(1-\phi) - \delta\underline{x}(1-q)}{1-\delta(1-q)}.$$

Given the upper bound of 1 for an offer, a peaceful MPE requires $x^*(p^{\max}, \underline{x}) \leq 1$. The offer in a maximum-threat period decreases in $\underline{x}$ because higher basement spoils increase the challenger's average consumption in future periods. We can rearrange to show that $x^*(p^{\max}, \underline{x}) \leq 1$ if and only if:

$$\underline{x} \geq 1 - \frac{1-\delta p^{\max}(1-\phi)}{\delta(1-q)} \equiv \underline{x}^{\text{peace}}. \tag{A.1}$$

This threshold is strictly less than 1, which means it is always possible to set $\underline{x}$ high enough to induce a peaceful equilibrium path of play.

Finally, we point out that there is no reason to believe that the core insights would not extend for the more general distribution of threats. However, the general case is difficult to characterize analytically. Intuitively, whenever p_t is lower than some bound $\underline{p}$, the ruler will offer exactly $x_t = \underline{x}$, and for all other periods the ruler will offer a higher value of x_t that makes the challenger indifferent between accepting and fighting. This breaks the linear structure of the offers in the baseline case. The specific complication is that the threshold $\underline{p}$ is endogenous to anticipated outcomes along the future path of play. This makes it difficult to characterize clean comparative statics on key parameters such as challenger strength.

A.3 Path-Dependent States

Despite the generality of our baseline model, one stark assumption is that Nature draws threat levels independently across periods. A simple way to introduce path-dependent states is to assume that with probability $r \in (0, 1)$, the challenger threat in period t is equal to p_{t-1}; and otherwise is drawn from the main distribution $F(p; s)$. Thus, higher values of r correspond to more persistent threat levels. The main findings here are that more persistent threats unambiguously make conflict less likely; and that when threats are sufficiently persistent, stronger challengers are unambiguously harder to buy off.

In this extension, the continuation value depends on the current value of p_t. Let $V^C(p_t)$ be the continuation value for entering the next period when the current threat is p_t. We can write the indifference condition as:

$$x_t(p_t) = \frac{\delta}{1-\delta} p_t(1-\phi) - \delta\Big(rV^C(p_t) + (1-r)V_n^C\Big), \tag{A.2}$$

where $V_n^C = \mathbb{E}[V^C(p_t)]$ is the continuation value if the threat is "new." We can write the continuation value with threat p_t as:

$$V^C(p_t) = x_t(p_t) + \delta\Big(rV^C(p_t) + (1-r)V_n^C\Big)$$

$$\implies V^C(p_t) = \frac{x_t(p_t) + \delta(1-r)V_n^C}{1-\delta r}.$$

Substituting this term back into Equation (A.2) yields:

$$x_t(p_t) = \frac{\delta}{1-\delta} p_t(1-\phi) - \delta\Big(r\frac{x_t(p_t) + \delta(1-r)V_n^C}{1-\delta r} + (1-r)V_n^C\Big)$$

$$\implies x_t(p_t) = \frac{\delta(1-\delta r)}{1-\delta} p_t(1-\phi) - \delta(1-r)V_n^C. \tag{A.3}$$

Importantly, and as in our baseline analysis, this expression is linear in p_t. As a result, we can solve for V_n^C as follows:

$$\begin{aligned} V_n^C &= \mathbb{E}[x_t(p_t)] + \delta V_n^C \\ \implies V_n^C &= \frac{\delta(1-\delta r)}{1-\delta}\bar{p}(1-\phi) - \delta(1-r)V_n^C + \delta V_n^C \end{aligned}$$

$$\implies V_n^C = \frac{\delta}{1-\delta}\bar{p}(1-\phi). \tag{A.4}$$

Note that this expression is the same as in the baseline case without path dependence, $r = 0$. Substituting Equation (A.4) back into Equation (A.3) provides an explicit characterization of the offer in each period:

$$\begin{aligned} x_t(p_t) &= \frac{\delta(1-\delta r)}{1-\delta}p_t(1-\phi) - \delta(1-r)\frac{\delta}{1-\delta}\bar{p}(1-\phi) \\ \implies x_t(p_t) &= \frac{\delta}{1-\delta}\Big((1-\delta r)p_t - \delta(1-r)\bar{p}\Big)(1-\phi). \end{aligned}$$

As $r \to 0$, we recover our baseline setup without path dependence. As $r \to 1$, threats do not change over time, and hence the optimal offer becomes the same as in a static version of the model, $(1-\phi)p_t$. This term is strictly less than 1, which means that any equilibrium path of play is peaceful. This is expected; the reason that fighting can occur along the equilibrium path in bargaining models with limited commitment is that threat levels fluctuate over time.

In general, peace is possible when:

$$\frac{1-\delta}{\delta(1-\phi)} \geq \underbrace{p^{\max} - \delta\bar{p}}_{\tau(s)} - \delta r(p^{\max} - \bar{p}) \equiv \tau(s, r) \tag{A.5}$$

This term clearly shows that higher r makes this inequality true for a wider range of parameter values; and at $r = 0$ it collapses to Equation (5). This inequality is harder to sustain for a stronger challenger if $\tau(s, r)$ increases in s:

$$(1-\delta r)\frac{\partial p^{\max}}{\partial s} - \delta(1-r)\frac{\partial \bar{p}}{\partial s} > 0.$$

When threats are sufficiently persistent, stronger challengers are unambiguously harder to buy off. To see this formally, if r is sufficiently large, then the second term in the preceding expression approaches zero, whereas the first term is strictly positive for any $\delta < 1$. Consequently, $\frac{\partial p^{\max}}{\partial s} > 0$ implies the preceding inequality must hold.

A.4 Fluctuating Costs of Conflict

In this section, we analyze a variant of the model in which the probability of winning is fixed but the cost of fighting fluctuates across periods. This more closely resembles the setup in Acemoglu and Robinson (2006), and the insights are qualitatively identical to our baseline model.

Suppose the probability of challenger victory is fixed at $p \in (0, 1]$, and the fraction of spoils that would permanently be destroyed by conflict is given by ϕ_t. (We rule out the trivial case $p = 0$, in which it is immediately apparent that the ruler survives while offering nothing in each period.) Each ϕ_t is iid and follows a distribution $G(\phi)$, with minimum value $\phi^{\min}$, maximum value $\phi^{\max}$, and average value $\bar{\phi}$.

By an identical logic as in our baseline model, the optimal transfer in every period must satisfy:

$$x^*(\phi_t) = \frac{\delta}{1-\delta}p(1-\phi_t) - \delta V^C.$$

In a peaceful MPE, the continuation value is written as follows. The first line is identical to the baseline setup except the integrand differs, and the final expression for V^C is identical except the average is taken over ϕ rather than p.

$$V^C = \frac{1}{1-\delta}\int_{\phi^{\min}}^{\phi^{\max}} \left[\frac{\delta}{1-\delta}p(1-\phi) - \delta V^C\right] dG(\phi)$$

$$\implies V^C = \frac{\delta}{1-\delta}p(1-\bar{\phi}).$$

Consequently, the optimal offer is:

$$x^*(\phi_t) = \frac{\delta}{1-\delta}p\Big(1-\delta-(\phi_t-\delta\bar{\phi})\Big).$$

The condition for a peaceful MPE is that it is possible to buy off the challenger when conflict destroys the smallest share of the pie, or:

$$\frac{\delta}{1-\delta}p\Big(1-\delta-(\phi^{\min}-\delta\bar{\phi})\Big) \leq 1.$$

This yields qualitatively identical comparative statics as the main analysis. If increasing challenger strength decreases the average amount destroyed by conflict but not the minimum amount, then this inequality is easier to meet, and so stronger challengers are easier to buy off peacefully. By contrast, if making the challenger stronger decreases $\phi^{\min}$ and $\bar{\phi}$ at an equal rate, then the opposite holds.

A.5 Endogenous Institutional Reform

In Appendix A.2, we extended the binary threat version of the model to incorporate an exogenous lower bound $\underline{x}$ on the ruler's per-period offer. Now we endogenize the choice of $\underline{x}$, which we interpret as endogenous institutional reform. In each period, after Nature realizes the challenger's threat, the ruler chooses $\underline{x}_t \in [\underline{x}_{t-1}, 1]$, with the initial level corresponding to that in the baseline game, $\underline{x}_0 = -\infty$. This means that the institutional choice in any period is a dynamic state variable and creates a floor for the offer in all future periods; the ruler can subsequently choose to raise this floor, but not lower it. This choice could capture a wide range of institutional reforms, such as a power-sharing agreement, expanding the franchise, or civil rights protections.

We begin by presenting three preliminary results. First, if the inequality in Proposition 1 is met, then the ruler will not set $\underline{x} \geq -\infty$. A deviant choice would either have no impact on the outcome the game, or would redistribute more surplus than needed to buy off the challenger. The interesting case is when the inequality in Proposition 1 is not met, and hence conflict will occur along the equilibrium path absent reform, on which we focus for the remainder of the analysis.

Second, the ruler never has a strict preference to reform institutions in a minimal-threat period. Doing so would simply deliver (weakly) more transfers to the challenger in a period in which they can already be induced to accept, and has no impact on the ruler's ability to buy off the challenger in a maximum-threat period (because, in such a period, the ruler can instantaneously increase the basement level of transfers).

Third, if the ruler makes institutional reforms, they will be "large." Recall from Equation (A.1) that $\underline{x}^{\text{peace}}$ is level of $\underline{x}$ at which the challenger is indifferent between accepting an offer of 1 and fighting in a maximum-threat period. This is the lowest level of $\underline{x}$ that induces a peaceful path of play. It is straightforward to rule out any finite choice $\underline{x}_t < \underline{x}^{\text{peace}}$ as the optimal level of institutional reform. Such a choice does not change the challenger's preference to fight in maximal-threat periods, and simply delivers weakly more spoils to the challenger in minimal-threat periods in which they would accept anyway.

Given these preliminary results, we ask: in a maximum-threat period, if conflict would otherwise occur, will the ruler make institutional reforms sufficiently large to buy off the challenger? The following proves that the answer is always yes. We already know the ruler's lifetime expected utility if a conflict occurs in a maximum-threat period:

$$\frac{\delta}{1-\delta}(1-p^{\max})(1-\phi). \tag{A.6}$$

Alternatively, upon choosing $\underline{x}_t \geq \underline{x}^{\text{peace}}$, the ruler's lifetime expected utility is:

$$1 - x^*(\underline{x}) + \frac{\delta}{1-\delta}\Big(q(1 - x^*(\underline{x})) + (1-q)(1-\underline{x})\Big), \tag{A.7}$$

where $x^*(\underline{x})$ is the offer that makes the challenger indifferent between accepting and fighting in a maximum-threat period, given institutions $\underline{x}$. Consequently, this term satisfies:

$$x^*(\underline{x}) + \frac{\delta}{1-\delta}\Big(qx^*(\underline{x}) + (1-q)\underline{x}\Big) = \frac{\delta}{1-\delta}p^{\max}(1-\phi). \tag{A.8}$$

Upon solving Equation (A.8) for $x^*(\underline{x})$ and then substituting back into Equation (A.7), we yield a lifetime expected utility for the ruler of:

$$\frac{1-\delta p^{\max}(1-\phi)}{1-\delta}. \tag{A.9}$$

Finally, we compare Equations (A.6) and (A.9) to yield a true inequality, thus completing the proof:

$$\frac{1-\delta p^{\max}(1-\phi)}{1-\delta} > \frac{\delta}{1-\delta}(1-p^{\max})(1-\phi) \implies 1 > \delta(1-\phi).$$

One notable attribute about the preceding proof is that conditional on making a large-enough institutional reform to induce peace, the ruler is in fact indifferent about the amount of institutional reform. There are a continuum of equilibrium choices in which the ruler chooses between a bit more institutional reform and somewhat fewer temporary transfers, and vice versa. We focus on the MPE with the minimum-necessary institutional reforms, which is consistent with microfoundations for such a choice posited in Dower et al. (2018) and Powell (2020).

Along the equilibrium path, the ruler does not choose institutional reform until the first maximum-threat period, when she implements reform. Formally, the ruler optimally sets $\underline{x}_t = \max\{-\infty, \underline{x}_{t-1}\}$ in every minimum-threat period and $\underline{x}_t = \max\{\underline{x}^{\text{peace}}, \underline{x}_{t-1}\}$ in every maximum-threat period; the max function accounts for the inability to lower basement spoils below those chosen in previous periods.

Given this result, the comparative statics on s are identical to those in the baseline game. We simply replace the conflict region in Figure 1 with a "reform" region. In other words, the parameter values in the baseline model for which conflict would ensue is identical to the parameter values in the present extension for which institutional reform will occur.

Therefore, higher $p^{\max}$ increases the range of parameter values in which any institutional reform occurs. An additional result is that higher $p^{\max}$ also increases the extent of institutional reforms (conditional on any occurring). To establish this result, we differentiate $\underline{x}^{\text{peace}}$ (see Equation (A.1)) with respect to $p^{\max}$. Increasing the challenger's opportunity cost to not fighting in a maximum-threat period causes them to demand more guaranteed concessions in the future.