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

Andrew T. Little* Jack Paine†

December 13, 2022

Abstract

Prominent theories propose that commitment problems drive phenomena such as war and democratization. However, existing work disagrees about a basic question: how does a challenger's coercive strength affect prospects for conflict and/or institutional reform? We establish that the relationship depends on how challenger strength affects their average and maximum threats. We analyze a formal model with a general distribution of the probability that the challenger would win a conflict in a given period ("threat"), and we conceptualize challenger strength as affecting the distribution of their threats. If the maximum threat is fixed and stronger challengers pose a higher average threat, then weak challengers will rebel (absent reform) during the rare periods they pose a high threat. However, if stronger challengers pose a greater maximum threat, then they are harder to buy off. Applying these insights advances theoretical and empirical debates about democratization.

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

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 threat that a challenger poses to overthrow the government 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. Even if a challenger poses a stark threat today, they may not tomorrow—which means they will lose their ability to compel the government for concessions. Consequently, a temporarily strong challenger will seize their window of opportunity and fight to overthrow a government that cannot commit to a favorable distribution of future spoils. Alternatively, the government may respond to a credible threat by reforming institutions (e.g., democratization, sharing power),which bolsters their commitment to share future spoils with the challenger.

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?

[1]For democratization, see Acemoglu and Robinson (2006); Ansell and Samuels (2014); Leventoğlu (2014); Castañeda Dower et al. (2018). For authoritarian power sharing and democratic separation of powers, see Helmke (2017); Christensen and Gibilisco (2022); Meng (2019); Powell (2021); 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).

How does a bigger non-elite class, a better-organized civil society, or a more advanced neighboring state affect these outcomes?

The core takeaway from this article is that different ways of conceptualizing the challenger's coercive strength yield divergent answers. Theoretically, we examine a more general distribution of threats—with "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 their maximum and average threats. To our knowledge, we are the first to consider any discrete or continuous distribution of threat levels or, more generally, any random variable that 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 is prohibitively destructive. By contrast, maximum threats arise when the challenger would win a (moderately costly) conflict with certainty, which creates a dire threat. In this setup with binary threat levels and fixed values for each, a natural way to capture coercive strength is the probability in any period that the challenger will pose their maximum threat—with stronger challengers posing the maximal threat more frequently. Counterintuitively, in this setup, *weaker challengers are more prone to fight*. Their rare windows of opportunity are too tempting to pass up and to forgo revolting, given their poor prospects to gain concessions in the future if the status quo regime remains intact. Thus, weak challengers lead to either more conflict or, if the ruler is able, more institutional reform.

We indeed recover this scenario as a special case in our model. The underlying commitment problem is less pressing for a challenger who poses a higher *average* threat. This source of strength enables the challenger to more frequently compel concessions from the ruler over time, which lessens their motives to fight.

However, as we demonstrate, this is not a general result about the consequences of stronger challengers. The seemingly innocuous assumption to construct the distribution of threats such that only the average threat fluctuates creates an unrecognized problem: it holds fixed a crucial margin, the *maximum* threat. If stronger challengers correspond with a higher maximum threat, then *stronger challengers are more prone to fight* when their maximum threat is realized because they have a greater opportunity cost to forgoing conflict. In this scenario, stronger challengers are also better positioned to force institutional reform.

Generally, coercively strong challengers should pose greater threats both on average and when maximally strong. This means that the overall effect of greater 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 (because the challenger does not enjoy the benefits of a higher average threat until future periods). For example, a uniform upward shift in the distribution of threats makes the challenger harder to buy off peacefully and makes institutional reform more likely. Even in a simple binary threats model, we can recover our core result 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 maximum and average threats.

After analyzing the model, we highlight new insights for debates about democratization and authoritarian power sharing. In models such as Acemoglu and Robinson (2006) and Castañeda Dower et al. (2018), coercively weak challengers trigger institutional reform. A low average threat makes the shadow of the future unfavorable. This feature, 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.

Our 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. Fortunately, our analysis highlights paths forward. For tests that compare challenger strength across cases, it is important to scrutinize the perceived permanence of the group who organizes against the government, as more durable organizations yield higher average threats. This helps to account for discrepancies between the empirical findings in Aidt and Franck (2015) and Castañeda Dower et al. (2018). Tests that compare challenger strength over time within cases are less subject to our main theoretical point because any model in this class predicts that conflict (or institutional reform) is more likely in periods in which the challenger's *current* threat is high relative to the *average* threat they pose. This applies to research designs that study, for example, rainfall-induced economic shocks (Brückner and Ciccone, 2011; Aidt and Leon, 2016). Yet more discussion is needed for why the maximum threat should tend to be high relative to the average threat in the settings they study, which is necessary for high-threat periods to produce institutional reform. We conclude by discussing the importance of formalization and directions for future empirical and theoretical research.

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 $x_t \leq 1$. This incorporates the standard assumption that the ruler cannot transfer more than the entire contemporaneous budget in any period, and hence cannot borrow across periods. 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 permanently destroys a fraction $\mu \in (0, 1)$ of total spoils, 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. Formally, assume 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.[2] To streamline the exposition, we suppress s when doing so does not cause confusion.

Following the model analysis, we discuss simplifying assumptions and summarize extensions presented in full in the appendix: imposing a lower bound on the bargaining offer, allowing for a path-dependent distribution of threats, modeling fluctuations in the cost of conflict rather than the probability of winning to more closely parallel the setup in Acemoglu and Robinson (2006), and allowing the ruler to engage in institutional reform.

2.2 Analysis: How the Distribution of Threats Affects Conflict

We start by asking when a Markov Perfect Equilibrium (MPE) exists in which conflict occurs with probability 0 in every period along the equilibrium path. We refer to this as a peaceful equilibrium.

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

[2] A sufficient but not necessary condition for this to hold is if $F(p, s_1)$ has first-order stochastic dominance over $F(p, s_2)$ for any $s_1 > s_2$.

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:

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

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:

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

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:

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

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

This term is simply the average probability of winning multiplied by the challenger's lifetime expected consumption from winning a conflict. Combining Equations (2) and (3) enables us to explicitly solve for the equilibrium per-period offer:

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

The two conditions for a peaceful MPE to exist are that (1) the ruler is able to make such an offer and (2) the ruler prefers making this offer in each period to experiencing conflict. We can see in Equation (4) that $x^*(p_t)$ strictly increases in p_t. Therefore, for the first condition to hold, we need the optimal offer at $p_t = p^{\max}$ to be less than 1, or $x^*(p^{\max}) \leq 1$. The second condition always holds as the ruler extracts all surplus saved from not fighting in a period of peaceful bargaining. The existence of bargaining surplus also means that a strategy profile with these features is the unique MPE when the first condition holds: in any strategy profile for which conflict occurs along the path of play, the ruler can deviate to make the challenger an acceptable offer, which would improve the ruler's payoff. We summarize this result in Proposition 1. If instead $x^*(p^{\max}) > 1$, then the equilibrium is qualitatively similar despite featuring conflict along the equilibrium path. Specifically, there is a unique MPE in which the ruler buys off the challenger whenever possible, but conflict occurs in the first period in which the contemporaneous threat p_t exceeds a critical threshold.

Proposition 1 (Equilibrium)**.** *If $\frac{(p^{max}-\delta\bar{p})(1-\mu)}{1-\delta} \leq 1$, then the following strategy profile is the unique MPE: in every period, the ruler proposes $x_t = x^*(p_t)$ according to Equation (4) and the challenger accepts any $x_t \geq x^*(p_t)$. Along the equilibrium path, the challenger accepts in every period. If $\frac{(p^{max}-\delta\bar{p})(1-\mu)}{1-\delta} > 1$, then any MPE has conflict along the equilibrium path of play (see Proposition A.1 for details).*

The inequality that forms the crucial scope condition for Proposition 1, $\frac{(p^{\max}-\delta\bar{p})(1-\mu)}{1-\delta} \leq 1$, enables us to analyze how changing challenger strength affects the possibility of a peaceful equilibrium by taking comparative statics on s. If we move the threat parameters to one side of the inequality and write them explicitly as a function of s, this condition becomes:

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

We say that the prospects for conflict are increasing in s if $\tau(s)$ increases in s because this expands the range of parameters in which conflict occurs along the equilibrium path. Similarly, we say the prospects for peace are decreasing in s if $\tau(s)$ decreases in s. The overall effect of increasing the challenger's strength on the prospects for peace and conflict can be summarized as follows:

Proposition 2. *Increasing challenger strength raises the prospects for conflict, that is, $\tau(s)$ strictly increases in s, if and only if:*

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

The proof follows directly from the following derivative:

$$\frac{\partial \tau(s)}{\partial s} = \frac{\partial p^{\max}(s)}{\partial s} - \delta\frac{\partial \bar{p}(s)}{\partial s}.$$

This result 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 needed for a peaceful path 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. A higher average threat lowers the opportunity cost of not fighting. The challenger expects more 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. Notice that $\frac{\partial \bar{p}(s)}{\partial s}$ is multiplied by δ whereas $\frac{\partial p^{\max}(s)}{\partial s}$ is not. The reason is that the challenger reaps the benefits of higher $p^{\max}$ in the present period whereas the consumption gains from higher $\bar{p}$ begin to accrue starting only in the next period.

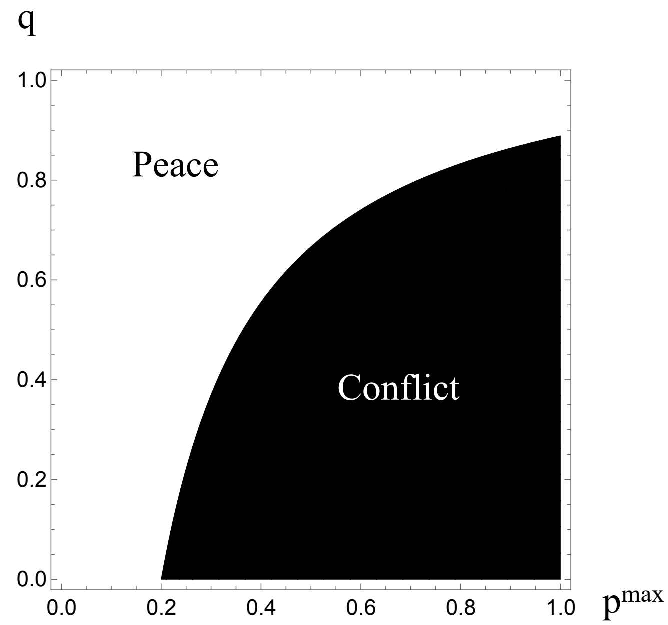
General binary distribution To connect this result more directly to past work, suppose the per-period threat takes one of two values. We write these as $p_t \in \{p^{\min}, p^{\max}\}$ with $0 \leq p^{\min} < p^{\max} \leq 1$ and $q = Pr(p_t = p^{\max})$. In this case, the average threat is $\overline{p}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

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

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 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

Figure 1: Peace and Conflict in the Binary Threats Model



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

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 by raising q and therefore boosting their average threat without altering the opportunity cost of fighting in the maximum-threat state, which is dictated by $p^{\max}$. 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 q remains 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 the distribution of threats F uniformly to the right. This corresponds to raising the challenger's probability of winning by a fixed amount in each period. Hence, the minimum and maximum threats each increase by some constant $d > 0$ but the per-period probability of each threat realization, q, is unchanged. In this case, facing a stronger challenger makes *peace harder to sustain* for the reason discussed after Proposition 2: the challenger immediately reaps the benefits of a higher maximum threat, but does not gain the benefits of a higher minimum (and therefore average) threat until future periods. Formally, we can see this 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 challenger strength and structures the parameters in the distribution of threats 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 bargaining offers. This makes it possible for the ruler to offer $x_t < 0$ and hence to demand a net transfer from the challenger. Substantively, we can interpret the "spoils" under consideration as only part of the resources available in the society, which enables the ruler in principle to extract more whenever the challenger poses a weak threat. Mathematically, the case without a lower bound is analytically simpler because the ruler can hold the challenger down to their reservation value in every period. This makes the optimal offer 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 lower bound of zero is binding. 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. With positive probability, Nature does not change the state of the world in the next period; but 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.

In our model, the probability of winning p_t is the random variable. Some related models instead fix the probability of winning and allow the cost of conflict to vary across periods, for example, Acemoglu and Robinson (2006). In their setup, the challenger is "stronger" when this cost, which we (and they) denote as μ, is more frequently low. In our view, allowing the probability of winning to vary is a more natural way to capture the notion of challenger strength. However, all that matters for our results is how the challenger's expected value to fighting changes over time—and the calculus here is identical regardless of whether the probability of winning or the costliness of fighting fluctuates over time. In Appendix A.4, we present a version of the model where the cost of conflict μ_t fluctuates over time to demonstrate qualitatively identical results.

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, has 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, to 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: Adjudicating Divergent Theoretical Implications

To illustrate the substantive importance of our findings, we engage with debates about causes of democratization and authoritarian power sharing. In this section we adjudicate divergent theoretical implications, and in the next we discuss implications for empirical research designs.

In Acemoglu and Robinson's baseline model of authoritarian politics,[3] 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 $p^{\text{min}} = 0$ and $p^{\text{max}} = 1$.[4] Thus, in maximum-threat periods, the masses can credibly threaten to stage a revolution, which succeeds with probability 1 and removes elites from power forever. In every maximum-threat period, elites would ideally like to buy off the masses with a temporary concession: setting a high tax rate and redistributing wealth in that period only. However, elites cannot credibly commit to make similarly generous concessions in any future periods 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, temporary concessions are insufficient to pacify the masses because their shadow of the future is unfavorable. The high frequency of minimum-threat period results in low consumption, given elites' limited commitment ability.

Acemoglu and Robinson then extend their framework to explain endogenous institutional reform.[5] 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

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

[4] Again, this is a slightly different interpretation of their stochastic cost-of-fighting parameter, but is conceptually equivalent (see Appendix A.4).

[5] 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.

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 to weak but not strong challengers.

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. In nineteenth-century European countries, industrialization created a stronger capitalist class that was better positioned to challenge landed elites who monopolized power. Rather than fix $p^{\max} = 1$, they parameterize the challenger's probability of winning in a similar fashion to our term $p^{\max}(s)$. They conclude that stronger challengers have greater 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. Castañeda 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 initially pose a high maximum threat anticipate a larger adverse shift in the future distribution of power. This makes stronger challengers more prone to stage a coup if the ruler does not share power with them at the outset, which induces power sharing with strong but not weak challengers. Here, greater challenger strength affects the maximum threat more than the average threat.

In Paine (2022), the relationship between challenger strength and prospects for both 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 reconcile seemingly inconsistent implications 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, and usually undiscussed, 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 theorizing. Seemingly technical model assumptions carry important substantive implications.

4 Application: Implications for Empirical Tests

Empirical examinations of democratization models with commitment problems typically address either of two questions. (1) Across cases, where should we expect democratic reform? (2) Within cases over time, under what conditions should we expect democratic reform? These models offer ambiguous implications for the first question for the reasons we have discussed. Challenger strength differs across cases, but without further specification, we do not know whether comparatively stronger challengers are more or less likely to fight or to gain institutional concessions (see Proposition 2). By contrast, these models offer an unambiguous prediction for the second question: a challenger is more likely to fight (or gain institutional reforms) in time periods when it poses a higher-than-average threat. This follows from the straightforward point, discussed in the analysis, that the opportunity cost of forgoing fighting strictly increases in p_t, and therefore $x^*(p_t)$ strictly increases in p_t (see Equation 4). We discuss existing empirical tests from this perspective.

4.1 Cross-Case Comparisons

Drawing conclusions from comparisons of challenger strength across cases is inherently difficult because the theoretical implications are ambiguous. Exemplifying this point, two leading empirical evaluations of the Acemoglu and Robinson (2006) model assess opposing hypotheses about challenger strength. In both cases, further discussion of key scope conditions is needed to more closely tie the empirical test to theoretical implications.[6]

Castañeda Dower et al. (2018) study endogenous representation for peasants in Imperial Russia. Reforms in 1864 created district-level assemblies, *zemstva*, which varied in the fraction of seats reserved for peasants. The authors use the frequency of protests in each district over the preceding decade to proxy for the average threat they would pose in the future, which we (and they) formalize as the q parameter. They find that high levels of past unrest engendered less representation for peasants, hence demonstrating that higher average threats led to less institutional reform. Alone among existing empirical tests, they explicitly engage with the idea the idea from the Acemoglu and Robinson (2006) model that the masses and institutional reformers consider not only the present threat, but also the expectation about future threats.

However, our analysis of a more general distribution of threats identifies a key scope condition for the theoretical implication: coercively stronger challengers (as proxied by the frequency of unrest in the 1850s) must have posed a higher average than maximum threat. Otherwise, if such districts posed (comparatively) very large threats in the mid 1860s when the reforms were implemented, then we might expect them to gain greater reforms. The authors' careful discussion of the

[6]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.

historical context does not address this point specifically.

By contrast, Aidt and Franck (2015) focus on the present threat posed by the masses. They leverage local variation in the intensity of Swing Riots to measure how British MPs perceived the threat level in their districts, and they analyze how these perceptions affected roll-call 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 had been temporarily overcome, that is, the masses posed their maximum threat and this threat was ominous (i.e., high $p^{\max}$). By focusing on the current threat, they anticipate that MPs were *more* likely to vote for reform when more riots and protests occurred in their district.

Yet comparing this hypothesis to the opposing one tested in Castañeda Dower et al. (2018) again highlights the additional steps need to link the theory to empirics. Aidt and Franck (2015) assume that strong challengers pose purely transitory threats and hence their average threat is low. However, suppose instead that riots and protests proxy for districts in which the masses posed *persistently* strong threats, even if not activated at all points in time. By refocusing on average threats, the model would anticipate that MPs in high-protest districts would be able to pacify the recalcitrant masses with temporary concessions rather than permanent reforms. Under these scope conditions, we would expect them to vote against the Reform Act.

The historical setting of each study differs in an important way that motivates the plausibility of the specific hypothesis assessed in each. For Aidt and Franck (2015), the Swing Riots were unprecedented attacks by peasants on agricultural infrastructure such as threshing machines and barns. Although the underlying economic grievances were long standing, the movement itself lacked coherent organization. Therefore, it is plausible that the contemporaneous (maximum) threat was high but the future (average) threat was low. By contrast, the types of anti-serfdom riots analyzed in Castañeda Dower et al. (2018) could, plausibly, have been sustained over longer periods, which would raise their average threat. More generally, it is important in empirical tests to identify which groups and organizations are inherently more transitory and which have greater staying power.

4.2 Within-Case Comparisons

Our analysis suggests a more direct specification to test theories of dynamic commitment problems with time series data: comparing within a single case an estimate of the threat *at a particular time* to an estimate of the expected threat in the future. Assuming one can come up with reliable estimates of both quantities (and setting aside other causal inference challenges), the current threat should be positively associated with conflict whereas the expected future threat should be negatively associated with conflict. Put another way, although models of dynamic commitment problems make ambiguous *cross-case* predictions about challenger strength, they make straightforward *within-case* predictions about what kinds of periods are most likely to involve conflict or institutional reform.

One example is empirical tests that examine how shocks in rainfall and the onset of droughts in Africa have influenced democratic reforms (Brückner and Ciccone, 2011; Aidt and Leon, 2016). These studies link their research designs to the Acemoglu and Robinson (2006) model and emphasize that transitory economic shocks create the conditions under which the passes pose a temporary, but not permanent, threat of revolution. Their empirical tests focus on Sub-Saharan Africa because these economies are predominantly agricultural by contemporary global standards. Therefore, fluctuations in rainfall should more greatly affect economic output.

These empirical tests analyze changes over time within cases, a design for which the ambiguous theoretical effects of challenger strength are less problematic. However, even in these cases, more extensive discussion of scope conditions would further tighten the connection between the empirical tests and the theoretical model. Transitory economic shocks lead to democratic reforms in equilibrium only in societies for which the average threat is low and the maximum threat is high. This scope condition appears plausible in this setting. During the Cold War, most African countries were closed dictatorships with no meaningful mass political participation, which made mobilization difficult (low average threat). Yet many of these states had tenuous control over their entire territory and weak command over the national military, which made the maximum threat high in the rare periods in which the masses could solve their collective action problem. In light of our theoretical discussion, we encourage authors to routinely address (and elaborate upon) points such as these.

5 Conclusion

How does a challenger's coercive strength affect prospects for conflict and/or institutional reform? We established that the relationship depends on how challenger strength affects their average and maximum threats. Higher average threats lead to less conflict and fewer institutional concessions, whereas higher maximum threats yield the opposite implications. In general, a stronger challenger—such as a bigger non-elite class, a better-organized civil society, or a more advanced neighboring state—poses a greater average and maximum threat, which yields ambiguous theoretical implications. We summarized existing theoretical debates and empirical tests from this perspective.

Our analysis underscores the importance of formalizing theoretical intuitions. We highlight that existing work conceptualizes challenger strength in different ways that lead to divergent theoretical implications. However, the ambiguous consequences of challenger strength are not product of nor a flaw with game-theoretic modeling per se. Instead, formalization enables us to clarify the conceptual difference between average and maximum threats, explain precisely why the implications are ambiguous, and characterize the conditions under which the implications cut in one direction or the other. Our modeling approach, in turn, is possible because of the advances in applied formal theory pioneered by the authors referenced throughout this article as well as many others. Clarifying theoretical implications is also crucial for empirical testing.

We highlight some fundamental impediments to empirically measuring key parameters from models of dynamic commitment problems, but we also provide some suggestions for which empirical tests are most convincing (at least with regard to the considerations raised here). Within-case comparisons elide the main source of theoretical ambiguity, although cross-case comparisons are viable if authors specify the extent to which the challenger they study has long-run staying power. Another possible approach in future work is to use structural models.[7] In general, this estimation technique is helpful when models make countervailing predictions about comparative statics and the theoretical parameters are difficult to measure directly (see, for example, Crisman-Cox and Gibilisco 2018). The goal would be to find the version of the model, in particular the distribution of threats $F(p)$, that best fits the data and then to compare the frequency of conflict and democratic reforms under counterfactual distributions. Of course, given limitations to available data, it may be difficult to distinguish distributions that vary only on the maximum threat level.

Our model can also be used as a foundation for future theoretical work. 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. Beyond "challenger strength" specifically, our theoretical results provide 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). Overall, future work can build on our insights to examine how factors such as repression, technology for mobilization, and economic fundamentals affect prospects for conflict and institu-

[7] We thank an anonymous reviewer for highlighting this point.

tional reform.

References

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

Aidt, Toke S and Gabriel Leon. 2016. "The Democratic Window of Opportunity: Evidence from Riots in Sub-Saharan Africa." *Journal of Conflict Resolution* 60(4):694–717.

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.

Brückner, Markus and Antonio Ciccone. 2011. "Rain and the Democratic Window of Opportunity." *Econometrica* 79(3):923–947.

Castañeda Dower, Paul, 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.

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. 2022. "How Budgets Shape Power Sharing in Autocracies." Working paper.

Crisman-Cox, Casey and Michael Gibilisco. 2018. "Audience Costs and the Dynamics of War and Peace." *American Journal of Political Science* 62(3):566–580.

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

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. 2021. "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 fully characterize the equilibrium, including a proof for Proposition 1 and a formal description of the strategies when conflict occurs along the equilibrium path. 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 Full Equilibrium and Conflictual Paths of Play

We first prove Proposition 1. We then finish the equilibrium characterization of the baseline model by analyzing the parameters in which conflict occurs along the path of play, and discuss the per-period probability of conflict along conflictual paths of play. Note that a Markovian strategy for the ruler specifies a bargaining offer $x(p_t)$ as a function of the current-period threat p_t, and a Markovian strategy for the challenger specifies a response function as a function of p_t and x_t.

Proof of Proposition 1. The proof of existence and uniqueness proceeds in six steps. We first provide some properties of any equilibrium while treating the continuation value for the challenger as fixed, and then show that these properties uniquely pin down that the continuation value must be the expected value of fighting in the next period. From there we can easily establish the optimal per-period offer in a peaceful MPE and that the MPE is peaceful if and only if these offers are feasible for all draws of p_t.

Step 1. In any MPE, the challenger accepts with probability 1 any offer that satisfies $x_t \geq \frac{p_t(1-\mu)}{1-\delta} - \delta V^C$ (see Equation 1) and accepts with probability 0 otherwise. We denote as V^C the challenger's continuation value given the strategy profile (which is not a function of past play, given the Markovian restriction). *Proof*: The proof follows immediately from a basic comparison of payoffs for all cases except that in which the expression holds with equality. For the latter case, we can rule out as equilibria all strategies in which the challenger, when indifferent, accepts with less than probability 1 because the ruler can profitably deviate to an infinitesimally larger offer, and therefore lacks a best response.

Step 2. Suppose $1 + \delta V^C \geq \frac{p_t(1-\mu)}{1-\delta}$. Then in any MPE, the ruler never offers an amount $x_t > \frac{p_t(1-\mu)}{1-\delta} - \delta V^C$. *Proof:* The ruler's future continuation value is identical for any offer such that $x_t \geq \frac{p_t(1-\mu)}{1-\delta} - \delta V^C$ because, as the previous step established, the challenger will accept any such offer with probability 1. Consequently, any offer satisfying this inequality affects only the ruler's current-period payoff. The ruler can profitably deviate from any strategy in which the offer strictly satisfies the inequality because an infinitesimally smaller offer would enable the ruler to consume more without triggering the challenger to fight. Finally, the assumed inequality $1 + \delta V^C \geq \frac{p_t(1-\mu)}{1-\delta}$ implies that $x_t = \frac{p_t(1-\mu)}{1-\delta} - \delta V^C$ is feasible.

Step 3. Suppose $1 + \delta V^C < \frac{p_t(1-\mu)}{1-\delta}$. Then in any MPE, the challenger rejects any offer. *Proof*: Follows directly from the assumed upper bound on x_t.

Step 4. In any MPE, $V^C = \frac{\bar{p}(1-\mu)}{1-\delta}$. *Proof*: This is clearly a lower bound for the continuation value because the challenger can always choose to fight in the next period, which yields this amount in expectation. To demonstrate that it is an upper bound, along an arbitrary equilibrium path, we can assign a mixed probability of acceptance $\alpha(p_t)$ for each state of the world (which also takes into account the ruler's optimal offer for that state). Per Step 3, acceptance can occur only if the condition in Step 2 is met. Therefore, from Step 2, we know that $\frac{p_t(1-\mu)}{1-\delta} - \delta V^C$ is an upper bound on current-period consumption. Therefore, the upper bound on the future continuation value is

$$\int_{p^{\min}}^{p^{\max}} \left[\alpha(p_t) \Big(\underbrace{\frac{p_t(1-\mu)}{1-\delta} - \delta V^C +}_{\text{Current period}} \underbrace{\delta V^C}_{\text{Future periods}} \Big) + \big(1-\alpha(p_t)\big) \frac{p_t(1-\mu)}{1-\delta} \right] dF(p).$$

This simplifies to $\int_{p^{\min}}^{p^{\max}} \frac{p_t(1-\mu)}{1-\delta} dF(p) = \frac{\bar{p}(1-\mu)}{1-\delta}$. Thus the upper bound equals the lower bound, completing the proof.

Step 5. Suppose $\frac{(p_t-\delta\bar{p})(1-\mu)}{1-\delta} \leq 1$. Then in any MPE, the ruler proposes $x_t = x^*(p_t) \equiv \frac{(p_t-\delta\bar{p})(1-\mu)}{1-\delta}$ (see Equation 4). *Proof*: Substituting V^C as solved in Step 4 into the inequality from Step 1 demonstrates that the challenger will accept this offer, and therefore the result from Step 2 establishes that the ruler cannot profitably deviate to a higher offer. We can also see that the challenger will reject any lower offer. Therefore, we need to check profitable deviations to an action that triggers conflict, which yields a lifetime expected utility for the ruler of $\frac{(1-p_t)(1-\mu)}{1-\delta}$. To demonstrate that this is not profitable, we need to demonstrate that the ruler can ensure herself a higher payoff from securing acceptance in the present period. Following a period of acceptance, we can bound the ruler's future continuation value from below at $\frac{(1-\bar{p})(1-\mu)}{1-\delta}$ because the ruler can always trigger the challenger to fight in the next period. Thus the claim requires showing

$$\underbrace{1 - \frac{(p_t - \delta\bar{p})(1-\mu)}{1-\delta}}_{\text{Peace now}} + \underbrace{\delta\frac{(1-\bar{p})(1-\mu)}{1-\delta}}_{\text{Conflict in next period}} > \underbrace{\frac{(1-p_t)(1-\mu)}{1-\delta}}_{\text{Conflict now}}.$$

This reduces to $\delta(1-\mu) > 0$, which always holds.

Step 6. There is a unique peaceful MPE if and only if $\frac{(p^{\max}-\delta\bar{p})(1-\mu)}{1-\delta} \leq 1$. *Proof*: Note that $x^*(p_t)$ strictly increases in p_t. Therefore, if $x^*(p^{\max}) \leq 1$, then this inequality holds for all p_t. Per the proceeding steps, the unique MPE features offers which are accepted in each period. If instead this inequality is violated, then conflict must occur along the equilibrium path because the challenger rejects all offers when $p_t = p^{\max}$. □

We now provide a full characterization of equilibrium strategies when the condition in Proposition 1 is violated.

Proposition A.1 (Conflictual equilibrium). *If $\frac{(p^{max}-\delta\bar{p})(1-\mu)}{1-\delta} > 1$, then there is a unique class of payoff-equivalent MPE in which conflict occurs along the path of play. In these MPE, there is a unique $p^* < p^{max}$ such that (i) when $p_t \leq p^*$, the strategies are the same as in Proposition 1, and (ii) when $p_t > p^*$, then the challenger rejects all offers and the ruler's strategy can involve making* *any offer.*

Proof The proof of Proposition 1 provides most of the elements needed to establish existence and uniqueness. The main additional step is to formally define p^* as:

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

The inequality that $p^* < p^{\max}$ follows from the present assumption that $\frac{(p^{\max}-\delta\bar{p})(1-\mu)}{1-\delta} > 1$.

The unique optimality of the challenger's accept/fight decisions follows immediately from the construction of p^* and from the steps in the proof of Proposition 1; as do the offers from the ruler when $p \leq p^*$. When $p > p^*$, all offers are rejected, and therefore the ruler is indifferent among all offers. □

How challenger strength affects the per-period probability of conflict Throughout the analysis in the article, 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 $1 - F\left(\frac{1-\delta}{1-\mu} + \delta\bar{p}\right)$.

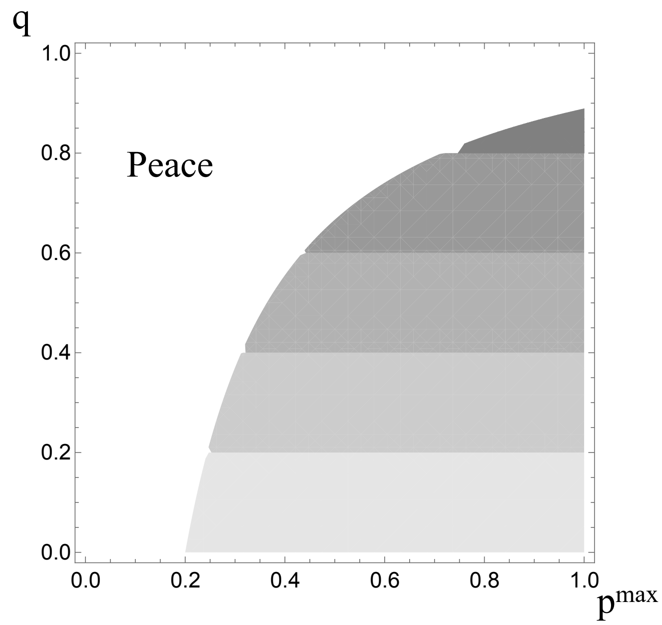
Increasing challenger strength changes two terms in this expression: $\bar{p}$ and the F function. Suppose we define an increase in challenger strength as a uniform upward shift in the probability of winning a conflict, such that this probability is $p_t + d$ for some constant $d > 0$. In this formulation, p_t is the "baseline" probability of winning, which still follows distribution F, and d is the increase in this baseline threat. 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 $Pr\Big(p_t + d > \frac{1-\delta}{1-\mu} + \delta(\bar{p} + d)\Big) = 1 - F\left(\frac{1-\delta}{1-\mu} + \delta\bar{p} - (1-\delta)d\right)$. This term strictly increases in d. Therefore, conditional on conflict occurring along the equilibrium path, a stronger challenger decreases the expected number of periods until conflict occurs. A uniform upward 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, as discussed following the statement of Proposition 6.

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. A high value of q guarantees 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 conflictual 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 in fact make conflict *more imminent*, whereas a large increase in q eliminates conflict entirely.

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



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

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, $p_t \in \{p^{\min}, p^{\max}\}$, with $q = Pr(p_t = p^{\max})$. 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.

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

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

If $\underline{x} \leq x^*(p^{\text{min}}, -\infty)$, then the lower bound never binds and the analysis is equivalent to the unbounded case. At the other extreme, if $\underline{x} > p^{\text{max}}(1-\mu)$, 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:

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

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

Given the upper bound of 1 for an offer, a peaceful MPE requires $x^*(p^{\text{max}}, \underline{x}) \leq 1$. 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 $x^*(p^{\text{max}}, \underline{x}) \leq 1$ if and only if:

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

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. This follows directly from the costliness of fighting.

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 in our baseline model. 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:

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

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:

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

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

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:

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

Solving for V_n^C gives:

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

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:

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

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, $(1-\mu)p_t$. 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:

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

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 μ_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 μ_t is iid and follows a distribution $G(\mu)$, with minimum value μ^{min}, maximum value μ^{max}, and average value $\bar{\mu}$.

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

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

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 μ rather than p.

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

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

Consequently, the optimal offer is:

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

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:

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

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 μ^{min} and $\bar{\mu}$ 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 strategic 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 $\underline{x}_t \geq -\infty$. 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 minimum-threat period. Doing so would deliver (weakly) more transfers to the challenger in a period in which they can already be induced to accept but 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 $\underline{x}_t$ 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 $\underline{x}_t$ 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 maximum-threat periods and simply delivers weakly more spoils to the challenger in minimum-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:

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

Alternatively, upon choosing $\underline{x}_t \geq \underline{x}^{\text{peace}}$ but not subsequently choosing $\underline{x}_z > \underline{x}_t$ in any period $z > t$, the ruler's lifetime expected utility is:

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

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

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

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

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

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

$$\frac{1-p^{\text{max}}(1-\mu)}{1-\delta} > \frac{(1-p^{\text{max}})(1-\mu)}{1-\delta} \implies \mu > 0.$$

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 exact amount of institutional reform. There are a continuum of equilibrium choices in which the ruler chooses between a bit more institutional reform (yielding less consumption for herself in a minimum threat period) and offering somewhat fewer temporary transfers in a maximum threat period, 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 Castañeda Dower et al. (2018) and Powell (2021).

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.