Futile or farsighted? Domestic politics, leader replacement, and the nature of lopsided conflicts

Forthcoming in *Defence and Peace Economics*

Michael Becker*

July 2021

Abstract

In an important subset of armed conflicts, there is a large disparity between the belligerents' capabilities. Although the outcomes of such wars cannot be predicted with absolute certainty, the chances that a weak state will triumph are often so low that private information is not a satisfying explanation for conflict. Under which circumstances, then, does the disadvantaged side give in rather than attempting to fight? And what explains variation in effort levels by strong states? I present a model of an environment in which a strong state faces political pressure to replace the losing government following a costly war. I show that there is a non-monotonic relationship between the effort the winning state devotes to the war and the probability that the losing state surrenders—when war effort is low, the weaker state fights because it has a decent chance of winning; when it is high, the weak state fights because it recognizes that it will be replaced upon surrender. Further results are derived to explain the domestic political foundations of variation in strong states' investments in lopsided conflicts. The results help rationalize a number of historical instances of lopsided conflict.

*Ph.D. Candidate, Department of Politics, New York University. Email: mb4959@nyu.edu. I thank Bruce Bueno de Mesquita, Eric Dickson, NYU seminar participants, and two anonymous reviewers for helpful comments. **Word count**: 7213

In an important subset of armed conflicts, fighting occurs despite the fact that there are wide gaps between belligerents' capabilities. Consider one case by way of illustration. In 1939, the Soviet Union demanded minor territorial concessions from Finland, which refused, leading to the Winter War. The Soviet forces outnumbered the Finns by hundreds of thousands, and also included thousands of tanks and planes; the Finnish military had only a few dozen tanks and aircraft, and soldiers had limited ammunition, fuel, and weapons. The Finnish military, benefitting from the unusually cold winter, mountainous terrain, and problematic Soviet tactics (none of which, except the terrain, could have been foreseen ex ante) did manage to hold off the Soviet advance for several months. But they were eventually overwhelmed and forced to cede territory that was not only more than the Soviets had initially demanded (and, moreover, not in exchange for a much larger amount of other land, as the Soviets had previously offered), but of significant economic and strategic value (Tillotson 1993; Trotter 2002).

Given that they were so outmatched and that Soviet demands were relatively minor, why did the Finns not simply concede at the outset? Although there is debate among historians, there is some evidence that the Finnish government inferred from the massive Soviet mobilization that the USSR's real war aims entailed more than mere territorial acquisition, including a full takeover of the country and installation of a puppet regime like those the Soviet Union subsequently set up in the Baltic States (Clemmesen and Faulkner 2013).[1] By focusing on the interaction between investments in war by strong powers and their domestic audiences, this paper provides a logical foundation for this inference.

In essence, since massive mobilizations are costly in terms of both resources and human lives, leaders typically choose them only when significant foreign policy goals—

[1]Although the Soviets did not, ultimately, depose the Finnish government, Stalin had intended originally to do so before deciding to limit his aims and divert resources to the Baltic States. Still, Soviet extractions were extreme, including much of the industrialized areas of Finland, its second-largest city, as well as thousands of cars, trucks, and ships (Clemmesen and Faulkner 2013).

regime change, or territory or resource appropriation—can be achieved. Domestic audiences may punish executives who raise taxes and send soldiers into battle for limited achievements. In autocracies, the relevant audience may be a small group of elites while leaders in democracies target important segments of the electorate. Leaders of weak states, then, may actually be *more* likely to fight when faced with a large mobilization, inferring that the stakes are high enough to merit fighting even when the chances of winning are low. In other words, weak states can see that strong states cannot afford politically to come away from a high-intensity conflict with little to show for it. At issue, then, is a particular species of commitment problem. My claim is not that this explanation accounts for all instances of lopsided conflict; weak states may also choose to fight because they think (rightly or wrongly) they can wear down even a militarily superior adversary, as the North Vietnamese did against the United States and the Afghans did against the Soviet Union. I merely argue that, because they are politically costly, large investments in war can be informative signals of the extent of a state's aims. In general, one key aim of this paper is to demonstrate that, even when states' capabilities are obviously different and when this is common knowledge, conflict can still be rationalized from the perspective of a weak state. Of course, in some conflicts, mismatches in disputants' capabilities are not foreseeable ex ante, but emerge endogenously over the course of the war. In such cases, the model presented here is better suited to explaining the prolongation of conflicts after military capacities have diverged, rather than their initiation.

In addition to explaining why especially *weak* states sometimes enter conflicts they are likely to lose, in this paper I also offer a rationalization for the wide variance in the effort *strong* states apply to lopsided conflicts. To do so, I provide a formal model that builds on existing work on commitment problems and the effects of domestic politics on international relations. In the model, leaders from one strong and one weak

state embroiled in a conflict make decisions that jointly determine not just the conflict outcome, but also their own career prospects. The strong state decides on a level of investment to devote to the war, and whether to replace the weak state's government in the event of victory. For its part, the weak state decides whether to surrender or continue the conflict after observing the strong state's effort.

The findings are consistent with several counterintuitive historical patterns. First, on the assumption that strong states face political pressure to replace the governments of their adversaries, and that this pressure is increasing in the cost and severity of the war, I show that the relationship between strong states' effort and the probability that the weak state continues the conflict is non-monotonic. This is because, if the strong state devotes too much effort to the war (i.e., above a specific cut-point), the weak state chooses not to surrender, recognizing that the gamble of war—even though defeat is likely—is superior to surrendering, because the strong state is certain to depose them. (This idea is related to the "gambling for resurrection" finding of Downs and Rocke (1994). In the next section I discuss how, although our accounts are potentially complementary, the mechanisms driving them are distinct.) On the other hand, if the strong state devotes too little effort to the war, the weak state will also fight, because, despite being disadvantaged, it has a chance to actually win. Finally, I show that, for most parameter values, there is an intermediate range between these two conflict regions in which the weak state surrenders.

I next derive equilibrium effort levels for the strong state, highlighting in particular the influence of two parameters: the weight that the winning state's representative voter places on the cost of the war (as opposed to winning); and the share of the war's cost that can be defrayed by expropriating the losing state's resources following the war. Overall, the contribution of this paper is to add to the literature on rationalist explanations for war by elucidating the logic driving a type of conflict that, given the

disparity in actors' capabilities, seems particularly perverse.

The remainder of the paper proceeds in the following manner. The next section situates the paper in the existing literature on conflict in international relations and the interaction between domestic politics and international conflict. I next present and analyze the model while discussing consistencies between the results and historical events. (I do not consider the discussed cases in any way a "test" of the model's implications, but rather as an indication of their plausibility.) The final section briefly summarizes and discusses planned extensions.

1 Related literature

The fact that ex-post inefficient war occurs at all was long a puzzle in the international relations literature (Fearon 1995). One possible explanation advanced by Fearon, among others, was that states fight because of private information about their relative strength or resolve. However, as Powell (2006) remarks, this explanation fails to account for the existence of long wars, when estimates of relative power should converge. It is similarly unsatisfying as an explanation for conflict between belligerents whose capabilities are manifestly disparate. Other work has shown how future expected shifts in capabilities can lead to the disappearance of peaceful bargains to which all sides can commit (Fearon 2004; Powell 2004). In this paper, by bringing in leaders' domestic political pressures, I extend the findings on the pernicious effect of commitment problems to conflicts between states whose capabilities are both far apart and unlikely to shift.

The model I construct also has implications for several other strands of the literature on conflict and international relations. Most obviously, this paper aims to contribute generally to the research program begun by Fearon by enriching the set of rationalist explanations for conflict—in this case, for the persistence rather than the onset of

conflict, and in a specific subset of cases. Recently, for example, formal-theoretic work has argued that, even in Fearon's bargaining model, conflict can persist when the expected duration of war is in a medium range and there is a disparity between the status quo and the expected post-war settlement: leaders "stall" in order to enjoy the short-term distribution of resources (Spaniel, Bils, and Judd 2018). Relatedly, Langlois and Langlois (2006) study "probabilistic war equilibria" in which, even when belligerents' capabilities diverge, for defender states, the costs of fighting can be mitigated by the utility from retaining the contested territory or resource while the conflict persists.

Second, there is an obvious connection to research on civil wars, which are often characterized by significant disparities in strength between governments and rebel groups and wide variation in war duration (Blattman and Miguel 2010; Fearon 2004; Sambanis 2004). Third, though I do not explicitly motivate the model this way, the costs associated with surrender by either actor, but in particular the advantaged one, are consistent with "audience costs" or other related domestic political explanations for conflict (e.g., Fearon 1994; Schultz 2001; Levendusky and Horowitz 2012).

Fourth, the model generates results that are interesting in light of existing information-based theories of conflict. Wagner (2000, p.477), formalizing an intuition from Clausewitz (1982), argues that weak states sometimes fight strong states "not to contribute to the enemy's defeat, but to influence instead his expectations of the future course of the war." Smith and Stam (2004) relax the "common priors" assumption, arguing that conflict results from disparities in adversaries' fundamental beliefs about the nature of the world, rather than disagreements over their relative military strength or private information. While the model is not, by any means, a substitute for informational accounts of conflict persistence, I provide a mechanism to explain wars between weak and strong states in a full-information environment.

But though this idea is inspired by the war inefficiency literature, some of which is

cited above, I abstract away from the bargaining between states that drives many of the existing models; the focus here is on the political economy of the strong state, and how this interacts with the weaker state's leader's desire to remain in office. Though the bargaining perspective has generated interesting theoretical predictions about the conditions that foster war, it remains characterized by logical gaps and a lack of empirical corroboration, including a problematic fit with historical narratives (Powell 2006; Garzke and Poast 2017).

This paper also draws on research examining the effect of domestic variables on war dynamics. So-called "second-image" explanations for war were long out of favor as scholars chose to treat war as a bargaining failure between unitary actors (Fearon 1995) or a logical consequence of international "anarchy" (Waltz 1959). More recently, however, both empirical and theoretical research has enhanced our understanding of the onset and conduct of war by relaxing the unitary actor assumption (e.g., Bueno de Mesquita et al. 1999). There is a two-way connection between leaders' foreign policy decisions and their domestic fortunes; for example, leaders who preside over war losses fare tend to fare more poorly domestically (e.g., Bueno de Mesquita, Siverson, and Woller 1992; Bueno de Mesquita and Siverson 1995; Goemans 2000). Bueno de Mesquita and co-authors (2003) also devote a chapter in their book to examining how domestic structures affect post-war outcomes, meaning the decision by the winning state of whether to install a "puppet" government or take land from the loser. The authors argue that large-coalition systems—democracies—usually prefer obtaining public goods in the form of policy gains by installing puppets, while autocracies prefer to take land and distribute the resulting private gains to their smaller coalitions. As described in the next section, in the model I present, the drive to replace a defeated government is consistent with either of these types of preferences.

Finally, it is worth noting the similarity between the explanation for conflict I ad-

vance in this paper and the "gambling for resurrection" argument of Downs and Rocke (1994) (See also Goemans 2008; Croco 2011.) In essence, this argument has it that leaders weakened in the eyes of *their own* domestic audiences prolong wars because such leaders are more likely to be deposed by their own people following defeat. Traditionally, this outcome is due to the informational pathologies of the principal-agent relationship (the principal, in this case, being the electorate that chooses the executive). Having more limited information than the executive, the electorate is unsure about whether an observed war was, ex ante, a good idea; likewise, if the executive refrains from starting a war, the electorate cannot tell whether this was, from their point of view, an optimal decision. To limit moral hazard (i.e., deter the executive from starting "bad" wars), the electorate may condition its reelection decision on the outcome of the war. This phenomenon sometimes induces the executive to continue fighting wars revealed (after initiation) to be bad ideas since "the alternative of escalation is attractive because, although the expected value of the outcome may be lower than that of peace, the greater variance of conflict holds out a better possibility of rising above the threshold for staying in office" (Downs and Rocke 1994: 375).

While the implication of their model—the occasional continuation of seemingly futile conflict—is similar to one of my main findings, the mechanism driving it is distinct. First, their findings derive from the information and preference structure, in which not only is there doubt ex ante about the likely course of a conflict, but there are potential gaps in preferences between the executive and electorate. My model, by contrast, produces this phenomenon when there is no ambiguity about the probability of victory or gap between the preferences of the actors on the losing side (indeed, I assume that the weak state is unitary). The desire to fight a losing battle, instead, emerges from a fear of replacement by a *foreign* power. In the next section, I proceed to the model setup.

2 Model

2.1 Setup

There are two actors, a winner W and a (potential) loser L, who are involved in an ongoing violent conflict. The sequence of actions is as follows (I also provide the extensive form in Figure 1):

1. W chooses a level of costly effort $x \in \mathbb{R}_+$ to devote to the war.
2. L observes x and decides whether to continue the conflict or surrender.
3. If L surrenders, W decides whether to replace the government of L. Either way, the game ends here and payoffs are realized.
4. If L chooses to fight, with probability $f(x)$ L manages to defeat W and the game ends. With complementary probability, L is decisively defeated (that is, defeated after attempting to fight).
5. If L is decisively defeated, W decides whether to replace the government of L. Either way, the game ends.

2.1.1 Preferences of the winning state

In keeping with the domestic politics focus of this paper, the leader of the winning state is primarily motivated by retaining office; his value for remaining in power is β. To do so, he must manage the tradeoff inherent in increasing the effort devoted to the war, which makes a decisive victory more likely but is also politically costly. The outcome of the war (if it happens) is determined by the winning state's investment, x, and the probability that W loses is given by the function $f : \mathbb{R}_+ \to [0, 1]$.

To make the idea of political cost more explicit, although voters are not strategic actors in this model, I assume the politician acts as if there is an electorate that votes retrospectively based on the outcome of the conflict and the amount of resources devoted to it. More specifically, suppose there is a representative voter whose preferences are a convex combination of two components. One share, $\lambda \in (0, 1)$, concerns the costliness of the war (x). As this share grows, the voter reacts more strongly to investments in war, x, and the probability of reelection declines. This might be thought of as the population's aversion to the material costs of war—higher taxes, battle deaths, etc. A complementary share of the voter's preferences $(1 - \lambda)$ relate to the conflict's outcome, reacting positively to W if L surrenders or loses decisively (that is, loses after attempting to fight), and replacing W if the latter is defeated.[2] This share of the representative voter's preferences reflects the fact that, other things equal, voters reward politicians who preside over war victories, and punish those who preside over defeats (Bueno de Mesquita, Siverson, and Woller 1992; Bueno de Mesquita and Siverson 1995; Bueno de Mesquita et al. 2003). To take two examples from American politics, President George H.W. Bush, who was in office during the Gulf War, which ended in a decisive victory for the U.S., had approval ratings near 90% as the conflict ended (though he was not ultimately reelected). On the other hand, the poor progression of the Vietnam War played an important role in making President Lyndon B. Johnson unpopular to the point where he did not seek reelection.

I also assume that to the extent that the voter is concerned about the costs of war (λ), he can be placated somewhat if W removes L from power. This could have multiple interpretations, but the one I have in mind is that the winning state may

[2]These parameters have multiple potential substantive interpretations. They might simply be thought of as the "weights" placed on different (though related) policy dimensions by a representative voter. Or they could be the relative shares of the poor and wealthy in a democratic electorate. It is the poor who overwhelmingly bear the physical and material costs of victory (λ) while the rich, perhaps, have more to gain economically and personally from victory.

be motivated by the opportunity to expropriate private resources from the defeated.[3] In this case, the parameter $\delta \in (0, 1)$ measures how much of the war's cost can be defrayed by pillaging L's resources after the war. As δ approaches zero, the cost of war is increasingly offset by the taking of L's resources, reflecting a victorious state that has obtained a politically valuable concession—for example, broad policy changes or significant reparations. If δ is close to one, there is less to gain from deposing the government and expropriating L's assets. As detailed in the game tree, since this parameter partially determines the financial burden of the war on the population W leads, it also affects the latter's reelection prospects.[4] It is important to note, however, that this type of *political* cost defrayal is only possible when the winning state deposes the weak state's leader, meaning that the model speaks better to cases in which winning states' electorates can only be appeased by significant foreign policy achievements such as radical, broad policy reversals (e.g., a dramatic economic liberalization), rather than incremental changes like minor territorial cessions. My assumption is that the former is more likely to occur, and to be durable, under new leadership beholden to a foreign actor for their office. Certainly, it is possible to imagine more modest types of extractions that would not require regime change (e.g., slight reductions in tariffs) but it is hard to envision such concessions exerting significant electoral influence in winning states. This

[3] As Bueno de Mesquita et al. (1999) demonstrate, the tendency for victorious states to appropriate resources is more characteristic of autocracies. Thus this interpretation may be more applicable to the foreign policies of small-coalition states.

[4] An alternative justification of the δ assumption is also offered in Bueno de Mesquita et al. (1999). The authors contend that leaders maintain power by successfully competing to construct winning coalitions, either by buying off elites (in autocracies) or voters (in democracies). In the authors' argument, following wars, large-coalition systems—i.e., democracies—are less motivated to seize private goods like territory, since the resulting gains are not large enough, once divided among the key coalition members (voters), for each to receive a substantial share. Instead, democracies tend to install puppet regimes to obtain policy gains like favorable trade terms. The reason, they argue, is that policy concessions more closely resemble public goods from which large numbers of voters can all benefit. Then $(1 - \delta)$ might represent the extent to which these policy goals were achieved. And since, in my model, a total victory is required to replace the leader of L (entailing an additional occupation/regime change cost $k > 0$), this interpretation is more applicable to democratic winning states (particularly as $\lambda \to 0$, meaning the "public good"—that is, winning the conflict—is more prominent in the winning electorate's preferences).

assumption is reflected as well in the k parameter, which is a constant, rather than a function of the nature of extraction.

Finally, if the conflict continues (i.e., if L does not surrender), W pays a cost c regardless of whether he retains office; this can be thought of as the personal, political, or reputational cost of having been a leader who presided over a war. Note that, following Filson and Werner (2007), I formalize a distinction between the costs of fighting and the costs of losing the war.

To give an example of W's payoffs (see below for the full game tree), suppose L surrenders. Then, upon leaving L in office, W obtains the following expected utility:

$$\beta[\lambda e^{-x} + (1-\lambda)(1)].$$

Here β is the value of winning reelection, and the terms in brackets capture the probability of winning: the share $(1-\lambda)$ of the population is sufficiently happy with the victorious outcome to support reelection, while the support of the share λ is a decreasing function of war effort, x. The overall probability of being reelected, then, is a convex combination of these components.

Alternatively, suppose that following surrender W elects to replace L. Then W's expected utility is:

$$\beta[\lambda e^{-(\delta x+k)} + (1-\lambda)(1)],$$

where the difference here is that W must pay an additional cost, $k > 0$, associated with regime change, but may also defray a fraction $(1-\delta) \in (0,1)$ of the war's cost through expropriation.[5]

[5]That is, δ is the share they are still politically punished for.

2.1.2 Preferences of the losing state

The preferences of the losing state are simple. L prefers to remain in power and values that status at β. In a situation in which the loser is replaced, he receives ϵ, which represents any of the potential negative outcomes a deposed leader might confront, including imprisonment, exile, or death. As is the case for W, L also pays a cost for continuing the war (c), but only if he remains in power. (To avoid uninteresting cases, I assume that the leader always values remaining in office over ending the war and being replaced, i.e., $\beta - c > \epsilon$.) Finally, there is also a parameter α, which represents the benefits or costs the loser attains in the aftermath of the conflict's conclusion.[6] These could reflect a possible strengthening or weakening of the leader's position in his ruling coalition following the war. If L fights and manages to win, it gains α. If it surrenders or is defeated, it loses α. I do not micro-found these costs, placing the focus instead on the domestic politics of the winning state. Further, I abstract away from the "effort" that L devotes to the war. This makes the model more tractable, but can also be justified on substantive grounds: in the cases the model is designed to speak to, the weaker side tends to have to "max out" its capabilities simply to have a chance of surviving; variation in the nature of the conflict tends to be driven more by how much effort the stronger side invests. Before proceeding to the analysis, below I provide a summary of the game's notation and a game tree.

[6]These are in a sense related to "audience costs," but with a meaning distinct from typical usage. In International Relations parlance, audience costs usually refer to reputational damage that leaders face from not following through on threats, or backing down during a crisis. The empirical support for such effects, however, is mixed. Schultz (2012), for example, refers to them as "dark matter" in that they are difficult to observe but whose existence it is convenient to assume. However, I mean the α parameter to refer to the political benefits or costs associated not with threats and whether they were made good, but with a war's outcome, which is both more easily observed and more materially consequential. The empirical record suggests that voters are quite responsive to war outcomes (Bueno de Mesquita, Siverson, and Woller 1992; Bueno de Mesquita and Siverson 1995; Bueno de Mesquita et al. 2003).

Summary of Notation		
Notation	Meaning	Range
e	Euler's number	
$f(.)$	Probability, conditional on conflict, that L defeats W. Functional form: $f(x) = e^{-x}$.	(0, 1]
β	Actors' value of holding office	$(0, \infty)$
λ	Share of the representative voter's preferences (in W) concerned by the war's cost (x)	(0, 1)
$(1 - \delta)$	Share of the war's cost that the winning state can recoup by expropriating L's resources; alternatively, public goods flowing from imposing regime change as a share of the war's cost	(0, 1)
x	Winning state's investment in the war	$[0, \infty)$
k	Cost borne by the winning state if it replaces the losing government, e.g., cost of effecting regime change	$(0, \infty)$
c	Opportunity cost of continuing the war	$(0, \infty)$
α	Domestic costs/benefits of the war's outcome for the loser	$(0, \infty)$
ϵ	Payoff to the loser if replaced	$(0, \beta)$

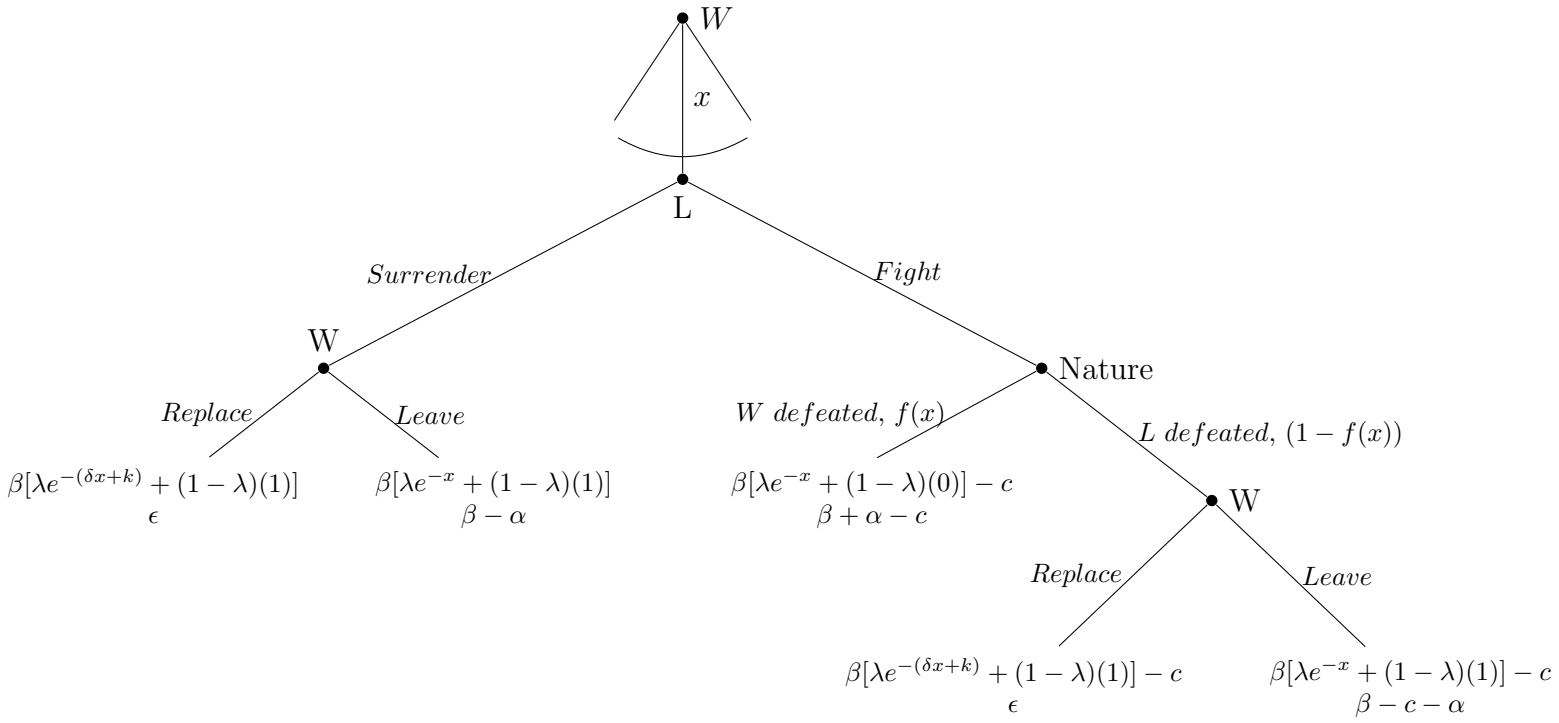

Figure 1: Extensive form (W payoffs listed first)

The environment is one of complete and perfect information, so the appropriate equilibrium concept is Subgame Perfect Nash Equilibrium (hereafter "equilibrium"), which requires players' strategies to be sequentially rational. I proceed by backwards induction.

2.2 Analysis

Begin with the winning state's decision of whether or not to replace the losing state's leader (in the subgames in which this is an option).[7] W replaces if

$$\beta[\lambda e^{-(\delta x+k)} + (1-\lambda)(1)] \geq \beta[\lambda e^{-x} + (1-\lambda)(1)]. \tag{1}$$

Simplifying the above inequality leads to the first result:

Proposition 1. *In equilibrium, W replaces L if $x^* \geq \frac{k}{1-\delta} \equiv \hat{x}$, where x^* denotes W's equilibrium choice of effort.*

Proof. Follows from the discussion in text.

This cutpoint, intuitively, is increasing in the costs of replacement (k). It also falls as the share of the war's costs that can be recouped by replacing L ($1-\delta$) increases. That is, if the political or economic benefit associated with replacing L is low, then W will not do so unless the endogenous costs of war are quite high relative to those associated with occupation.

Since L can anticipate W's replacement decision, in equilibrium, it must make a choice on both sides of the cut point, $\hat{x}$. If $x^* > \hat{x}$, i.e., if W will replace L, then L

[7] W's utility function is not exactly the same in the two subgames in which this decision must be made. If L surrenders, W does not have to consider the direct costs of war (c) when making his decision. However, since in the subgame after L chooses to fight, this cost must be paid no matter what, it plays no role in the decision-making process.

chooses to continue the fight if

$$(1 - f(x^*))(\epsilon) + f(x^*)(\beta + \alpha - c) \geq \epsilon.$$

For tractability, I assume the following functional form for $f(.)$, the probability, conditional on conflict, that the underdog, L, wins: $f(x) = e^{-x}$. This form satisfies $\frac{\partial f(.)}{\partial x} < 0$ and $\frac{\partial^2 f(.)}{\partial x^2} > 0$. Assuming L values office sufficiently highly (β), and his cost of war is sufficiently low (c), this condition is trivially satisfied for any value of $f(.)$, meaning that if W chooses a high enough value of x, L will fight no matter how low the chances are of winning. Here L fights because, even though the chances of winning are low (because x is so high), W cannot commit to leaving L in office.

If, on the other hand, $x^* < \hat{x}$, i.e., if W will leave L in office, then L chooses to fight if

$$(1 - f(x^*))(\beta - c - \alpha) + f(x^*)(\beta - c + \alpha) \geq \beta - \alpha. \tag{2}$$

Using the functional form of $f(.)$ specified above, this inequality simplifies to

$$x^* \leq \ln\left(\frac{2\alpha}{c}\right) \equiv \tilde{x}.$$

For most parameter values that satisfy intuition, it will be the case that $\tilde{x} < \hat{x}$. This leads to the second result:

Proposition 2. *If $\tilde{x} < \hat{x}$, then the relationship between war effort and the equilibrium probability that L surrenders is non-monotonic, as depicted in Figure 2.*

Proof. Follows from the discussion in text.

This result is somewhat counterintuitive, since it says that the likelihood of L sur-

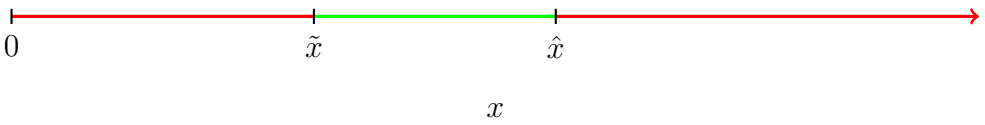

Figure 2: L's equilibrium surrender-or-fight decision as a function of W's war effort (green region denotes surrender)

rendering is lower above a certain threshold of (x)—when the probability of being decisively defeated is higher. The logic driving this result is the following: at low levels of x, L fights because there is a decent chance of achieving victory; at high levels, L fights because, even though the chances of getting to a stalemate are vanishingly small, this gamble is worth it since, given the high cost of the war, W will be under more pressure to depose L, giving the latter the deposition payoff, ϵ. This result is consistent with a number of historical cases in which outmatched states continued warring long after it was clear they would likely lose: Nazi Germany after 1943, for example. This may also help explain the persistence of many civil wars in which a rebel group is too weak to contest a state directly, but instead of bargaining, prefers to maintain a weak hold on a subset of the state's territory, knowing that the state's investment in the war makes it politically unable to commit to following through on a peace deal. This was arguably the case in Sri Lanka and Colombia, for example. Thus there is a particular type of commitment problem standing in the way of peace, grounded in the assumption that winning domestic audiences can benefit from the loser's regime change following costly wars. On the other hand, by committing an intermediate amount of resources to the conflict—but enough that winning decisively is sufficiently likely—the winning state can signal that the war is not costly enough to warrant replacing L, inducing the latter to back down and avoid prolonging the war.

Finally, consider W's equilibrium choice of x. There are three possible cases: choose a value of $x \leq \ln\left(\frac{2\alpha}{c}\right)$—inducing L to fight—and leave L in power if W wins decisively; choose a value of $x \geq \frac{k}{1-\delta}$—in which case L will also fight—and replace L in the case

of victory; or choose an $x \in [\ln(\frac{2\alpha}{c}), \frac{k}{1-\delta}]$—which leads L to surrender—and leave L in power. W's maximization problem, then, is the following:

$$\max \begin{cases} \max_{x \geq \frac{k}{1-\delta}} \begin{array}{r} (1-f(x))\Big[\beta[\lambda e^{-(\delta x+k)} + (1-\lambda)(1)] - c\Big] \\ +f(x)\Big[\beta[\lambda e^{-x} + (1-\lambda)(0)] - c\Big] \end{array} \\ \max_{x \in [\ln(\frac{2\alpha}{c}), \frac{k}{1-\delta}]} \beta[\lambda e^{-x} + (1-\lambda)(1)] \\ \max_{x \leq \ln(\frac{2\alpha}{c})} \begin{array}{r} (1-f(x))\Big[\beta[\lambda e^{-x} + (1-\lambda)(1)] - c\Big] \\ +f(x)\Big[\beta[\lambda e^{-x} + (1-\lambda)(0)] - c\Big] \end{array} \end{cases}. \tag{3}$$

Taking the effort levels that solve the above sub-maximization problems, subject to the appropriate constraints on x, and substituting these levels back into the utility functions allows us to characterize *equilibrium* effort levels for different parameter values. I provide the details of these derivations in the appendix, and summarize the key results here and in Figure 3.

Proposition 3 (effect of high costs of war). For c sufficiently high, W's best response is to choose an intermediate effort level just high enough to induce L to surrender; that is $x = \ln(\frac{2\alpha}{c})$.[8]

Proof. See appendix.

[8]Note that this requires assuming that, in equilibrium, L is willing to play an undominated strategy (surrendering, in this case), since when $x = \ln(\frac{2\alpha}{c})$, L is actually indifferent between fighting and surrendering. Assuming otherwise would create a problem for W wherein at high levels of c it would like to choose the smallest number greater than $\ln(\frac{2\alpha}{c})$, which does not exist.

The intuition here is simple: as c increases, the opportunity costs of war make conflict less profitable—for both players. The best that W can do is spend as little as possible while still inducing L to surrender. The interesting part of this result is that, even as the costs of war grow prohibitively high, W cannot simply set a very low x and avoid further losses. To do so would signal (correctly) to L that it might win and derive the domestic benefit, α, in which case W also incurs a political cost that is decreasing in λ. The phenomenon of devoting just enough to the war to induce surrender, but not so much as to telegraph a planned regime change, is perhaps characteristic of the first Gulf War and the 2008 war between Russia and Georgia. In both cases, the stronger state's leaders needed, politically, to force their rivals to back down, but anticipated prohibitively high costs from a prolonged conflict. So they devoted enough investment to force a truce, and opted against replacing their adversaries' leaders.

Finally, Proposition 4 describes the effort tradeoff shown in Figure 3.

Proposition 4 (effect of λ and δ on equilibrium effort levels). W's optimal choice of effort, x, is a function of λ and δ, as depicted in Figure 3.

Proof. See appendix.

Specifically, in Figure 3, I show whether high, medium, or low effort is optimal for W for different values of λ and δ. The figure has several key takeaways. First, toward the right of the graph (as $\delta \to 1$), effort levels are trivially driven by the share of the public concerned by the war's cost; as this share increases, effort levels decrease. Second, as the gains from replacement grow (as $\delta \to 0$), medium effort—which induces surrender—becomes less profitable since the costs of war can be defrayed. Notice, too, that the range in which high effort is optimal grows relative to low since it makes victory more likely while a lower δ insulates W more from the cost of the war.

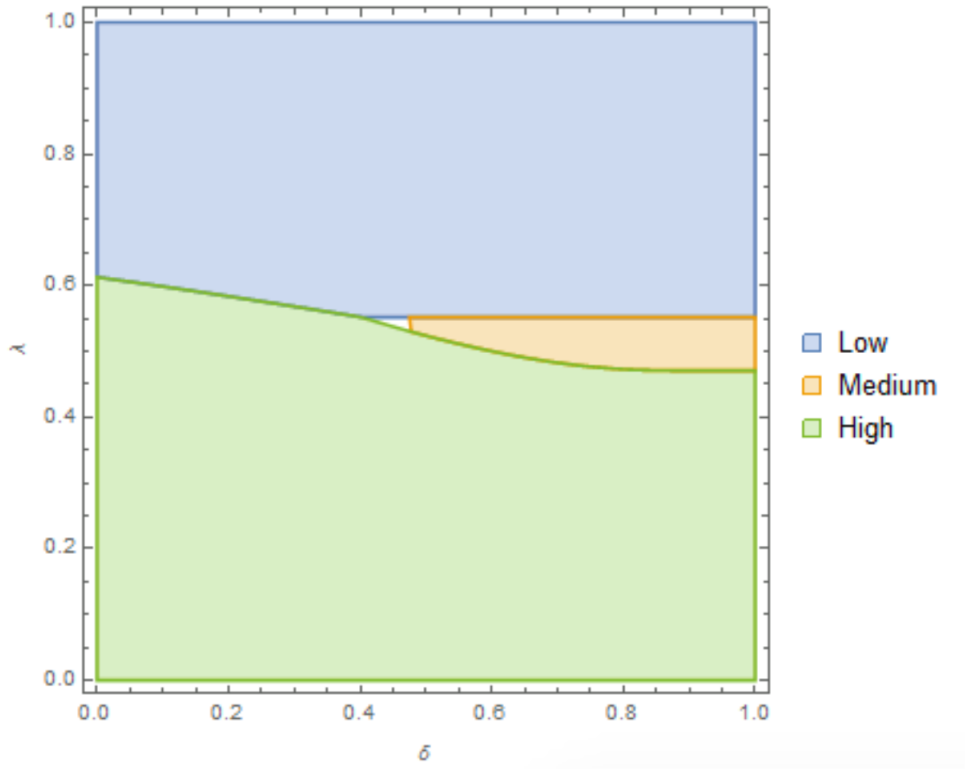

Figure 3: W's equilibrium effort decision as a function of δ and λ. Parameter values: $\beta = 10$, $\alpha = 4$, $k = 1$, $c = 3$.

As an illustration of the public's influence on the executive's choice of military effort in war—and the downstream effect on weak states' surrender-or-fight calculus—consider again the case of the Gulf War. Following the defeat in Vietnam, American politicians had grown sensitive to "Vietnam syndrome"—the American electorate's aversion to protracted ground wars and casualties. Since Vietnam, American presidents had been reluctant to commit ground troops to protracted conflicts. Indeed, after the war, President Bush referred publicly to the syndrome, saying "the ghosts of Vietnam had been laid to rest beneath the sands of the Arabian desert" (Anderson and Ernst 2014). In the model, this state is captured by a value of λ close to one. Historical evidence suggests that the syndrome had a powerful effect on President Bush's decision-making and the American military's strategy (Freedman and Harsh 1991). The intense media attention

on the conflict, theatrical use of airpower, and high nominal numbers of ground troops mask the fact that the American military's effort level was actually rather low relative to its potential.

By "low" effort, in this case, I refer to both a restraint in the use of resources and a choice of military tactics that minimized casualties for the Allies, consistent with the model's prediction about the effect of a high λ. For one thing, American resource commitments were limited. Admittedly, hundreds of thousands of troops were involved, but the costs of war were overwhelmingly paid by other countries. American taxpayers ended up responsible for about 20% of the total costs of war.[9]

Second, the coalition's tactics heavily emphasized airpower over the use of ground troops. In fact, American officials debated whether to use ground troops at all. When possible, they avoided pitched battles with Iraqi units, and ended the war without destroying many units of Saddam Hussein's elite Republican Guard (Freedman and Harsh 1991). In the end, American casualties were fewer than 1,200, and the Iraqis surrendered after only 100 hours of ground engagements.

It would be an exaggeration to suggest that Saddam surrendered because he inferred perfectly from the Allied conduct that they didn't intend to remove him from power or attempt to expropriate Iraqi oil. And Allied decision-making was driven in part by the anticipation that Saddam's ouster would lead to post-war destabilization in the region and high costs of occupation (k). But to the Iraqis, the coalition's restraint was a tangible signal of limited war aims. In contrast, in the 2003-2011 Iraq War the Americans both funded the majority of the war and supplied the bulk of the soldiers, suffering almost 37,000 casualties (though many of these came after the main Iraqi military had been defeated). Further, the conflict entailed an aggressive ground invasion in which the coalition moved to seize the country's territory from Iraqi control.

[9]See the American military's report available here: `https://apps.dtic.mil/dtic/tr/fulltext/u2/a249270.pdf`.

In general, the result shown in Figure 3 illustrates how—at certain fixed values of the other parameters—public opinion and the spoils of war affect executive decision-making regarding war investment. Of course, shifting the other parameters affects the shape and relative size of the three regions. As mentioned, as c grows, the medium effort region starts to crowd out the high and low regions until, at a certain point, it becomes a dominant strategy for any values of the other parameters. As k—the cost of occupation/expropriation from the point of view of the strong state—grows, high effort is less likely to be optimal; while an increase in α (domestic costs/benefits of the war's outcome for the weak state) privileges low and high effort at the expense of medium. In the final section, I briefly recapitulate and discuss possibilities for future research to solidify the model's internal logic and fit with real-world instances of lopsided conflict.

3 Conclusion

In this paper, I provide one explanation for a recurring, and puzzling, empirical pattern: weak states choosing to fight wars that they are very likely to lose. I show theoretically that, since domestic audiences are concerned by the costs of war, weak states can infer something about strong states' intentions from the investments they make. In essence, strong states invest large amounts in wars when they expect to get something out of it: natural resources, territory, or foreign policy concessions. But these are significant goals, attainable only at cost—and sometimes, only by overthrowing the weak state's government. If the strong state invests "too much", that is, enough to attain significant concessions or regime change, the weak state knows it must fight to survive, since the strong state cannot credibly commit to limited extractions. The findings help explain why conflict occurs even in cases where casual intuition would suggest that it should be least likely to occur. On the other hand, while in some cases the lopsided nature

of a war can be anticipated ex ante, in others the mismatch emerges as a *result* of the conflict. The model, admittedly, can speak more to why the latter type of conflict *continues* rather than why it began.

Yet while the findings are consistent with several historical cases, in future work a number of further steps can be taken both to solidify the internal logic of the model and to incorporate additional relevant variables, the better to maximize external validity. In the current setup, for example, it is assumed that leaders in strong states can limit the electoral costs associated with war by despoiling weak states' resources or replacing their leaders. Moreover, it enters into the voter's calculus multiplicatively, as a share of the overall investment in war. While I believe this assumption is substantively defensible, in future work the ability to extract political concessions need not *require* regime change. Such research can also enrich the structure of domestic politics in the strong state by microfounding electoral effects, that is, specifying voters' preferences and incorporating their actions into the model. In one possible instantiation, in which voters could choose between their current leader and a challenger based on their offers of public and private goods (some of which may have been taken from the weak state and distributed to coalition members after the war), there would be a natural opportunity to link this model with selectorate theory, which has been successful generally in explaining the relationship between regime types and leaders' foreign policy (e.g., Bueno de Mesquita et al. 2004). In a related area, in the current model I leave unspecified the *nature* of δ—it may entail literal resources taken from the defeated state; symbolic, domestically important gains such as control of a religious site; or more nebulous foreign policy gains obtained by replacing their leaders. Formalizing these distinctions could lead to interesting findings about how the political economy of the strong state (i.e., voter preferences over intangible vs. extractable goods derived from war) interacts with the nature of the spoils of war to influence leaders' war-fighting decisions. Another

possibility is to merge a model of this type with a signaling game in which a leader must infer the intent of her counterpart, rather than the intent emerging quasi-mechanically from the structure of the game. Perhaps, for example, leaders engage in large mobilizations not always to achieve grand ambitions, but rather to raise the probability of winning substantively minor concessions.[10] Finally, to enhance the model's verisimilitude and generate richer empirical predictions, future research can allow for differential opportunity costs of war, and broaden the weak state's action space by allowing its leader to choose a level of effort or a negotiating stance instead of whether to fight in a binary sense.

But overall, though the theoretical story told in this paper is a simple one, the logic of the main findings—that war investment decisions, because determinative of politicians' career outcomes, are also indicative of their ultimate intentions in war—is consistent with several historical cases as well as some existing international relations theories. The theory offered is an incremental contribution to a more general understanding of international conflict.

[10]I thank an anonymous reviewer for suggesting this possibility.

4 References

1. Anderson, David L., and John Ernst (eds). *The war that never ends: New perspectives on the Vietnam War*. University Press of Kentucky, 2014.

2. Blattman, Christopher, and Edward Miguel. "Civil war." Journal of Economic literature 48, no. 1 (2010): 3-57.

3. Bueno De Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair Smith. "An institutional explanation of the democratic peace." American Political Science Review 93, no. 4 (1999): 791-807.

4. Bueno De Mesquita, Bruce, Randolph M. Siverson, and Gary Woller. "War and the fate of regimes: A comparative analysis." American Political Science Review 86, no. 3 (1992): 638-646.

5. Bueno De Mesquita, Bruce, and Randolph M. Siverson. "War and the survival of political leaders: A comparative study of regime types and political accountability." American Political Science Review 89, no. 4 (1995): 841-855.

6. Bueno De Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair Smith. "Testing novel implications from the selectorate theory of war." World Politics 56, no. 3 (2004): 363-388.

7. Clausewitz, Carl. *On war*. Vol. 20. Penguin UK, 1982.

8. Clemmesen, Michael H., and Faulkner, Marcus (eds). *Northern European Overture to War, 1939-1941: From Memel to Barbarossa*. 2013.

9. Croco, Sarah E. "The decider's dilemma: Leader culpability, war outcomes, and domestic punishment." American Political Science Review 105, no. 3 (2011): 457-477.

10. Downs, George W., and David M. Rocke. "Conflict, agency, and gambling for resurrection: The principal-agent problem goes to war." American Journal of Political Science (1994): 362-380.

11. Eichenberg, Richard C., Richard J. Stoll, and Matthew Lebo. "War president: The approval ratings of George W. Bush." Journal of Conflict Resolution 50, no. 6 (2006): 783-808.

12. Fearon, James D. "Domestic political audiences and the escalation of international disputes." American political science review 88, no. 3 (1994): 577-592.

13. Fearon, James D. "Rationalist explanations for war." International organization 49, no. 3 (1995): 379-414.

14. Fearon, James D. "Why do some civil wars last so much longer than others?" Journal of peace research 41, no. 3 (2004): 275?301.

15. Filson, Darren, and Suzanne Werner. "Sensitivity to costs of fighting versus sensitivity to losing the conflict: Implications for war onset, duration, and outcomes." Journal of Conflict Resolution 51, no. 5 (2007): 691-714.

16. Freedman, Lawrence, and Efraim Karsh. "How Kuwait Was Won: Strategy in the Gulf War." International Security 16, no. 2 (1991): 5-41.

17. Gartzke, Erik, and Paul. Poast. "Empirical Implications of Bargaining Theory: An Agenda for Research." Working paper. (2017).

18. Goemans, Hein E. "Which way out? The manner and consequences of losing office." Journal of Conflict Resolution 52, no. 6 (2008): 771-794.

19. Langlois, Catherine C., and Jean-Pierre P. Langlois. "Does attrition behavior help explain the duration of interstate wars? A game theoretic and empirical

analysis." International Studies Quarterly 53, no. 4 (2009): 1051-1073.

20. Langlois, Jean-Pierre P., and Catherine C. Langlois. "Does the principle of convergence really hold? War, uncertainty and the failure of bargaining." British Journal of Political Science (2012): 511-536.

21. Langlois, Catherine C., and Jean-Pierre P. Langlois. "When fully informed states make good the threat of war: Rational escalation and the failure of bargaining." British Journal of Political Science (2006): 645-669.

22. Levendusky, Matthew S., and Michael C. Horowitz. "When backing down is the right decision: Partisanship, new information, and audience costs." The Journal of Politics 74, no. 2 (2012): 323-338.

23. Powell, Robert. "The inefficient use of power: Costly conflict with complete information." American Political science review 98, no. 2 (2004): 231-241.

24. Powell, Robert. "War as a commitment problem." International organization 60, no. 1 (2006): 169-203.

25. Sambanis, Nicholas. "What is civil war? Conceptual and empirical complexities of an operational definition." Journal of conflict resolution 48, no. 6 (2004): 814-858.

26. Schultz, Kenneth A. "Looking for audience costs." Journal of Conflict Resolution 45, no. 1 (2001): 32-60.

27. Schultz, Kenneth A. "Why we needed audience costs and what we need now." Security Studies 21, no. 3 (2012): 369-375.

28. Smith, Alastair, and Allan C. Stam. "Bargaining and the Nature of War." Journal of Conflict Resolution 48, no. 6 (2004): 783-813.

29. Spaniel, William, Peter Bils, and Gleason Judd. "Stalling." Working paper. (2018).

30. Tillotson, Henry M. *Finland at Peace and War, 1918-1993.* Norwich: Michael Russell, 1993.

31. Trotter, William R. *The winter war: The Russo-Finnish war of 1939-40.* Aurum Press, 2002.

32. Wagner, R. Harrison. "Bargaining and war." American Journal of Political Science (2000): 469-484.

33. Waltz, Kenneth Neal. *Man, the state, and war: A theoretical analysis.* Columbia University Press, 1959 [2001].

5 Appendix

Derivation of equilibrium effort levels:

I first consider the sub-maximization problems in turn. Begin with W's maximization problem at a low level of effort, i.e., one that induces L to fight since, not only will it not be replaced following a loss, but there is also a decent chance of L winning. This problem is

$$\max_{x \leq \ln(\frac{2\alpha}{c})} \begin{array}{c} (1-f(x))\Big[\beta[\lambda e^{-x} + (1-\lambda)(1)] - c\Big] \\ +f(x)\Big[\beta[\lambda e^{-x} + (1-\lambda)(0)] - c\Big] \end{array}. \tag{4}$$

Substituting in the functional form for $f(.)$, the partial derivative with respect to x is

$$\frac{\partial}{\partial x} = -2\lambda\beta e^{-x} + \beta e^{-x},$$

implying that for $\lambda < \frac{1}{2}$, W's optimal choice is to set $x^* = \ln(\frac{2\alpha}{c})$, which (substituting this value back into (4)) yields the following expected utility at the optimal level of (low) effort:

$$EU_W\left(\text{Low effort} \mid \lambda < \frac{1}{2}\right) = 2\beta\lambda\frac{c}{2\alpha} - \beta\frac{c}{2\alpha} + \beta - \beta\lambda - c; \tag{5}$$

whereas for $\lambda > \frac{1}{2}$, W should set $x^* = 0$, receiving

$$EU_W\left(\text{Low effort} \mid \lambda > \frac{1}{2}\right) = \beta\lambda - c. \tag{6}$$

Next, consider W's maximization problem in the "middle" range, i.e, one that induces L to surrender since it is high enough to make L's defeat in battle likely, but not high enough to warrant removal from office:

$$\max_{x \in [\ln(\frac{2\alpha}{c}), \frac{k}{1-\delta}]} \beta[\lambda e^{-x} + (1-\lambda)(1)].$$

The partial derivative with respect to x is

$$\frac{\partial}{\partial x} = -\beta\lambda e^{-x} < 0,$$

so in this range the optimal x^* is $\ln(\frac{2\alpha}{c})$, yielding the following expected utility at the optimal level of medium effort:

$$EU_W\Big(\text{Medium effort}\Big) = \beta\lambda\frac{c}{2\alpha} + 1 - \lambda. \tag{7}$$

Finally, consider W's maximization problem at a high level of effort, i.e., one large enough to induce L to fight since it anticipates being replaced if it surrenders. This problem is

$$\max_{x \geq \frac{k}{1-\delta}} \begin{array}{l} (1-f(x))\Big[\beta[\lambda e^{-(\delta x+k)} + (1-\lambda)(1)] - c\Big] \\ +f(x)\Big[\beta[\lambda e^{-x} + (1-\lambda)(0)] - c\Big] \end{array}.$$

The partial derivative with respect to x is

$$\frac{\partial}{\partial x} = -\delta\beta\lambda e^{-(\delta x+k)} + \delta\beta\lambda e^{-(\delta x+x+k)} + \beta\lambda e^{-(\delta x+x+k)} + \beta e^{-x} - \beta\lambda e^{-x} - 2\beta\lambda e^{-x}.$$

Except at implausibly low[11] levels of λ this expression is negative. So in this case, W is best off setting $x^* = \frac{k}{1-\delta}$ and receiving the following expected utility at the optimal level of high effort:

$$EU_W\Big(\text{High effort}\Big) = \beta - \beta\lambda - c + \beta\lambda e^{-(\delta\frac{k}{1-\delta}+k)} - \beta\lambda e^{-(\delta\frac{k}{1-\delta}+k+\frac{k}{1-\delta})} - \beta e^{\frac{-k}{1-\delta}} + 2\beta\lambda e^{\frac{-k}{1-\delta}}. \tag{8}$$

I form the Figure displaying *equilibrium* effort levels by comparing equations (5),

[11] Substantively, these levels would correspond to an electorate so indifferent to the costs of war that W can nearly ensure victory by spending extreme amounts. This seems empirically implausible, so I rule these cases out by assumption.

(6), (7), and (8) at fixed values of the parameters β, k, c, and α (specified under Figure 3). □

Proof of Proposition 3: Proposition 3 states that, for a sufficiently high c, W is best off setting $x^* = \ln(\frac{2\alpha}{c})$ (on the border between low and medium effort), just enough to induce L to surrender. Consider W's utility in the three effort intervals (at the optimal level of effort within each interval).

$$EU_W\left(\text{Low effort} \mid \lambda < \frac{1}{2}\right) = 2\beta\lambda\frac{c}{2\alpha} - \beta\frac{c}{2\alpha} + \beta - \beta\lambda - c,$$

$$EU_W\left(\text{Low effort} \mid \lambda > \frac{1}{2}\right) = \beta\lambda - c,$$

$$EU_W\left(\text{Medium effort}\right) = \beta\lambda\frac{c}{2\alpha} + 1 - \lambda,$$

and

$$EU_W\left(\text{High effort}\right) = \beta - \beta\lambda - c + \beta\lambda e^{-(\delta\frac{k}{1-\delta}+k)} - \beta\lambda e^{-(\delta\frac{k}{1-\delta}+k+\frac{k}{1-\delta})} - \beta e^{\frac{-k}{1-\delta}} + 2\beta\lambda e^{\frac{-k}{1-\delta}}.$$

Fix the values of all parameters but c. W's utilities at both low and high effort levels are strictly decreasing in c (recall that the first low effort equation only holds when $\lambda < \frac{1}{2}$), while its utility at a medium level of effort is increasing in c. Since there is no bound on the range of c, for any set of values of the other parameters, there must exist a level of c beyond which medium effort is strictly preferred. □