

# The Armed Peace: A Punctuated Equilibrium Theory of War

**Bahar Leventoğlu**

*Department of Political Science, Stony Brook University*

**Branislav L. Slantchev\***

*Department of Political Science, University of California – San Diego*

October 6, 2005

---

**Abstract.** According to a leading rationalist explanation, war can break out when a large rapid shift of power renders unbelievable a rising state's promise to compensate its declining opponent, causing the latter to attack preventively. This mechanism does not provide a complete and coherent explanation of war because it does not specify how inefficient fighting resolves this commitment problem. We present a complete information model of war as a sequence of battles and show that although opportunities for a negotiated settlement arise throughout, the very desirability of peace creates a commitment problem that undermines its likelihood. Because players have incentives to settle as soon as possible, they cannot credibly threaten to fight long enough if an opponent launches a surprise attack. This decreases the expected duration and costs of war and causes mutual deterrence to fail. Fighting's sheer destruction improves the credibility of these threats by decreasing the benefits from continuing the war. Equilibrium fighting may involve escalating costs that exceed the value of the stakes by the time peace is negotiated and that leave both players worse off than when the war began.

---

---

\*Corresponding author (slantchev@ucsd.edu). First draft: February 9, 2004. This work has been generously supported by the National Science Foundation (Grant SES-0518222 to Slantchev and SES-0518185 to Leventoğlu). Presented at the 2005 annual meetings of the Midwest Political Science Association (Chicago), and the American Political Science Association (Washington, DC). We thank Ken Schultz for comments.

To understand why wars begin, we have to know why they end—the termination of war must involve the resolution of its causes (Blainey 1988, x). Imagine a situation in which a powerful state faces a rising opponent. The declining state could wage costly preventive war while it is still strong or it can wait and be forced to accept an unpalatable revision tomorrow. Because the challenger is still weak, it would rather live in peace, and so it must compensate the declining state for foregoing an advantageous war. If the amount of this compensation exceeds the total benefits states can divide today, the challenger must promise to deliver the rest in the future. If the power shift is sufficiently rapid and large, however, the now dominant challenger will renege on that promise. As a result, the declining state cannot believe it today and goes to war.

This commitment problem is one of the leading rationalist explanations of war (Fearon 1995), the underlying mechanism is quite general (Powell 2004*b*), and there are very good substantive reasons to insist on a theory that does not rely on incomplete information to explain war (Powell 2005). Unfortunately, the theory does not specify how inefficient fighting can resolve this commitment problem short of the total obliteration of one of the warring states. Because outcomes of such finality are very rare, the theory does not provide a satisfactory answer to this puzzle. But if fighting does not resolve the commitment problem, it is pointless to bear its costs and risks in this context: in the end, the commitment problem will still exist and both states would have suffered tremendously. Moreover, if the opponents can terminate the war without resolving the commitment problem, then what does it mean to say that the war was caused by this problem in the first place?

We have several goals in this article. First, we explain why it is imperative that our explanations of war are complete and coherent, and then discuss how the commitment problem mechanism fails these requirements and why. This leads us to a model with stylized representation of war that simultaneously addresses our objections to the traditional approach and allows for the commitment problem to arise. Our second goal is to demonstrate using this model that large rapid shifts of power are only necessary but not sufficient to cause war. Instead, peace fails when states cannot credibly threaten to impose large enough costs to deter each other from trying to exploit the advantages of surprise attack. The credibility problem arises from the very desirability of peace; states cannot threaten to prolong fighting more than absolutely necessary. Third, we argue that whereas most existing models are forced to assume that the costs of war are low relative to the value of the stakes, our results suggest that war can quickly become a net loss for all participants and by the time peace is negotiated states can end up paying cumulative costs that far exceed any possible benefit they can obtain from fighting. Finally, our theory may help shed light on why, in the absence of military victory, wars tend to be settled when at least one of the actors is nearing exhaustion, and why negotiating peace may only be possible at specific junctures during the conflict.

## **1 Completeness and Coherence of the Bargaining Model of War**

Building on ideas of von Clausewitz (1832) and Schelling (1966), the recent work on the causes of war sees military force as an instrument that can be strategically applied in pursuit of political objectives. In this view, actors are interested in dividing some flow of benefits over which they have conflicting preferences. They can do so peacefully through negotia-

tions, but they can also attempt to impose a solution by force. Because such an option is always potentially available to the actors, they can use threats to exercise it to influence the expectations of their opponents, and perhaps secure concessions.

Although there has been much analysis of the conditions that make threats useful, a fundamental puzzle that arises from this approach was overlooked until Fearon's (1995) path-breaking analysis. The problem with the use of force is that it destroys resources and therefore reduces the benefits available for division relative to what actors have before fighting begins. Because war is *ex post* inefficient, it should be possible to locate *ex ante* agreements that would leave both actors better off compared to actually fighting the war. The *bargaining model of war* seeks to identify conditions that make such *ex ante* bargains impossible. That is, it seeks to explain why actors would rather fight than accept any settlement that their opponents are willing to offer.<sup>1</sup> There are two canonical rationalist explanations for war: one arises from asymmetric information, and the other from inability to credibly commit to fulfill promises.<sup>2</sup>

One reason bargaining can end in war is that leaders possess private information about their expected payoffs from war and peace, and they have incentives to misrepresent this knowledge to extract bargaining advantage. War can break out when actors bargain in the shadow of power and engage in the risk-return trade off: they run a slightly higher risk of war in return for obtaining slightly more at the bargaining table. When private information exists, actors may press their opponents beyond their tolerance thresholds (Fearon 1995, Powell 1996).

For the other explanation, imagine an environment in which an actor that is weak today will become strong tomorrow. Because the opponent is relatively strong now, its expected payoff from fighting is rather high, so the transfer that the temporarily weak actor must agree to has to be correspondingly large. With resource constraints, this amount may exceed the benefits currently available, and so the weak actor must commit to transfer the remainder over several periods. However, when power shifts to that actor's advantage, its incentives to follow through on that agreement will be undermined, rendering today's commitment incredible. In such a situation, the weak may be unable to prevent its opponent from resorting to force (Fearon 1995, Powell 2004b).<sup>3</sup>

Slantchev (2004a) argues that although these two mechanisms are very appealing, neither provides a satisfactory explanation for war as originally stated. To see why, we need to step back and ask ourselves what we would want a theory of war to be able to account for.

---

<sup>1</sup>See Powell (1999) for a lucid explanation of the inefficiency puzzle. Powell (2002) and Reiter (2003) provide good overviews of the bargaining model of war and its most recent developments. See Wagner (2000) and Slantchev (2003a) for critiques of the assumptions made in the early generations of this theory.

<sup>2</sup>Fearon (1995) also proposes territorial indivisibilities as a separate explanation but is highly skeptical. Powell (2005) demonstrates that these must be understood as a species of the commitment problem. Two other potential explanations do not fit the above dichotomy. If arming is costly, then defending the status quo by deterring the opponent may be costlier than fighting. If peace is too expensive, bargaining can break down in war (Powell 1993). Slantchev (2005a) shows how private information may lead to arming decisions that can cause war in this way. The other explanation focuses on coordination problems: when there are multiple equilibria, actors may fail to coordinate on an efficient one. Slantchev (2003a) analyzes a situation where fear of early settlement drives inefficient behavior.

<sup>3</sup>This logic does not come from formal models. It is essentially analogous to the causal mechanism in power transition theory (Organski 1968, Organski and Kugler 1980).

Obviously, we want it to tell us what can cause the war to begin. Perhaps less obviously, we also want it to tell us what can cause the war to end. If one is willing to assume that wars end with the unconditional surrender of the loser who is defeated militarily and is unable to continue fighting (or in its outright extermination), then the second requirement is moot: wars end when one side cries “Uncle!” and there is nothing more to explain.

As it turns out, however, such an approach is empirically untenable: many wars end in negotiated settlements. For example, out of 104 interstate wars between 1816 and 1991, 67 (or 64%) ended short of total military victory for one side.<sup>4</sup> In other words, many wars end because both sides *agree* to cease hostilities while both could fight on. That is, they strike a bargain where previously they have been unable to do so. The question then immediately becomes: what has changed during the war? Why can these opponents find a mutually-acceptable deal only after costly fighting? Any non-trivial theory of war must be able to explain war termination without military collapse.

Another related requirement for a theory of war is that its account of war termination explains how fighting has resolved its causes (Blainey 1988, x). To see why we must insist on such a requirement, suppose that actors agree to end the war while the original cause has not been resolved. This may happen if both sides are simply too exhausted to continue and peace is really a temporary truce. However, since most wars do not get renewed as soon as the former belligerents recuperate from the last bout, such an explanation is not satisfactory. This leaves us with the one remaining possibility: actors end the war while still able to continue fighting. But if they are now able to locate a mutually-acceptable bargain without resolving the original problem, then this problem cannot be said to have caused the war: after all, opponents do find a peaceful settlement even in its presence! Any compelling theory of war must explain how the termination of war involves the resolution of its causes.

Slantchev (2004a) proposes two additional criteria that a rationalist explanation for war must meet: completeness and coherence. A theory of war is *complete* if it can account both for the war’s outbreak and for its termination. That is, the theory must identify a set of factors that may cause bargaining breakdown, and it must also identify a set of factors that can subsequently open up the bargaining range, thereby enabling the opponents to reach a negotiated settlement and end the fighting. In addition, a theory of war is *coherent* if it can explain how fighting and intrawar diplomacy may resolve the problem that caused the bargaining breakdown in the first place. In other words, a theory of war is complete and coherent if it can identify a cause of war that fighting then resolves.

Both canonical rationalist explanations for war are incomplete as originally stated (which implies that they are also incoherent). Because they treat war as a game-ending move and represent it with a costly lottery over exogenously fixed outcomes, these models do not even deal with war termination, especially not in terms of a negotiated settlement. It turns out, however, that the asymmetric information explanation can be extended into a complete and coherent theory of war. According to the original account, war occurs because actors have no way of revealing credibly their private information without fighting. This suggests

---

<sup>4</sup>Our calculations using Slantchev’s (2004b) data set which separates war outcomes into those where one side is unable to continue the military contest and those where both sides agree to a settlement while still physically able to continue fighting. For more on war outcomes, see Kecskemeti (1958), Iklé (1971), and Pillar (1983).

that for peace to occur, they must be able to learn enough about each other so that this is no longer an issue. Slantchev (2003*b*) provides just such a rationale for war termination: battlefield outcomes and intrawar diplomacy allow actors to acquire information that eventually causes their expectations to converge sufficiently to permit an agreement. By eliminating private information, fighting essentially resolves the original cause of war.<sup>5</sup>

Unfortunately, we have no analogous complete and coherent theory of war under complete information, and especially one that involves the credible commitment problem (CCP). This is a distressing state of affairs for two main reasons: (1) the asymmetric information explanation has serious empirical deficiencies, and (2) the CCP mechanism is very general. That is, the one good theory we have is limited in what it can substantively explain, while the one that promises to account for inefficiency in a wide variety of settings is not satisfactory.

What are some of the substantive objections to the informational theory of war? Surely, uncertainty is pervasive in international relations, and certainly private information plays a huge role in crisis bargaining. It also stands to reason that fighting enables both sides to correct their initial beliefs about each other. The theory has received empirical support too, at least for interstate wars (Slantchev 2004*b*). However, the very pervasiveness of incomplete information is a major weakness of this theory: after all, uncertainty is ubiquitous and shared by almost all crises, and yet only few of these escalate into war. In other words, whereas the theory can show how private information may cause war, it is less well suited to explain why opponents often manage to avoid conflict despite being asymmetrically informed. Second, the information revelation mechanism cannot deal very well with long wars. Whereas the median interstate war has a duration of less than six months, civil wars tend to last years, sometimes decades (Fearon 2004). Although technological innovations may prolong the fighting while military leaders attempt new strategies to take advantage of them, and although organizational idiosyncracies may delay the incorporation of new information (Gartner 1997), to think that it may take years of near constant interaction for opponents to learn enough about each other is surely stretching the theory.<sup>6</sup> Hence, we need an approach that does not rely exclusively on incomplete information.

The CCP theory seems to offer just such an explanation. Powell (2004*b*) shows that the same mechanism is at work in a wide variety of seemingly unrelated models. In all of them, bargaining breaks down under complete information because large rapid shifts of power undermine the credibility of the promises that temporarily weak actors have to make to avoid fighting. In a related article, Powell (2005) further demonstrates that bargaining indivisibilities, first-strike advantages (Fearon 1995) and bargaining over objects that are sources of military power (Fearon 1996) also turn out to be manifestations of the same fundamental mechanism. In other words, almost every complete-information explanation for war involves this commitment problem at its heart (excepting the two discussed in footnote 2).

As Gartzke (1999, pp. 571-72) rightly notes, however, the commitment problem explanation raises a puzzle itself: “States that can rationally choose to fight because of commitment problems cannot rationally choose to terminate contests until the commitment problems are

---

<sup>5</sup>This convergence of expectations during war can be obtained in other settings. See Filson and Werner (2004), Powell (2004*a*), and Smith and Stam (2005).

<sup>6</sup>For a similar point, see Powell (2005) who also provides additional reasons the informational approach may be somewhat limited substantively.

somehow resolved.” Because wars that end in a Carthaginian “peace” are relatively uncommon, the resolution cannot be reduced to the elimination of one of the belligerents. That is, the theory must be coherent and it must explain how fighting eliminates the commitment problem.

Gartzke argues that it cannot do so, at least in the context of large rapid shifts of power, because to end, a war must have eliminated the potential for a future increase in bargaining power for both sides. However, if it does so and there is complete information about the expected outcome of fighting, then actors should accept the consequences of their prospects instead of fighting. To wit, they should be able to obtain a similar settlement without a costly war. As he puts it, “if a solution exists to the commitment problem and states are assumed to be fully informed, then *ex ante* bargaining can occur” (p. 573). He then concludes that uncertainty is necessary for war to happen.

We shall show that this logic breaks down and that the reason it does so is its dependence on the original specification of the CCP model which treats war as an end-game. We shall demonstrate that when we model war as a process, large rapid shifts of power by themselves are not sufficient to produce fighting although they are necessary for it. Indeed, our results suggest that the CCP is much more subtle than that, and that fighting can break out and then end without the potential for such shifts disappearing. In other words, whereas the commitment problem may arise because of these shifts, the mechanism through which it does so has to do with how actors can credibly commit to fight if one of them reneges on a peace agreement rather than with how one can lock in benefits now instead of risking losing them in the future. It is this CPP mechanism—still intimately related to the original specification—that can serve as the basis for a coherent theory of war under complete information.

To summarize, we must require complete and coherent theories of war. Whereas the informational theory meets these requirements, it faces serious substantive objections. On the other hand, the complete-information theory based on credible commitments is incomplete. Because the mechanism that lies at its heart is so general, it is imperative to study it more closely and extend it to a coherent explanation.

We now enumerate several minimal features that the model must have to address the puzzle we have identified. First, we must represent war as a sequence of costly engagements rather than a game-ending costly lottery over exogenous outcomes. We must allow for negotiations during fighting, and we must allow for the game to end with a military victory in addition to a peace settlement. Second, we must incorporate large rapid shifts of power to create an environment where the original CCP can manifest itself. Third, contrary to most existing bargaining models, we should not assume that an agreement automatically ends the game with the division of benefits. Rather, we must allow for the possibility that actors can renege on the agreement and attempt to use the newly acquired resources to extract further concessions. In other words, peace, if it happens, must be endogenous to the model. Finally, ideally we would not want our results to be dependent on a particular choice of a bargaining protocol, and the model must have complete information.

## 2 The Model

Two players, each initially endowed with some capital  $K_i > 0$ , dispute a prize worth  $v \geq 0$ .<sup>7</sup> Each period  $t$  ( $t = 1, 2, \dots$ ) consists of two rounds: bargaining and fighting. Let  $k_i(t) > 0$  denote the amount of resources available to player  $i$  at the beginning of period  $t$ . In the bargaining round, players negotiate over the distribution of resources comprising the entire surplus  $S(t) = v + k_1(t) + k_2(t)$ . We leave the bargaining protocol unspecified and only require that negotiations during the round end in finite time. If players reach an agreement,  $(x_1, x_2)$  such that  $x_i \in [0, S(t)]$  and  $x_1 + x_2 \leq S(t)$ , they implement it and  $k_i(t) = x_i$  (resources become immediately available for fighting). If they fail to reach an agreement, they keep whatever resources they had at the beginning of the period. After the bargaining round, players simultaneously choose whether to fight. If both choose not to fight, the game ends and the negotiated distribution of resources remains. If at least one player chooses to fight the other, a battle occurs and each loses  $c_i \geq 1$  units of his resources. The game transits to the next period with probability  $(1 - p)$ , or at least one of the players collapses with probability  $p < 1$ . How this probability is distributed between the two players depends on their actions during the period: if they choose to fight, then player  $i$ 's opponent collapses with probability  $p_i$ , where  $p_1 + p_2 = p$ ; if  $i$  chooses to fight but  $j$  chooses not to, then  $j$  collapses with probability  $p$ , and  $i$  collapses with probability 0 (that is, if there is going to be a battle, it pays to participate in it). Since players will never fight if neither values the prize more than at least one battle, we assume that:

ASSUMPTION 1. The prize is worth fighting at least one battle for:  $p_i v > c_i \Rightarrow pv > C$ , where  $C = c_1 + c_2$  is the total cost of one battle.

How long players can survive fighting depends on their resource endowment. Let  $K_i > 0$  be player  $i$ 's initial capital stock. Because capital stocks are finite, the game ends in a finite number of periods. The payoffs are as follows. A player who collapses fighting in period  $t$  derives utility 0, and the surviving player gets  $S(t) - C$ ; that is, the victor absorbs the loser's remaining resources. If player  $i$ 's capital stock falls to zero in  $t$ , then  $i$  collapses automatically without an additional battle with a payoff of 0, and  $j$ 's payoff is  $S(t)$ . If both players collapse simultaneously, they split the total surplus equally, and each obtains  $S(t)/2$ . If both players choose peace at time  $t$ , the game ends and each keeps the  $k_i(t)$  as agreed upon. Since there is no confrontation in this case, there is no capital loss.

It is worth discussing briefly some of the assumptions in this model and how they relate to the requirements we outlined in the previous section. First, war is a sequence of engagements rather than an one-shot event. We shall refer to these engagements as battles with the understanding that we mean any period of time during which fighting occurs and actors do not negotiate (e.g., campaigns). Military victory can be achieved in two ways: either by causing the military collapse of the opponent during a campaign or by exhausting the opponent's resource base. However, the war can end as soon as both actors agree to a division of the benefits and as long as neither then engages in additional fighting after the distribution. The model thus allows for negotiated outcomes without assuming away the commitment to uphold the resulting distribution in peace. In addition, we have assumed

---

<sup>7</sup>For simplicity, we shall refer to player 1 (or a generic player  $i$ ) as "he," and player 2 as "she."

that when actors bargain, they can use their resources for side-payments too. We allow for a general bargaining protocol as long as it ends in finite time and does not artificially preclude settlement (we explain later what we mean by that). Furthermore, by allowing actors to turn around and use the resources obtained by negotiation for further war, we have effectively endogenized peace: any equilibrium that involves a successful negotiated settlement will necessarily incorporate disincentives to renege from it.

Second, we have created an environment where the original CCP can arise by allowing for large rapid shifts of power when an actor surprises its opponent by attacking when the other is not. The shift of power arises from the increased probability of military collapse of the actor caught by surprise. This is very similar to the first-strike advantage notion that Fearon (1995) uses. The difference is that instead of conferring an advantage for the entire war, it only does so temporarily on the tactical level. We find this assumption much more tenable for two reasons. Strategic surprise, although possible to achieve, is often indecisive for the entire war (the two most famous examples are the Japanese attack on Pearl Harbor in 1941 and the Egyptian/Syrian attack on Israel in 1973). On the other hand, tactical surprise often is decisive for the particular engagement, and such an engagement could end the war. For example, the Spartan surprise destruction of the Athenian fleet at Ageospotami ended the Peloponnesian War, the surprise Soviet invasion of Manchuria coupled with the dropping of the American atomic bombs ended the War in the Pacific, and the surprise Israeli crossing of the Suez Canal with the resulting encirclement of the Egyptian army in the Sinai ended the Yom Kippur War. The assumption that the side which achieves surprise has zero probability of collapse in that engagement is made for simplicity and does not affect the results as long as that probability is well below the one for the surprised opponent.<sup>8</sup> Finally, it is worth stressing that the costs are per engagement, not for the entire war. This means that Assumption 1 is not very demanding, and makes the model a bit more attractive on substantive grounds too: we do not have to assume, like most models do, that war costs are relatively small. Indeed, as we shall find, cumulative war costs here can be gargantuan, far exceeding the value of the prize actors are fighting over.

### 3 Total War

Total war occurs when players fight until one of them collapses from exhaustion. In this section, we establish the conditions for such an equilibrium. (All proofs are in Appendix A.) We first show that if there exists a period from which players will fight at least one battle, then they will also fight in all previous periods provided  $v$  is large enough. We then prove that players will fight to the end if at least one of them is sure to collapse after one battle. Together, these results imply that if  $v$  is large enough, total war will be inevitable: in any subgame perfect equilibrium (SPE), players will fight without redistributing resources until one of them collapses from exhaustion.

---

<sup>8</sup>We have further simplified the environment by keeping the probability of collapse constant. Whereas it might be natural to assume that it should increase as the actor nears exhaustion, the low expected payoff from fighting at such a stage is partly captured by the decreasing resources. We conjecture that allowing for variable collapse probabilities and tying these in a predictable manner to resources should not alter our results appreciably.

LEMMA 1. *If players fight for  $T \geq 1$  periods when the resources are  $(k_1, k_2, S)$ , then they fight  $T + 1$  periods when the resources are  $(k_1 + c_1, k_2 + c_2, S + C)$  provided that  $v$  and  $p$  are large enough.*

It is worth elaborating what this lemma means. We are not claiming that if players fight for some  $(k_1, k_2; S)$ , then they would fight for an arbitrary distribution  $(\hat{k}_1, \hat{k}_2; S + C)$  where  $\hat{k}_i \neq k_i + c_i$ . Rather, we prove that if the resource distribution  $(k_1, k_2; S)$  is such that players fight, then players will also fight when the resource distribution is exactly  $(k_1 + c_1, k_2 + c_2; S + C)$  as long as  $v$  and  $p$  are large enough. The next step is to locate a period in which players will fight for sure.

LEMMA 2. *Consider a period with  $(k_1, k_2; S + C)$ . If  $k_i \in [c_i, 2c_i)$  and  $k_j \geq 2c_j$ , then players fight in this period provided that  $p_j S > C$ .*

In words, if one of the players has just enough resources for exactly one additional battle, then no peace is possible.<sup>9</sup> If players have not redistributed resources until the period prior to  $i$ 's imminent collapse, then  $S > v$ , and hence  $p_j v > C$  will be sufficient to ensure war in that period. This means that if  $v$  is large enough and players do not redistribute, then the last battle is inevitable. We now show that if the condition of Lemma 1 is satisfied, then players would not, in fact, redistribute before the penultimate period, which implies that they will fight a total war.

PROPOSITION 1 (Total War). *If the conditions of Lemma 1 and Lemma 2 are satisfied, then players never redistribute resources and fight until the weaker one collapses from exhaustion.*

From our bargaining perspective, this case does not help answer the puzzle: the outcome here involves total annihilation of one of the players, which resolves the commitment problem in a trivial way. As Stalin famously (reportedly) quipped, "Death solves all problems—no man, no problem."

## 4 Limited War

Limited war occurs when players fight for some length of time and then settle on a negotiated redistribution. This is the most interesting case because it involves inefficient use of power under complete information: the opponents waste resources and then manage to negotiate the peace even though they agree in expectation on how the war will evolve from the very beginning. The results in this section demonstrate that peace is crucially dependent on the ability to threaten war, and in particular, the ability to threaten to impose sufficient costs by prolonging the fight. The problem actors face when negotiating peace is that these threats may not be credible: when it is common knowledge that negotiations will be available in the future, and actors will be tempted to reach a peaceful agreement then, the incentive to prolong the war in the future is undermined, and hence the threat to impose costs today becomes unbelievable.

---

<sup>9</sup>The lemma does not handle the case where both players are about to collapse simultaneously. We prove the analogous result in Claim 1 in Appendix A.

We now proceed in several steps. First, we show that for peace to occur, actors must credibly threaten to fight if it is violated. This suggests that the most permissive conditions for peace are those where players can credibly make the most deterrent threats, that is, threats that would impose the highest costs on the opponent. Since the most deterrent threats are those where players fight until the bitter end (when one of them collapses from exhaustion), the second step is to derive the equivalent to Powell’s (2004b) sufficiency condition: if peace cannot be attained when even the most deterrent threats are credible, then peace cannot be attained in any subgame perfect equilibrium (i.e., we have a sufficient condition for war). We demonstrate that fighting can lead to violation of this condition, opening up the road to peace. The next step is to show that peace can be attained once this condition fails. In the process, we demonstrate that although Powell’s condition is quite useful for what it was derived (ensuring that all equilibria are inefficient), it is less so for answering when peace becomes possible. To wit, this condition is “too strong” for war and therefore “too weak” for peace. We derive a stronger condition for peace, explain how it relates to Powell’s, and then construct an example equilibrium that has limited war, and which showcases the difference between the two conditions.

#### 4.1 Strongest Mutually Deterrent Threats

We begin with the following lemma, which makes it very clear that peace is sustained by the threat of fighting off the equilibrium path.

LEMMA 3. *Suppose  $(x_1, x_2)$  is a peaceful bargain when resources are  $(k_1, k_2; S)$  with  $S = v + k_1 + k_2$ . Then there is no peaceful bargain when resources are  $(x_1 - c_1, x_2 - c_2; S - C)$ .*

Lemma 3 shows that if players are able to conclude a bargain that is peaceful in equilibrium and one of them tries to deviate and attack after the redistribution of resources according to that agreement, then the next period surely involves fighting as well. The longer the fighting one can threaten with, the higher the expected costs for both players, and hence the less each would be willing to accept at the negotiating table, and the better the prospects for peace. The most deterrent threat is to fight to the bitter end. Generally, these threats will not be subgame perfect. However, the minmax strategies to fight to the end do form a Nash equilibrium. In other words, *if we allow players to commit to playing the Nash equilibrium that involves fighting to the end, then we are allowing them to make the most severe threats possible, which creates the strongest incentive to make peace today.* If for some period peace is impossible with the commitment to fight to the end should they fail to reach an agreement, peace certainly will not be possible in any subgame perfect equilibrium either.

We now turn to investigation of the conditions for war provided players can commit to their most deterrent threats. Suppose that given the current stocks  $(k_1, k_2; S)$ , the war can last at most  $T$  more periods. That is, if both players fight in each of the following periods without reallocating, then at least one of them will collapse after  $T = \min\{T_1, T_2\}$  fights, where  $T_i = \lfloor k_i/c_i \rfloor$ .

Denote the current period as the first and let  $F_i(t|T)$  denote player  $i$ ’s expected payoff from rejecting all offers and fighting to the end starting in period  $t$  and fighting up to, and including, period  $T$ . For example, suppose that one of the players would collapse after  $T = 7$  battles (so if they start fighting now, he would collapse in period  $t = T + 1 = 8$ ).

Then  $F_1(3|7)$  would denote player 1's expected payoff in period 3 from fighting to this end (that is, fighting five more battles). When  $T = T_i < T_j$ , if players never reallocate and fight in each period, eventually player  $i$  will collapse in period  $T + 1$ . Hence,  $F_i(T|T) = p_i(v + k_1 + k_2 - TC) = p_i(v + k_j - Tc_j)$  and  $F_j(T|T) = (1 - p_i)(v + k_j - Tc_j)$ . For  $t \in \{1, 2, \dots, T - 1\}$ , define the following recursive equation:

$$\begin{aligned} F_i(t|T) &= p_i(S - tC) + (1 - p)F_i(t + 1) \\ &= p_i \sum_{n=0}^{T-t-1} (1 - p)^n [S - (n + t)C] + (1 - p)^{T-t} F_i(T|T). \end{aligned} \quad (1)$$

This is player  $i$ 's expected payoff in period  $t$  if both players fight without redistribution until the weaker player collapses. It is also player  $i$ 's reservation value: the payoff that can be unilaterally guaranteed. The lower this payoff for player  $i$ , the more deterrent  $j$ 's threat. Since we want to find the most permissive condition for peace, we need to derive the most deterrent threats for both players. To do this, we show that the joint expected payoff from fighting until one player collapses from exhaustion is strictly decreasing in the number of periods the war is expected to last:

LEMMA 4.  $\sum_i F_i(t|T)$  is strictly decreasing in  $T$ .

In other words, if the distribution  $(x_1, x_2)$  enables players to fight  $T$  periods at most, and the distribution  $(y_1, y_2)$  enables them to fight one more period, then the sum of their expected fighting payoffs under  $(y_1, y_2)$  is strictly worse, and hence  $(y_1, y_2)$  involves more deterrent threats and is more conducive to peace. As the proof shows, the joint loss if peace fails is precisely  $(1 - p)^{T-t+1}C$ , which is the cost of the additional battle times the probability of having to fight it.

The result in Lemma 4 is intuitive: the longer players expect the war to last, the more resources they expect to waste waging it. Hence, their joint payoffs must necessarily decrease in expected duration. The most deterrent threats then are those where players fight for the longest possible time. Given total resources  $S$ , the maximum number of battles players can fight after some distribution is:

$$\bar{T} = \frac{S}{C}. \quad (2)$$

The largest number of battles under  $(x_1, x_2; S)$  is  $T_1 = x_1/c_1 = x_2/c_2 = T_2 \Rightarrow x_i = c_i S/C$ , where we used  $x_1 + x_2 = S$ . Therefore, if players equalize resources, they can make the most mutually deterrent threats possible, and the resulting environment is most conducive to peace. Since fighting is inefficient,  $\sum_i F_i(1|\bar{T}) < S$ , so bargains that improve on the minmax payoffs always exist. However, as we shall now see, this is not enough for peace to occur.

## 4.2 The Analogue to Powell's Condition

Simply offering a player his minmax payoff is not enough to induce him to agree to peace. Peace requires that *both players forego the advantages of surprise attack*. To see this, suppose a player accepted a division that exactly matched his minmax payoff,  $x_i = F_i(1|T)$ ,

and suppose in equilibrium players achieve peace immediately. Since they achieve peace,  $j$ 's strategy after this distribution must be not to attack. But then surprise attack yields  $i$  at least  $p(S - C) + (1 - p)F_i(2|T) > p_i(S - C) + (1 - p)F_i(2|T) = x_i$ . Therefore, attacking is a best response for  $i$ , which contradicts the assumption that  $x_i$  is accepted in a peaceful equilibrium. Because a peaceful bargain must deter surprise attacks, it must exceed the minmax payoffs.

To see the minimum demands that players would make, suppose they have divided everything such that  $(x_1, x_2; S)$  with  $x_1 + x_2 = S = k_1 + k_2 + v$ . Let  $T = \min\{x_1/c_1, x_2/c_2\}$  denote the largest number of battles they can fight under the new distribution without reallocation until one of them collapses from exhaustion. Player  $i$ 's payoff from sneak attack after the negotiated division is *at least*:

$$A_i(x_1, x_2) \geq p(x_1 + x_2 - C) + (1 - p)F_i(2|T).$$

As we have established, peace requires that surprise attacks are not profitable. Therefore, fighting a battle is going to be unavoidable if there are not enough resources to satisfy minimum deterrent demands. That is, players are sure to fight if:

$$A_1(x_1, x_2) + A_2(x_1, x_2) > S \quad \text{for all } (x_1, x_2; S). \quad (3)$$

This, of course, is the logic of Powell's (2004a) sufficiency condition which guarantees that complete-information bargaining will break down in any stochastic game, a general category that encompasses our model. To see that this is the case, note that  $F_i$  denotes player  $i$ 's minmax payoff in any period  $t$  because  $i$  can always guarantee himself this payoff by rejecting all offers and fighting in each period. Further, since  $A_i = p_j(S - C) + F_i(1|T)$ ,  $p_j(S - C)$  reflects the increase in  $i$ 's payoff if he catches his opponent by surprise and results from the temporary "power shift" in  $i$ 's favor whenever his opponent is expected not to fight. Hence, in any peaceful equilibrium,  $i$  must obtain at least  $A_i$ , leaving at most  $S - A_i$  to meet the minimal demand of the other player. In a peaceful equilibrium both players must have their minimal demands satisfied, yielding the condition in (3).

To ensure that peace will not be possible, we have to establish that no possible distribution can violate (3). By Lemma 4,  $\sum_i F_i(2|T)$  is minimized by taking the largest number of battles, which implies that  $\sum_i A_i \geq 2p(S - C) + (1 - p)\sum_i F_i(2|\bar{T})$ . Since the right-hand side of this inequality is the *worst* players can jointly expect, if this amount exceeds the available surplus, then fighting is guaranteed. In other words, we obtain a sufficient condition of war in the current period:

$$2p(S - C) + (1 - p)\sum_i F_i(2|\bar{T}) > S, \quad (4)$$

which reduces to:

$$p^2(\bar{T} - 1) + (1 - p)\bar{T} > 1. \quad (\text{P})$$

Condition (P) is sufficient to guarantee that peace will not be possible in period  $t = 1$ . It is analogous to Powell's (2004a), and we have emphasized this with our choice of nomenclature. The condition is defined entirely in terms of the fixed exogenous parameters and the total resources available at the beginning of the period. This means that we can apply

this condition to each period of the game by taking  $S_t = S - (t - 1)C$  to be the surplus in period  $t$ , and  $\bar{T}_t = S_t/C$  to be the maximum number of battles that can be fought until some player collapses from this period on. Even if (P) is satisfied and peace is impossible at  $t = 1$ , fighting reduces the available surplus and therefore the condition will eventually fail in some period provided that both players last that long. In other words, destruction seems to open up the road to peace.

In sum, fighting, whenever it occurs, depends on large rapid power shifts due to surprise attack advantages. The underlying mechanism parallels the one that causes inefficiency in the class of models covered by Powell's (2004*b*) condition. As we shall now see, however, these shifts are not the primary cause of war—they may be necessary, but certainly not sufficient, for fighting to occur.

### 4.3 The Conditions for Peace

Whereas (P) shows that limited war is possible in principle, it does not prove that it can happen. With a sufficient condition for fighting, we only have a *necessary condition* for peace in its converse. Although we know that the condition for peace can be achieved through fighting, we do not know whether players will be able to commit to peace once it is satisfied. Perhaps whenever they fight in equilibrium, they always end up in a total war? If this is the case, then our arguments do not take us very far. Therefore, it is imperative to demonstrate that limited war can happen in equilibrium. That is, that players fight and then settle, all with complete information.

The problem with using (P) to establish conditions for peace is that it is too strong for war. That is, it is designed to ensure that fighting will occur and is correspondingly too weak for peace: after all, when fighting is not guaranteed to break out, we cannot necessarily assume that peace will take hold. To see that, observe that (P) assumes that players fight to the end following a sneak attack. This minimizes the joint continuation value of the game and makes deviations to such attacks as unprofitable as possible. As we noted, these minmax threats will generally not be credible (subgame perfect) because actors may not be able to threaten to continue fighting if they are offered an acceptable peace deal. But if peace becomes possible after fewer fights and long before players are exhausted, then the joint continuation value of the game today increases, and with it, the deviation payoff goes up as well. In other words, surprise attack becomes more attractive, and this undermines the incentives to keep the peace.

Letting  $V_i(t)$  denote player  $i$ 's equilibrium payoff in the continuation game from period  $t$ , the necessary and sufficient condition for fighting in period  $t$  is:

$$2p(S - C) + (1 - p) \sum_i V_i(t + 1) > S. \quad (\text{F})$$

Because  $\sum V_i(t + 1) \geq 0$  and  $S > C$ , it follows that for  $p$  and  $S$  large enough, (F) will be satisfied. Bargaining in this period will break down and at least one battle will be guaranteed in any SPE. However, whereas (P) implies (F), there will be instances where (F) is satisfied but (P) is not. In other words, if condition (P) fails, we are not guaranteed that peace will occur because (F) may still be satisfied. This is why (P) is too loose for our purposes.

We now use condition (F) to construct an example SPE that demonstrates the main theoretical results that we use for our substantive discussion. We shall assume that the bargaining protocol does not artificially preclude peace—if (F) is violated, then players will achieve a peaceful redistribution in SPE—and that the entire available surplus is distributed between the players. Otherwise, we leave the protocol unspecified. This means that we cannot compute individual continuation values ( $V_i$ ), but we can compute joint ones ( $\sum_i V_i$ ).

Assume that players are *symmetric*, that is,  $K_1 = K_2 = K$ ,  $c_1 = c_2 = c$ , and  $p_1 = p_2 = p/2$ . Suppose the game begins with  $v > 0$ . After  $T$  battles without redistributing, each player has  $k_i = K - Tc$  resources left, so the total is  $S = 2k_i + v$ . If they distribute now, players face a situation with surplus  $S$  and  $v = 0$ , and therefore in any SPE the relevant behavior is that in these continuation games. Consequently, we now explore games with symmetric players and  $v = 0$ . We first show that in this setup, subgames with equalization of resources are particularly helpful because they capture all relevant aspects and are easy to analyze. We then use these results to construct a numerical example SPE.

The following lemma proves that in games with symmetric players who have equal resources, we do not lose any generality by restricting analysis to subgames in which players do not redistribute along the equilibrium path. Using this result, the lemma further derives the necessary and sufficient condition for fighting in any arbitrary period.

LEMMA 5. *Assume  $v = 0$  and symmetric players with  $k_i = nc$ , where  $n \geq 1$ . Then, without loss of generality, the players do not redistribute in equilibrium. Suppose that they find a peaceful settlement when  $k_i = nc$ . Then:*

(A) *they fight with  $k_i = (n + 1)c$  if, and only if,  $pn > 1$ ;*

(B) *if they fight with  $k_i = (n + t)c$ ,  $t = 1, \dots, T$ , then they fight with  $k_i = (n + T + 1)c$  if, and only if,*

$$p^2(n + T) + (1 - p)^{T+1} > 1. \quad (5)$$

Observe now that by Lemma 5, if players are symmetric and have equal resources, there is no loss of generality if we consider only SPE where players do not redistribute. This now means that if they cannot achieve peace in equilibrium with  $(nc, nc)$ , they cannot achieve it under any alternative allocation. This yields the following helpful result:

COROLLARY 1. *Suppose players are symmetric and cannot achieve peace if they redistribute such that  $k_i = nc$  and  $S = 2k_i$ . Then they cannot achieve peace under any alternative distribution.*

This is a powerful result: for any distribution  $(x_1, x_2)$  with  $v = 0$  and  $S = x_1 + x_2$ , we only need to check if players can achieve peace by equalizing their shares. That is, we use the conditions in Lemma 5 to check if players can achieve peace had they distributed  $(S/2, S/2)$ . This saves us a lot of work because otherwise we had to compute continuation values for any subgame with shares  $x_i - c$ , rather than just with  $S/2 - c$ . But since in any period with  $v > 0$ , the credibility of peace will depend on what happens after they redistribute, this result provides the key to unraveling the SPE in the entire game. The algorithm we use is to take any period, suppose they equalize resources, and check if they

will still fight using Lemma 5. If they do, then no alternative distribution can produce peace (by Corollary 1), and we know a battle is inevitable in any SPE. If they do not, then they will certainly achieve peace in this period (because we have identified at least one distribution that can do it and because we assumed that the bargaining protocol does not artificially limit their ability to utilize the opportunity).

We now construct an example SPE with symmetric players that exhibits the difference between our condition (F) and (P) and shows how players can fight and then settle in equilibrium under complete information.

**PROPOSITION 2 (Limited War).** *Assume  $v = 12$  and  $p = .18$ , and symmetric players with  $K_i = 30$ ,  $c_i = 1$ , and  $p_i = p/2$ . The players fight 12 battles without distributing resources and achieve peace in  $t = 13$  provided neither collapses in the interim.*

Since the construction is illustrative, we provide the proof here. The parameter values satisfy Assumption 1. Letting  $S$  denote the surplus in period  $t$ , condition (F) is  $S < .36(S - 2) + .82 \sum_i V_i(t + 1)$ . If (F) fails in some period  $t$ , then  $\sum_i V_i(t) = S$  because the entire surplus is peacefully distributed by our assumption that the bargaining protocol does not artificially preclude peace. If, on the other hand, (F) is satisfied, then peace is impossible in that period and  $\sum_i V_i(t) = p(S - C) + (1 - p) \sum_i V_i(t + 1)$ . Note that we are silent about the exact division that will occur because we have left the bargaining protocol implicit.

Recall from the proof of Proposition 1 that if players are sure to fight in some period  $t$ , they cannot improve matters by redistributing, and therefore there is no strict incentive to do so in equilibrium. In other words, we can restrict attention to SPE where players do not redistribute in any period in which they expect to fight for sure. This now allows us to backward-induct along the no-distribution path using condition (F) and Lemma 5.

Consider now the no-redistribution path at  $t = 30$ , where players have fought 29 battles already, and so  $S = 14$ . If they distribute and equalize resources now, they could fight  $n = 7$  more battles. By part (A) of Lemma 5, players will not fight with any  $n \leq 6$  but will fight with  $n = 7$  (because not fighting with  $n = 6$  results in  $p = .18 > .17 = 1/n$  in the current period). Therefore, players will surely fight at  $t = 30$  and settle in the next period (by collapsing simultaneously if they do not redistribute now or by redistributing if they do), and so  $\sum_i V_i(30) = p(14 - 2) + (1 - p)(12) = 12$ .

Going to the previous period along the no-distribution path, we have  $t = 29$  with  $S = 16$ . Equalization would yield  $n = 8$ , and since players fight with  $n = 7$  but settle with  $n = 6$ , part (B) of Lemma 5 applies with  $T = 1$  and  $n = 6$ . Solving (5) shows that players can achieve peace in this period. Therefore,  $\sum_i V_i(29) = 16$ . Continuing in this way, we construct Table 1, which shows the SPE outcomes for all periods of the game along the no-distribution path.

Observe that (P) is satisfied for all  $t \leq 4$ . This implies that peace is impossible in the first four periods even according to the stricter criterion.<sup>10</sup> However, even though (P) fails for all  $t > 4$ , there are many periods where peace cannot be achieved because (F) still holds. If, for example, this game started out with  $K_i = 25$ , then (P) would have no bite at all but

<sup>10</sup>That is, if we assume  $K > 30$ , the expected duration of war will be longer. The first period where peace is possible remains  $t = 13$ . Note that whereas the right-hand side of (F) always increases by  $C$  only, the left-hand side increases in increments larger than  $C$  once (P) is satisfied. Hence, the condition will never fail from this point on.

$t$	$(k_1, k_2; S)$	$\sum_i A_i$	Lemma 5	(P)	outcome	$\sum_i V_i(t)$
1	(30, 30; 72)	74.516	1.226	1.135	fight	61.92
2	(29, 29; 70)	72.381	1.214	1.103	fight	60.14
3	(28, 28; 68)	70.296	1.207	1.070	fight	58.42
4	(27, 27; 66)	68.271	1.204	1.038	fight	56.75
5	(26, 26; 64)	66.320	1.209	1.006	fight	55.16
6	(25, 25; 62)	64.459	1.221	0.974	fight	53.66
7	(24, 24; 60)	62.707	1.244	0.942	fight	52.27
8	(23, 23; 58)	61.088	1.278	0.910	fight	51.01
9	(22, 22; 56)	59.632	1.327	0.879	fight	49.91
10	(21, 21; 54)	58.375	1.394	0.847	fight	49.02
11	(20, 20; 52)	57.360	1.482	0.816	fight	48.36
12	(19, 19; 50)	56.640	4.320	0.785	fight	48.00
13	(18, 18; 48)	47.939	0.994	0.754	<b>peace</b>	48.00
14	(17, 17; 46)	46.187	1.017	0.723	fight	38.27
15	(16, 16; 44)	44.568	1.051	0.693	fight	37.01
16	(15, 15; 42)	43.112	1.100	0.663	fight	35.91
17	(14, 14; 40)	41.855	1.167	0.634	fight	35.02
18	(13, 13; 38)	40.840	1.256	0.606	fight	34.36
19	(12, 12; 36)	40.120	3.060	0.579	fight	34.00
20	(11, 11; 34)	33.672	0.971	0.553	<b>peace</b>	34.00
21	(10, 10; 32)	32.415	1.037	0.528	fight	27.02
22	(9, 9; 30)	31.400	1.126	0.505	fight	26.36
23	(8, 8; 28)	30.680	2.340	0.483	fight	26.00
24	(7, 7; 26)	25.335	0.940	0.465	<b>peace</b>	26.00
25	(6, 6; 24)	24.320	1.029	0.449	fight	20.36
26	(5, 5; 22)	23.600	1.800	0.437	fight	20.00
27	(4, 4; 20)	19.600	0.964	0.429	<b>peace</b>	20.00
28	(3, 3; 18)	18.880	1.440	0.427	fight	16.00
29	(2, 2; 16)	14.880	0.899	0.431	<b>peace</b>	16.00
30	(1, 1; 14)	14.160	1.080	0.444	fight	12.00

Table 1: Limited War SPE (No-distribution Path).

we will still get 7 battles in equilibrium. This illustrates our claim that (P) is too demanding for establishing the conditions for peace.

## 5 Discussion

We now turn to the substantive implications from the analysis of the Limited War SPE in Proposition 2. One very general result is a vindication for Varro’s dictum “if you want peace, prepare for war.” As Lemma 3 shows, a peaceful settlement must be sustained by the threat to punish attempts to exploit it; that is, peace necessarily involves a credible deterrent to surprise attacks. This is in keeping with results in Powell (1993) and Slantchev (2005a) who show that peace may require substantial military investments to produce such

mutual deterrence. That the notion of the *armed peace* has emerged in three rather different stylized conflict environments and that the logic behind it is essentially equivalent across them signify that it may be a very general phenomenon.

## 5.1 How War Resolves the Commitment Problem

Recall Fearon's (1995, 402-4) discussion of how first-strike advantages can close the bargaining range and cause war. As Powell (2005) has shown, the mechanism that causes inefficiency in that model is equivalent to the general commitment problem resulting from large rapid shifts of power. Intuitively, foregoing the advantages of a first strike produces a power shift in favor of one's opponent. The "declining" actor needs to be compensated for not striking first but the "rising" actor cannot credibly promise to deliver the rest of the compensation tomorrow when he finds himself in a strong position. As a result, the declining actor wages preventive war today.

The logic of the general power shift mechanism implicitly relies on a stylized representation of war as a one-shot game-ending move: if the declining state cannot be bought off, it goes to war. But what does such an explanation for war actually tell us? In reality, war is not once-and-for-all decision: it is a sequence of battles that may be concomitant with diplomacy, a process that ends either with the military destruction of one of the participants or with a negotiated settlement. An explanation that shows that a state can fight a battle under some circumstances is a rather weak explanation of war. More importantly, the mechanism through which this happens begs the question: *if states go to war because of a commitment problem caused by first-strike advantage, then how do they agree to end the war while this technology is still present?* How does war resolve this commitment problem?

Our model shows that the reason we cannot answer this puzzle with the original models has to do with a subtlety in the explanation that these models cannot uncover. We incorporated the equivalent to a first-strike advantage by assuming that surprise attack yields a temporary battlefield advantage and whereas we did find that war is possible in such an environment, it is not really caused by the resulting power shift. As Proposition 2 shows, players can begin fighting, wage war, and then negotiate a stable peaceful settlement while this advantage remains constant throughout. War is caused not by the power shift but by an inability to threaten with sufficiently harsh punishment should a player risk a surprise attack for immediate gain. This *war commitment problem* arises from the incentives to settle at the first opportune moment that both players have and know they have. The closure of the bargaining range does not occur because of a shift in the distribution of power but because both players expect to gain from fighting the war just a bit longer. It is this war commitment problem that fighting resolves. Whereas the surprise attack advantage (i.e., power shift) does not vary, the credibility of the threat to punish sneak attacks does: when the threat becomes sufficiently deterrent, peace can obtain. Obviously, any model that does not construct an environment where players have opportunities to settle the war cannot provide any purchase on this problem and its solution.

The example in Table 1 helps follow the logic. At  $t = 1$ , players could fight 30 battles without redistributing. Unfortunately, even threatening to fight all of them until they collapse from exhaustion cannot prevent fighting in the first five period (condition (P) holds). A rather dark implication of this analysis is that there may exist situations where even a

credible threat to fight to the end may not avert war. Sometimes the stakes can be so high that neither player can impose enough costs on its opponent to deter him from risking a few battles to win them.

From  $t = 6$  onward, however, players could achieve peace in any period if they only could threaten to fight to the bitter end. These threats are not credible because players know that if the war does not end by  $t = 13$ , they will negotiate peace then. To wit, surprise attack at, say  $t = 7$ , does not risk a total war but a limited one, and both players know this. Because of this, neither player can impose sufficient costs on the opponent to deter him from sneak attack, and the incentives to strike a bargain in all prior periods dissipate. To wit, players are not credibly prepared for war, and therefore cannot obtain peace.

It is worth comparing the sufficient condition for fighting in (5) with (P). Recall that  $\bar{T}$  is defined as the maximum number of battles players can fight starting in the current period. By the definitions of Lemma 5, players settle peacefully when  $k_i = nc$ , that is, when they can, in principle, fight  $n$  more battles. Further, by assumption they will fight  $T$  more battles starting tomorrow if they fight today, so the maximum number of battles they can fight from the current period is  $n + T + 1$ . In other words,  $\bar{T} - 1 \equiv n + T$ , which means that the coefficient on  $p^2$  is identical in both expressions. This implies that (5) and (P) differ in the second term only. Since  $1 - p < 1$  and  $T + 1 \leq \bar{T}$ , it follows that  $(1 - p)^{\bar{T}} \leq (1 - p)^{T+1}$ . In other words, whenever (P) is satisfied, so will (5) be, but the converse is not true. If players could fight many battles but will settle only after a few,  $T + 1$  will be much larger than  $\bar{T}$ , and it will be quite possible for (5) to hold even if (P) does not.

The difference between the two is intuitive: whereas (P) uses the largest number of battles, (5) only uses the number that players are actually expected to fight *in an equilibrium with credible threats*. This highlights our main conclusion: the ability to achieve peace critically depends on the credibility of the mutually deterrent threats that players can make. Even if (P) could be satisfied with Nash threats, players have no reason to believe them but will instead only take into account how many battles the opponent is actually prepared to fight before settling. As we have seen, this unhappy calculation can undermine the incentives for peace today. Conversely, if one can credibly threaten to punish an opportunistic move by imposing very large costs on the attacker, then peace can be sustained.<sup>11</sup>

Intuitively, inability to threaten with prolonged war causes two things. First, one's own expected payoff from continuing to fight *increases* because in no event would one accept peace terms in the future that would be worse than one's expected payoff from continuing to fight then. In other words, the expectation of a future settlement makes one more intransigent today and decreases the concessions one would be willing to make. Second, it also *increases* the opponent's payoff from continuing the war because it decreases the expected level of destruction he will have to suffer. In other words, because one cannot credibly threaten to continue to the bitter end, one's opponent expects to have to fight fewer battles, and correspondingly to have to pay less for foregoing peace today. This makes the opponent more intransigent as well. Ironically, one's rational incentives to obtain peace by negotiation may actually undermine the ability to do so. Conversely, one's seemingly belligerent

---

<sup>11</sup>See Slantchev (2005b) on how credible threats to punish potential revisionists arose from the territorial distribution at the Congress of Vienna in 1815, and how these sustained peace in Europe during the first half of the 19th century.

and unyielding threat to fight to the end may considerably improve the prospects for peace.

How does fighting resolve this war commitment problem then? Unhappily, it does so by war's very nature, its sheer destructiveness. Initially, both players are rich in resources and the stakes are high. What are a few battles compared to the possibility of obtaining these riches should a military operation prove successful? As war progresses, the pie shrinks and continuation becomes less and less tempting, which in turn means that it takes weaker threats to deter participants from surprise aggression. Eventually, players can credibly commit to fight a handful of battles and this becomes sufficient to ensure that peace can be maintained. Every war carries the seeds of its own peace.

## 5.2 War Costs and Rational Escalation

Can two players rationally escalate war and end up agreeing to peace terms that are worse than what each originally started the war with? That is, can rational players pay war costs that exceed the value of the prize?

In the Dollar Auction game introduced by Shubik (1971), two players alternate in bidding for a prize of one dollar. The highest bidder wins the prize but both have to pay their bids. This game often gives rise to escalatory behavior: as soon as a player is outbid, he attempts to defend his, now losing, bid. If he escalates, he might yet win and at least deduct the value of the prize from his final cost, whereas if he stays put, he loses his entire investment.

O'Neill (1986) analyzes a discrete complete-information version of the dollar auction with budget constraints. He finds a unique subgame perfect equilibrium in which no escalation occurs—player 1's initial bid forces player 2 out of the game. This result differs markedly from experimental studies of the Dollar Auction in which players often escalate, sometimes bidding more than the value of the prize itself (Teger 1980). O'Neill concludes that such escalatory behavior must be irrational.<sup>12</sup> He offers nine "entrapping" features that promote escalation in international settings but are not captured by his model, and most of them are forms of irrationality: bounded rationality (e.g., players cannot look far enough ahead), psychology (e.g., desire to win for sake of winning), and misperception.<sup>13</sup>

The Dollar Auction has been used extensively to shoehorn interstate crises and wars, among other events, into its interpretive framework of loss-avoidance, a case of "we have invested too much to quit now." For many analysts, the game provides an especially apt analogy that illustrates the pernicious consequences of various forms of irrationality. They see escalation as pathological, and the policy prescriptions derived usually take the form of a wishlist: if only players could foresee. . . , if only they knew. . . , if only they could admit

---

<sup>12</sup>In this he echoes Shubik's (1971, 111) original warning that "the game's [experimental] play appears to depend upon virtually only the social-psychology of the players." Shubik attributes escalation to the players' inability to use signals and make commitments, and so implicitly points to incomplete information as the culprit.

<sup>13</sup>Leininger (1989) takes issue with this conclusion. He does find that *any* subgame perfect equilibrium of the dollar auction game with discrete bids and the unique SPE in a continuous variant have the same qualitative features as O'Neill's solution. The version of the model where escalation does happen in SPE involves an assumption about what happens when nobody wins the prize because players bid indefinitely (since bids are continuous, they could always outbid each other without reaching their budget constraints). This, of course, is a rather unrealistic feature of the model and perhaps should have motivated its reassessment instead of forming the basis for substantive conclusions.

their mistakes. . . .<sup>14</sup> The general implication is clear: if only players were fully rational and knew everything about each other, then we would never see “senseless” escalation that ends with both of them paying more than the prize itself is worth.

This conclusion seems to be supported by O’Neill’s (1986) analysis and is implicitly endorsed by a great many formal models of conflict which assume that the costs of war are less than the value of the prize. In fact, even the canonical bargaining model of war (Fearon 1995, Powell 1996) is forced to make this assumption if conflict is ever to occur with positive probability in equilibrium. After all, if the costs of war are expected to exceed the value of victory, then a rational player would never go to war (Bueno de Mesquita 1981).<sup>15</sup> In other words, most of our rationalist theories of war tacitly agree with the view that escalation of the sort that happens in experimental plays of the Dollar Auction is inherently irrational.

Leaving aside the question of whether the Dollar Auction is actually a good model of crises and wars (after all, it admits no bargaining and no negotiated outcomes), we argue that excessive escalation can happen with fully informed rational players even when they are virtually unconstrained by the bargaining protocol. In this, we do not select an inefficient equilibrium among many efficient ones (Slantchev 2003a) and we do not rely on strategic uncertainty induced by mixed strategies (Langlois and Langlois 2005). While we do not mean to discount psychological explanations, we want to emphasize that there is no need to resort to irrationality to understand situations in which both players end up paying more than the prize itself is worth.

When we replace the heroic assumption that war costs are low relative to the value of the stakes with the milder one that the stakes are worth at least one battle, the tragedy of war reveals itself: *If players survive to make a peaceful settlement, war is sure to have become a net loss for both!* When players settle at  $t = 13$  in the Limited War SPE, the amount to be divided is  $S = 48$ , whereas it started out at 72. They have collectively paid war costs of 24 even as the prize itself is only worth 12. Escalation here happens because costs already paid are sunk (and so irrelevant for the current calculation), because players have incentives to risk a small additional amount (costs of one battle) to obtain the whole prize should they happen to win militarily, and because they cannot be deterred from taking these risks.<sup>16</sup>

At this point one may quibble that whereas the prize in our model is worth only 12, the actual prize is much larger because if a player wins militarily he obtains the resources of the loser. This misses the thrust of our argument, which is that *at the time of peace*, players accept shares that leave them worse off relative to just conceding the prize from the outset and they willingly escalate to that point. For example, if players share equally

---

<sup>14</sup>See, for example, Poundstone (1992, 261-65) and Teger (1980) for some examples. Deutsch’s (1973, 356-57) analysis of U.S. involvement in Vietnam essentially follows the same logic.

<sup>15</sup>We thank Ken Schultz for forcefully arguing against this shortcoming of existing models. The reason for this empirically unsustainable assumption is inherent in modeling war as a single-shot lottery over exogenous outcomes. See Wagner (2000) and Slantchev (2003a) for critiques of this approach.

<sup>16</sup>One can think of the probability of winning outright as being the equivalent of a player’s belief that his opponent would quit if he escalates his bid slightly in the Dollar Auction game. The major difference is that here we are dealing with an exogenous uncertainty inherent in military engagements rather than private information whose revelation can be strategically manipulated. In the experimental Dollar Auction players escalate, at least in part, to signal resolve. In our complete information model communication is meaningless because players already know everything there is to know.

at  $t = 13$  (not unlikely given their symmetry), each would obtain a payoff of 24. This is worse than immediate concession at the outset and it is less than the initial resources the player had. Each has paid war costs of 12 for the dubious privilege of obtaining a benefit of 6. Furthermore, since (P) would hold if we increase the size of initial stocks, richer players would fight longer wars. For example, letting  $K = 35$  results in five additional battles before peace, and total individual costs of 17 at the time of negotiated peace.

War becomes unprofitable very quickly: from  $t = 7$  on, at least one of the players cannot recover the resources he started the war with. In our example, negotiated peace can occur only after war has lasted nearly twice the duration that could be potentially profitable for one of the players. However, the piling costs do not deter players from fighting because they are sunk, and hence all that matters are the expected future gains and losses. It is that forward-looking aspect of war that may make it a dead loss to both players if they fight into peace. Pillar (1983, 173) argues that it can be rational to continue fighting even after the war “has already escalated well out of proportion to the value of the objectives at stake” because one cannot manipulate past costs, only future ones. The model vindicates this logic even in an environment where all cost manipulation is only implicit in the threat to continue to fight. It also provides rationalist theoretical foundations the empirical findings by Orme (2004) and what he terms the “paradox of peace”: peace is most likely when the threat of costly conflict is greatest.

These results show that it is rational to risk small escalatory steps that eventually may accumulate enough costs to exceed the value of the issue at stake. This highlights the problematic assumption in traditional models of war and suggests that we may need to rethink some of the causal mechanisms derived from such theories.

### 5.3 War as Punctuated Equilibrium

Our analysis suggests that one can usefully view war as a mutually coercive process that involves continuous fighting punctured by occasional opportunities for peace. Even though peace negotiations are available throughout the war, a credible commitment to a settlement is only possible at specific junctures. There are specific *windows of opportunity* to end the conflict, and if such a window closes, players are stuck fighting until the next one comes along. In a sense, it seems true that conflicts have to be “ripe for resolution” (Young 1967, Zartman 1985). However, in our formulation ripeness is not a battlefield property of conflict that appears when players reach a “mutually hurting stalemate” (Zartman and Berman 1982). Rather, it is a function of the credibility of threats to punish attempts to take advantage of the peace negotiations.

Zartman (2001) notes that when parties are locked in a mutually hurting stalemate, they would seek a way out, and the sense of impending catastrophe may urge them to take action now. Something similar then happens in our model: if players fail to cease the fleeting opportunity to end the war, they will be condemned to fight it out until another window presents itself. In our example, should players for some reason be unable to negotiate at  $t = 13$ , they have to carry on the war for seven more periods until  $t = 20$  opens up the possibility for peace again. At this puncture in fighting, the terms of peace are much worse for both: players would have jointly paid costs of 40 to divide the prize, and each can hope to live with a little more than a third of his original resources. The desire to avoid this

additional fighting and worse outcome in turn induces players to agree to peace at  $t = 13$ . Observe, however, that nowhere in our model are players stalemated (they can always risk a battle that gives a chance of outright victory) and neither does inability to negotiate a termination of war hinge on problems with perception.

These windows of opportunities are rarer when the stakes are higher. The more resource-rich the warring parties, the longer the fighting spells between these windows. Their frequency, however, increases the longer the war lasts, and their closure gets ever shorter as opponents approach exhaustion. That is, the more weakened the actors are from fighting, the more willing to negotiate they become. As they approach collapse, the terms of peace begin to approximate the expected payoff from continued fighting, obviating the incentive to risk it. As the terms of peace deteriorate, so does the expected payoff from prolonging the war, and hence the prospects for war termination improve. Ironically, the better the expected terms of the settlement, the worse the prospects for immediate peace. This is because the peace settlement itself is a function of the available benefits to be divided and since fighting may secure these benefits completely, the stronger the incentive to risk it.

It is this pattern of windows of opportunities for peace, which cluster toward the military end of war, that leads us to view war as a *punctuated equilibrium*.

## 6 Conclusion

One of the canonical rationalist explanations of war is that opponents cannot credibly commit themselves to follow through on the terms of agreement because a change in relative power renders such promises against their interests. Actors then may prefer to start a war today rather than face the unpalatable consequences of peace tomorrow. However fundamental and intuitive, this mechanism is incomplete and incoherent because it does not explain how fighting alleviates that commitment problem. We argued that unless we view war as a process, we will not be able to resolve this puzzle.

The analysis uncovered a subtlety that essentially turns the original commitment problem on its head. In our account, an actor's inability to promise credibly to fight for long lowers the costs of war and causes his opponent to demand so much today that he prefers to continue fighting rather than concede. The credibility problem arises from the opportunities for peace in the future: when both actors know that they want to settle the costly conflict as soon as possible, threats to extend fighting beyond such an opportunity for peace become unbelievable. Actors are tempted to risk some more fighting because they cannot deter each other by threatening not to negotiate in the future. In showing how fighting resolves that problem, we provide a complete and coherent rationalist explanation of war that does not require asymmetrically informed players.

Ironically, the very desirability and possibility of peace make war more likely because they decrease its expected duration and costs. An obvious tactic then suggests itself: if one could conceal such temptation to negotiate and somehow commit not to seek peace until a military resolution of the conflict, the likelihood of being able to negotiate an early termination will increase. Of course, it should also be obvious how difficult it will be to pull such a trick: one must simultaneously demonstrate complete resolve to fight to the bitter end and willingness to negotiate peace. The problem becomes worse in societies where leaders might be constrained by the public to fight short wars: all else equal, democracies

may be unable to mount credible threats to fight to the end, and this may embolden their opponents and needlessly prolong the wars they fight. After all, it is not at all clear that democratic leaders cannot mobilize for a long haul in spite of widespread opposition. But then again, neither it is clear that these leaders will persevere against popular opinion for too long.

## A Proofs

*Proof of Lemma 1.* Let  $V_i(k_1, k_2; S)$  denote  $i$ 's equilibrium payoff in the game that begins with total surplus  $S$  and capital stocks  $(k_1, k_2)$  such that  $k_1 + k_2 < S$ . Suppose that when the resources are  $(k_1, k_2; S)$ , players will fight for  $T \geq 1$  more periods. That is, after  $T$  more battles, either one of them will collapse from exhaustion or else they will reach some peaceful bargain. Consider now the distribution  $(k_1 + c_1, k_2 + c_2; S + C)$  and any feasible  $(x_1, x_2)$  such that  $x_1 + x_2 = S + C$  and  $x_i \geq 0$ . If  $i$  rejects this distribution and fights this period, then they continue with  $(k_1, k_2; S)$  and fight  $T$  periods by our supposition. Let  $V_i(T + 1) = V_i(k_1 - Tc_1, k_2 - Tc_2; S - TC)$  be the short-hand expression for the payoff players expect to obtain in the period after  $T$  fights starting tomorrow. Recalling that today's surplus is  $S + C$ , fighting today and then for another  $T$  periods starting tomorrow can be compactly expressed as:

$$\begin{aligned} W_i(1|T + 1) &= p_i S + p_i \sum_{n=1}^T (1-p)^n (S - nC) + (1-p)^{T+1} V_i(T + 1) \\ &= p_i \sum_{n=0}^T (1-p)^n [S - nC] + (1-p)^{T+1} V_i(T + 1). \end{aligned}$$

Player  $i$  can expect  $W_i(1|T + 1)$  if he rejects a deal and fights today. Hence, if  $(x_1, x_2)$  is a peaceful bargain, it must be the case that:

$$x_i \geq W_i(1|T + 1) \text{ for } i = 1, 2. \quad (6)$$

We want to show that for any feasible distribution  $(x_1, x_2)$  that satisfies (6), there exists  $i$  such that:

$$x_i < pS + (1-p)V_i(x_1 - c_1, x_2 - c_2; S). \quad (7)$$

That is, such a distribution is not peaceful because at least one player expects to do better by deviating and fighting at least one battle. We proceed by contradiction. Suppose that there does exist  $(x_1, x_2)$  such that  $x_1 + x_2 = S + C$ , and (6) is satisfied, but (7) is not. Suppose now that  $(x_1 - c_1, x_2 - c_2)$  is a peaceful bargain for  $(x_1 - c_1, x_2 - c_2; S)$ . In this case,  $V_i(x_1 - c_1, x_2 - c_2; S) = x_i - c_i$ , which means that  $x_i \geq pS + (1-p)(x_i - c_i)$  by our original supposition. We now obtain:

$$\sum_i x_i = S + C \geq 2pS + (1-p)(x_1 - c_1 + x_2 - c_2) = 2pS + (1-p)S \Rightarrow C \geq pS,$$

which contradicts Assumption 1 because  $S > v \Rightarrow pS > C$ . Therefore,  $(x_1 - c_1, x_2 - c_2)$  cannot be a peaceful bargain when resources are  $(x_1 - c_1, x_2 - c_2; S)$ . That is, it must be

the case that for some  $i$ ,

$$x_i - c_i < p(S - C) + (1 - p)V_i(x_1 - 2c_1, x_2 - 2c_2; S - C). \quad (8)$$

We now show that if  $v$  and  $p$  are large enough, then (8) implies (7), which will establish our claim. Because  $(x_1 - c_1, x_2 - c_2)$  yields a fight, it follows that:

$$V_i(x_1 - c_1, x_2 - c_2; S) = p_i(S - C) + (1 - p)V_i(x_1 - 2c_1, x_2 - 2c_2; S - C).$$

Combining this with (8) yields:

$$V_i(x_1 - c_1, x_2 - c_2; S) > p_i(S - C) + x_i - c_i - p(S - C) = x_i - c_i - p_j(S - C).$$

Multiplying both sides by  $(1 - p)$  and adding  $pS$  to each produces the equivalent inequality:

$$pS + (1 - p)V_i(x_1 - c_1, x_2 - c_2; S) > pS + (1 - p)[x_i - c_i - p_j(S - C)].$$

We now show that  $pS + (1 - p)[x_i - c_i - p_j(S - C)] > x_i$ , which will complete the proof because it will establish (7) directly. Using some algebra to simplify the last expression, we obtain:

$$(p_i + pp_j)S + p_j(1 - p)C > px_i + (1 - p)c_i.$$

From (6), we have  $x_j \geq W_j(1|T + 1)$ , which implies that  $x_i \leq S + C - W_j(1|T + 1)$ , and so  $px_i + (1 - p)c_i \leq p[S + C - W_j(1|T + 1)] + (1 - p)c_i$ . Hence, showing that

$$(p_i + pp_j)S + p_j(1 - p)C \geq p[S + C - W_j(1|T + 1)] + (1 - p)c_i$$

would be sufficient. Rearranging terms yields the inequality we need to establish:

$$pW_j(1|T + 1) \geq p_j(1 - p)(S - C) + pC + (1 - p)c_i. \quad (9)$$

Simplifying  $W_j(1|T + 1)$  yields:

$$\begin{aligned} W_j(1|T + 1) &= p_j \left[ S \cdot \sum_{n=0}^T (1 - p)^n - C \cdot \sum_{n=0}^T (1 - p)^n n \right] + (1 - p)^{T+1} V_j(T + 1) \\ &= p_j \left[ S \cdot \frac{1 - (1 - p)^{T+1}}{p} - C \cdot \frac{1 - p - (1 - p)^{T+1}(pT + 1)}{p^2} \right] \\ &\quad + (1 - p)^{T+1} V_j(T + 1) \\ &= p_j \left( \frac{1 - (1 - p)^{T+1}}{p} \right) (S - C) \\ &\quad + p_j \left( \frac{2p - 1 + (1 - p)^{T+1}[p(T - 1) + 1]}{p^2} \right) C \\ &\quad + (1 - p)^{T+1} V_j(T + 1), \end{aligned}$$

where the last step was obtained by adding and subtracting  $\frac{1 - (1 - p)^{T+1}}{p} \cdot C$  to the bracketed term in the expression above it and distributing terms. Comparing the coefficients on  $(S - C)$  in (9) produces  $p \geq (1 - p)^{T+1}$ , which holds for any  $T$  as long as  $p \geq 1/2$ . (Note that as  $T$  increases, the smallest value of  $p$  that satisfies the inequality decreases.) Consequently, (9) is satisfied provided this is the case and  $v$  is sufficiently large.  $\square$

*Proof of Lemma 2.* Let the distribution be  $(k_1, k_2; S + C)$ . Without loss of generality, let  $k_1 \in [c_1, 2c_1)$ . That is, without a redistribution of resources player 1 has enough to fight just one last battle. Assume that  $p_2S > C$ , and note that this implies  $pS > C$  as well. Assume also that  $k_2 \geq 2c_2$ , so that player 2 will outlast player 1 if they fight.

We now establish the bounds on player 1's share in any equilibrium bargaining division, and then show that no peaceful division exists within these bounds. Consider some arbitrary division of the surplus  $(x_1, x_2)$  such that  $x_1 + x_2 = S + C$ . If 2 rejects this and fights, her expected payoff is  $p_2S + (1 - p)S = (1 - p_1)S$ . Hence, player 2 will not accept any share  $x_2 < (1 - p_1)S$ , which implies  $x_1 \leq S + C - (1 - p_1)S = p_1S + C$ . Similarly, player 1 can guarantee  $p_1S$  by rejecting any division and fighting. Hence,  $x_1 \in [p_1S, p_1S + C]$  for any feasible equilibrium bargaining division.

Suppose now that a peaceful bargain  $(x_1, x_2)$  exists. Then  $x_1 \geq pS + (1 - p)V_1(x_1 - c_1, x_2 - c_2; S) \geq pS$ . Since  $x_1 \leq p_1S + C$ , it follows that  $p_1S + C \geq pS \Leftrightarrow C \geq p_2S$  must hold. This contradicts our assumption and hence no peaceful  $(x_1, x_2)$  exists. In this case, player 2 is better off without a redistribution that expands player 1's expected life span, and player 1 will not accept any distribution that decreases it. Hence, after any potential distribution, player 1's life span must remain unchanged, and so we have  $(\hat{k}_1, \hat{k}_2)$  with  $\hat{k}_1 \in [c_1, 2c_1)$ . If this distribution is peaceful, then  $\hat{k}_1 \geq pS$  and  $\hat{k}_2 \geq pS + (1 - p)S = S$ . Adding these inequalities yields  $S + C \geq (1 + p)S \Rightarrow C \geq pS$ . This contradicts our assumption, and therefore players fight.  $\square$

CLAIM 1. *Lemma 2 establishes the condition that ensures that players will fight if one of them is about to collapse. We noted that this ignores the special case where both players are about to collapse. We now provide the conditions that ensure no peace is possible in this case either.*

*Proof.* Take  $(k_1, k_2; S + C)$ . Suppose  $k_1 \in [c_1, 2c_1)$  and  $k_2 < 2c_2$ , so that both players will collapse after one fight. Player  $i$  can guarantee  $p_iS + (1 - p)S/2$  by rejecting any division and fighting. Because  $x_j \geq p_jS + (1 - p)S/2$ , we also have  $x_i \leq S + C - p_jS - (1 - p)S/2$ . This means that any feasible equilibrium distribution must satisfy  $x_i \in [(1 + p_i - p_j)S/2, (1 + p_i - p_j)S/2 + C]$ .

Suppose now that a peaceful bargain  $(x_1, x_2)$  exists. Then  $C + (1 + p_i - p_j)S/2 \geq x_i \geq pS + (1 - p)V(x_1 - c_1, x_2 - c_2; S) \geq pS$ . Summing over  $i$ , we obtain  $C \geq (p - 1/2)S$ . If  $p > 1/2$  and  $S$  large enough, this inequality is not satisfied. (Note that if players have not redistributed resources, as they would not in a total war, then  $S \geq v$ , and hence a large enough  $v$  will break the inequality.) This means that for some  $i$ ,  $pS > C + (1 + p_i - p_j)S/2$ . This now yields  $pS + (1 - p)V(x_1 - c_1, x_2 - c_2; S) \geq pS > C + (1 + p_i - p_j)S/2 \geq x_i$ . That is,  $i$  strictly prefers to attack when  $j$  does not attack, and so no peaceful  $(x_1, x_2)$  exists.

We now show that rejecting any redistribution  $(\hat{x}_1, \hat{x}_2)$  such that  $\hat{x}_i \notin [c_i, 2c_i)$  for  $i = 1, 2$  is a weakly dominant strategy for each player. No player accepts  $\hat{x}_i < c_i$ . Further,  $i$  can guarantee  $S/2$  in the next period if they both collapse. If  $\hat{x}_i \geq 2c_i$  for  $i = 1, 2$ , then the sum of expected payoffs is no more than  $S$  in the next period, and is strictly less than  $S$  if they fight in that period. Hence, there is at least one player who obtains no more than  $S/2$  in the continuation game. This means rejecting  $(\hat{x}_1, \hat{x}_2)$  and fighting is a weakly dominant strategy for that player. Therefore, players do not redistribute to expand their life spans in

equilibrium. This implies that  $(\hat{k}_1, \hat{k}_2)$  with  $\hat{k}_i \in [c_i, 2c_i)$  for  $i = 1, 2$  is the only possible equilibrium distribution. If this distribution is peaceful, then  $\hat{k}_1 \geq pS + (1-p)S/2$  and  $\hat{k}_2 \geq pS + (1-p)S/2$ . Adding these inequalities yields  $S + C \geq (1+p)S$ , which implies  $C \geq pS$ , which contradicts Assumption 1, and therefore players fight.  $\square$

*Proof of Proposition 1.* Consider the extensive form of the game and take the path where players fight in each period without redistributing resources. Its terminal node is the collapse from exhaustion of one of the players, say player 1. Hence, at the penultimate node resources are  $(k_1, k_2; S+C)$  such that  $k_1 \in [c_1, 2c_1)$  and  $k_2 \geq 2c_2$ , with  $S+C = v+k_1+k_2$ . By Lemma 2, players will fight at that node and will not redistribute resources. Consider now the node prior to that, with resources  $(k_1+c_1, k_2+c_2; S+2C)$ . Since players fight one battle with  $(k_1, k_2; S+C)$ , Lemma 1 implies that they will fight at that node as well. Note that the lemma does not say that they will not redistribute, only that they will fight even if they redistribute.

We now show that players will not redistribute at that node. For any arbitrary  $(k_1, k_2; S)$ , a redistribution  $(x_1, x_2)$  is feasible if, and only if,  $V_i(x_1, x_2; S) \geq V_i(k_1, k_2; S)$  for  $i = 1, 2$  because otherwise at least one player will not accept it and will instead continue with  $(k_1, k_2; S)$ . Now consider some  $(k_1, k_2; S)$  that is sure to result in a fight. We claim that if  $(x_1, x_2)$  is a feasible bargain for  $(k_1, k_2; S)$ , then  $(x_1 - c_1, x_2 - c_2)$  is a feasible bargain for  $(x_1 - c_1, x_2 - c_2; S - C)$  as well, and so players do not have a (strict) incentive to redistribute today. To see this, consider any  $(x_1, x_2)$  that is feasible for  $(k_1, k_2; S)$  that does not have a peaceful solution. Then:

$$\begin{aligned} V_i(x_1, x_2; S) &= p_i(S - C) + (1 - p)V_i(x_1 - c_1, x_2 - c_2; S - C) \\ V_i(k_1, k_2; S) &= p_i(S - C) + (1 - p)V_i(k_1 - c_1, k_2 - c_2; S - C) \end{aligned}$$

because neither  $(k_1, k_2)$  nor  $(x_1, x_2)$  can avoid a fight today. Combining these inequalities with the fact that  $(x_1, x_2)$  is feasible yields  $V_i(x_1 - c_1, x_2 - c_2; S - C) \geq V_i(k_1 - c_1, k_2 - c_2; S - C)$ . That is,  $(x_1 - c_1, x_2 - c_2)$  is a feasible bargain for  $(k_1 - c_1, k_2 - c_2; S - C)$ . Hence, players will not redistribute at  $(k_1, k_2; S)$  if they are sure to fight there.

Returning to our argument, players will fight without redistribution at the node with  $(k_1 + c_1, k_2 + c_2; S + 2C)$ . Repeated application of that lemma to all preceding nodes along this path unravels the game back to the initial node and ensures that players will not attempt redistribution there either. In other words, the only equilibrium outcome is total war.  $\square$

*Proof of Lemma 3.* Seeking a contradiction, suppose  $(x_1 - c_1, x_2 - c_2; S - C)$  admits a peaceful bargain. Since  $(x_1, x_2)$  is peaceful  $x_i \geq p(S - C) + (1 - p)V_i(x_1 - c_1, x_2 - c_2; S - C)$  for  $i = 1, 2$ . This implies  $x_1 + x_2 = S \geq 2p(S - C) + (1 - p)\sum_i V_i(x_1 - c_1, x_2 - c_2; S - C) = 2p(S - C) + (1 - p)(S - C) = S - C + p(S - C)$ . This now implies  $C \geq p(S - C)$ . Noting that  $S = k_1 + k_2 + v$  and  $k_1 + k_2 \geq C$ , it follows that  $S - C = k_1 + k_2 + v - C \geq v$ , and so the previous inequality implies  $C \geq p(S - C) \geq pv$ , which contradicts Assumption 1. Hence, there is no distribution of resources that can sustain peace when the surplus is  $S - C$  if players can agree to a peaceful bargain when the surplus is  $S$ .  $\square$

*Proof of Lemma 4.* Note that  $\sum_i F_i(T|T) = v + K_1 + K_2 - TC = S - TC$ , where  $S$  denotes the total amount of resources available in the first period. We can then write the

sum of expected fighting payoffs after division in period  $t$  as:

$$\sum_i F_i(t|T) = p \sum_{\tau=0}^{T-t-1} (1-p)^\tau [S - (\tau+t)C] + (1-p)^{T-t} (S - TC).$$

We now show that this expression is strictly decreasing in  $T$ . Take an arbitrary  $T \geq 2$  and consider another division that enables players to fight one more period. Then  $\sum_i F_i(T+1|T+1) = v + K_1 + K_2 - (T+1)C = S - (T+1)C$  and:

$$\sum_i F_i(t|T+1) = p \sum_{\tau=0}^{T-t} (1-p)^\tau [S - (\tau+t)C] + (1-p)^{T-t+1} [S - (T+1)C].$$

We can now express  $\sum_i F_i(t|T+1)$  in terms of  $\sum_i F_i(t|T)$  as follows:

$$\begin{aligned} \sum_i F_i(t|T+1) &= p \sum_{\tau=0}^{T-t-1} (1-p)^\tau [S - (\tau+t)C] \\ &\quad + p(1-p)^{T-t} [S - (T-t+t)C] + (1-p)^{T-t+1} [S - (T+1)C] \\ &= \sum_i F_i(t|T) - (1-p)^{T-t+1} C. \end{aligned}$$

Therefore,  $\sum_i F_i(t|T+1) - \sum_i F_i(t|T) = -(1-p)^{T-t+1} C < 0$ . Since we picked  $T$  arbitrarily, this establishes the claim.  $\square$

*Proof of Lemma 5.* We first show that, without loss of generality, players will not reallocate in equilibrium. If  $(nc, nc)$  achieves peace without reallocation, then players will not agree to reallocate because doing so necessarily leaves one of them worse off, whether or not such reallocation is peaceful. Hence, in equilibrium  $(nc, nc)$  may involve reallocation only when players cannot achieve peace unless they reallocate. So, suppose that  $(nc, nc)$  results in a fight in the continuation game. Choose the smallest such  $n$ ; that is, for all  $k_i < nc$ , they will not reallocate in equilibrium (we shall later see that  $n > 2$ ). So, unless they reallocate in equilibrium when  $k_i = nc$ , players will fight  $t \geq 1$  periods and stop fighting at  $(m_1c, m_2c)$ . Without loss of generality, assume  $m_1 \geq m_2$ . If  $t = n$ , then players must equalize resources at some point (otherwise one of them would collapse sooner). But the player who accepts the larger share initially would not agree to this. Therefore,  $t < n$  (that is they will not fight until they both collapse). If they fight until one of them collapses, i.e.  $m_2 < 1$ , then  $t < n$  implies player 2 would not agree to the initial reallocation because she can fight longer without reallocating. Therefore,  $m_2 \geq 1$  and players achieve peace when they are both still alive and have resources  $k_i = m_i c$  each. If  $m_1 = m_2$ , then they are indifferent between initial reallocation and no reallocation at all, hence our claim of no reallocation in equilibrium holds without loss of generality. So, assume that  $m_1 > m_2$ . Because each player can guarantee  $m_i c$  in that peaceful period by accepting an initial reallocation of  $(m_i + t)c$ , rejecting all offers, and fighting for  $t$  periods, it is a weakly dominant strategy for each player not to agree to anything less than  $(m_i + t)c$  in the initial reallocation. Hence, without loss of generality assume they reallocate initially such that each player has  $(m_i + t)c$ .

If player 2 disagrees with the initial reallocation, then our choice of  $n$  implies that they will fight for at least one period. If they achieve peace in the following period, then 2's share will be  $(n - 1)c$ . But since  $t \geq 1$ , it follows that  $n - 1 \geq m_1 > m_2$ , and therefore player 2 is better off disagreeing with that initial reallocation. This means that they must be fighting at least two periods. Because by our choice of  $n$  there is no reallocation on the equilibrium path in the continuation game, we have

$$V_i((n - 1)c, (n - 1)c) > V_i((m_1 + t - 1)c, (m_2 + t - 1)c) \quad (10)$$

for some  $i$ . Otherwise, players would agree to reallocate and continue with  $((m_1 + t - 1)c, (m_2 + t - 1)c)$ , contradicting our choice of  $n$ . But if  $i$  disagrees with the initial reallocation, his payoff is:

$$\begin{aligned} & p_i(n - 1)2c + (1 - p)V_i((n - 1)c, (n - 1)c) \\ & > p_i(n - 1)2c + (1 - p)V_i((m_1 + t - 1)c, (m_2 + t - 1)c) \\ & = V_i((m_1 + t)c, (m_2 + t)c), \end{aligned}$$

where the inequality follows from (10) and the equality follows from the fact that players fight for  $t \geq 1$  periods after the reallocation  $(m_i + t)c$ . In other words, player  $i$  does better by rejecting the reallocation, which contradicts our assumption that players reallocate in equilibrium. Hence, without loss of generality, players do not reallocate in equilibrium when they start out with equal resources.

Part (A). Because there is no reallocation in equilibrium, we can find the SPE by simple backward induction. First note that when  $n = 1$ , fighting destroys all resources and both players get a payoff of zero, which is worse than just agreeing to keep the current resources. Therefore, players achieve peace in this period. If players achieve peace in  $k_i = nc$ , then player  $i$  would sneak attack at  $k_i = (n + 1)c$  if, and only if,  $p(2(n + 1)c - 2c) + (1 - p)nc = (p + 1)nc > (n + 1)c$ , that is,  $pn > 1$ . This is necessary and sufficient when players do not redistribute, and since they do not redistribute in equilibrium, this establishes our claim.

To prove part (B), label the period where we want to see if players would fight as  $t = 1$ . If they fight in this period, by assumption they fight  $T$  battles, and then settle in  $t = T + 2$  on  $k_i = nc$ . We now have  $S = 2(n + T + 1)c$ , and so  $S - C = 2(n + T)c$ . Further,  $F_i(T + 1|T + 1) = p_i(2(n + 1)c - 2c) + (1 - p)nc = nc$ . Player  $i$  will sneak attack at  $t = 1$  if, and only if,

$$\begin{aligned} A_i &= 2p(n + T)c + (1 - p)F_i(2|T + 1) \\ &= 2p(n + T)c + (1 - p) \left[ p_i \sum_{s=0}^{T-2} (1 - p)^s [S_1 - (s + 2)C] + (1 - p)^{T-1} nc \right] \\ &= 2p(n + T)c + 2cp_i \sum_{s=1}^{T-1} (1 - p)^s (n + T - s) + (1 - p)^T nc \\ &> (n + T + 1)c = k_i, \end{aligned}$$

or, dividing through by  $c$ , noting that  $2p_i = p$ , and adding and subtracting  $p(1 - p)^T n$ ,

$$2p(n + T) + p \sum_{s=1}^T (1 - p)^s (n + T - s) + (1 - p)^{T+1} n > n + T + 1.$$

Noting now that:

$$\sum_{s=1}^T (1-p)^s (n+T-s) = (1-p) \left[ \frac{n(1-(1-p)^T)}{p} + \frac{pT-1+(1-p)^T}{p^2} \right],$$

we can further simplify that to condition (5) stated in the lemma. □

## References

- Blainey, Geoffrey. 1988. *The Causes of War*. 3rd ed. New York: The Free Press.
- Bueno de Mesquita, Bruce. 1981. *The War Trap*. New Haven: Yale University Press.
- Deutsch, Morton. 1973. *The Resolution of Conflict: Constructive and Destructive Processes*. New Haven: Yale University Press.
- Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (Summer): 379–414.
- Fearon, James D. 1996. "Bargaining Over Objects that Influence Future Bargaining Power." Paper presented at the annual meeting of the American Political Science Association, Washington, D.C., August 28-31.
- Fearon, James D. 2004. "Why Do Some Civil Wars Last So Much Longer Than Others?" *Journal of Peace Research* 41 (May): 275–301.
- Filson, Darren, and Suzanne Werner. 2004. "Bargaining and Fighting: The Impact of Regime Type on War Onset, Duration, and Outcomes." *American Journal of Political Science* 48 (April): 296–313.
- Gartner, Scott Sigmund. 1997. *Strategic Assessment in War*. New Haven: Yale University Press.
- Gartzke, Erik. 1999. "War Is in the Error Term." *International Organization* 53 (Summer): 567–87.
- Iklé, Fred Charles. 1971. *Every War Must End*. New York: Columbia University Press.
- Kecskemeti, Paul. 1958. *Strategic Surrender: The Politics of Victory and Defeat*. Stanford: Stanford University Press.
- Langlois, Catherine C., and Jean-Pierre P. Langlois. 2005. "Fully Informed and On the Road to Ruin: The Perfect Failure of Asymmetric Deterrence." *International Studies Quarterly* 49 (September): 503–527.
- Leininger, Wolfgang. 1989. "Escalation and Cooperation in Conflict Situations: The Dollar Auction Revisited." *Journal of Conflict Resolution* 33 (June): 231–254.
- O'Neill, Barry. 1986. "International Escalation and the Dollar Auction." *Journal of Conflict Resolution* 30 (March): 33–50.

- Organski, A. F. K. 1968. *World Politics*. 2nd ed. New York: Knopf.
- Organski, A. F. K., and Jacek Kugler. 1980. *The War Ledger*. Chicago: University of Chicago Press.
- Orme, John David. 2004. *Paradox of Peace: Leaders, Decisions, and Conflict Resolution*. New York: Palgrave Macmillan.
- Pillar, Paul R. 1983. *Negotiating Peace: War Termination as a Bargaining Process*. Princeton: Princeton University Press.
- Poundstone, William. 1992. *Prisoner's Dilemma*. New York: Doubleday.
- Powell, Robert. 1993. "Guns, Butter, and Anarchy." *American Political Science Review* 87 (March): 115–32.
- Powell, Robert. 1996. "Bargaining in the Shadow of Power." *Games and Economic Behavior* 15 (August): 255–89.
- Powell, Robert. 1999. *In the Shadow of Power*. Princeton: Princeton University Press.
- Powell, Robert. 2002. "Bargaining Theory and International Conflict." *Annual Review of Political Science* 5: 1–30.
- Powell, Robert. 2004a. "Bargaining and Learning While Fighting." *American Journal of Political Science* 48 (April): 344–361.
- Powell, Robert. 2004b. "The Inefficient Use of Power: Costly Conflict with Complete Information." *American Political Science Review* 98 (May): 231–241.
- Powell, Robert. 2005. "War as a Commitment Problem." *International Organization* .
- Reiter, Dan. 2003. "The Bargaining Model of War." *Perspectives on Politics* 1 (March): 27–43.
- Schelling, Thomas C. 1966. *Arms and Influence*. New Haven: Yale University Press.
- Shubik, Martin. 1971. "The Dollar Auction Game: A Paradox in Noncooperative Behavior and Escalation." *Journal of Conflict Resolution* 15 (March): 109–111.
- Slantchev, Branislav L. 2003a. "The Power to Hurt: Costly Conflict with Completely Informed States." *American Political Science Review* 47 (March): 123–33.
- Slantchev, Branislav L. 2003b. "The Principle of Convergence in Wartime Negotiations." *American Political Science Review* 47 (December): 621–32.
- Slantchev, Branislav L. 2004a. "Completeness and Coherence in the Bargaining Model of War." Manuscript, University of California, San Diego.
- Slantchev, Branislav L. 2004b. "How Initiators End Their Wars: The Duration of Warfare and the Terms of Peace." *American Journal of Political Science* 48 (October): 813–829.

- Slantchev, Branislav L. 2005a. "Military Coercion in Interstate Crises." *American Political Science Review* 99 (December).
- Slantchev, Branislav L. 2005b. "Territory and Commitment: The Concert of Europe as Self-Enforcing Equilibrium." *Security Studies* .
- Smith, Alastair, and Allan C. Stam. 2005. "Bargaining and the Nature of War." *Journal of Conflict Resolution* .
- Teger, Allan I. 1980. *Too Much Invested to Quit*. New York: Pergamon Press.
- von Clausewitz, Carl. 1832. *On War*. Princeton: Princeton University Press.
- Wagner, R. Harrison. 2000. "Bargaining and War." *American Journal of Political Science* 44 (July): 469–84.
- Young, Oran R. 1967. *Intermediaries: Third Parties in International Crises*. Princeton: Princeton University Press.
- Zartman, I. William. 1985. *Ripe for Resolution*. New York: Oxford University Press.
- Zartman, I. William. 2001. "The Timing of Peace Initiatives: Hurting Stalemates and Ripe Moments." *The Global Review of Ethnopolitics* 1 (September): 8–18.
- Zartman, I. William, and Maureen R. Berman, eds. 1982. *The Practical Negotiator*. New Haven: Yale University Press.