Intimidation: Linking Negotiation and Conflict

Sambuddha Ghosh, Gabriele Gratton, and Caixia Shen*

April 9, 2018

Abstract

A challenger wants a resource initially held by a defender, who can negotiate a settlement by offering to share the resource. If challenger rejects, conflict ensues. During conflict each player could be a tough type for whom fighting is costless. Therefore non-concession intimidates the opponent into conceding. Unlike in models where negotiations happen in the shadow of exogenously specified conflicts, the rejected offer determines how conflict is played if negotiations fail. In turn, how players are expected to play during conflict determines their negotiating positions. In equilibrium, negotiations always fail with positive probability, even if players face a high cost of conflict. Allowing multiple offers leads to brinkmanship—the only acceptable offer is the one made when conflict is imminent. If negotiations fail, conflict is prolonged and non-duration dependent.

Keywords: Intimidation, reputation, terrorism, negotiation, brinkmanship, costly war-of-attrition. **JEL Classification Numbers:** D74, D82.

*Ghosh: School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, China 200433, sghosh@mail.shufe.edu.cn. Gratton: School of Economics, UNSW Business School, UNSW Sydney, NSW 2052, Australia, g.gratton@unsw.edu.au. Shen: School of International Business Administration, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, China 200433, shencaixia@gmail.com. We are grateful to three anonymous referees, Masaki Aoyagi, Gaurav Aryal, Laurent Bouton, Alessandra Casella, Micael Castanheira, Costas Cavounidis, Jimmy Chan, Pauline Grosjean, Anton Kolotilin, Hongyi Li, Qingmin Liu, Youming Liu, Massimo Morelli, Santiago Oliveros, Juan Ortner, Carlos Pimienta, Ronny Razin, Santiago Sanchez-Pages, Balázs Szentes, John Tang, and Adam Wong. We thank seminar participants at the Australian National University, Boston University, Columbia University, European University Institute, Georgetown University, University of Montreal, the 2014 ASSA Meetings, the 2014 Econometric Society Australasian Meetings, the 2013 Australasian Economic Theory Workshop, and the 2014 SAET Conference. Barton Lee provided excellent research assistance.

1 Introduction

Interstate conflicts begin when negotiations end. But why do negotiations fail to prevent conflict even when compromise solutions are available, commitment is possible, and conflict is likely to be long and painful?[1] In the crisis bargaining literature, two parties negotiate in the shadow of an exogenously given conflict (Esteban and Sàkovics, 2008; Fearon 1992; 1994; Özyurt, 2014; Powell, 2004; Sechser, 2010). If the parties' costs of fighting the conflict are private information, then each party has an incentive to misrepresent its real cost so to build a "reputation." This process leads to prolonged crises and, possibly, to war. What drives the choices made during negotiations is the type of conflict the parties expect, its expected length, and the probability of ultimate victory of each party.

However, how conflict unfolds (its length and the probability of ultimate victory of each party) is also determined by past negotiations. In fact, by rejecting a generous ultimatum offer, one party may signal that its cost of fighting is low, which in turn makes its opponent more cautious once the conflict begins. Therefore, generous offers may backfire. In this paper we show that this damning effect of generous offers weakens the effectiveness of negotiations: parties may deliberately make offers which are rejected by their opponent with positive probability. For example, the Rambouillet Agreement offered by NATO to Yugoslavia before the onset of the Kosovo War were described as "a provocation, an excuse to start bombing"[2] that "deliberately set the bar higher than the Serbs could accept."[3]

To understand this two-way feedback relation between negotiation and conflict, we develop a model where negotiation and conflict are interlinked. In our model, Challenger (she) and Defender (he) want a resource that yields flow utility. Defender, who initially holds the resource, can try to negotiate peace by offering a share to Challenger. If all offers are rejected, conflict begins and Challenger has the option to repeatedly attack Defender to convince him to concede the resource. Our model of conflict draws from the literature on reputation building in bargaining (and war of attrition) (Abreu and Gul, 2000; Chatterjee and Samuelson, 1988; Ponsatì and Sàkovics, 1995). It is built on the concept of *intimidation*: repeatedly inflicting losses on one's opponent to scare him that further losses will be incurred unless a concession is made. But attacking could be costly too. What allows room for intimidation to function is that, in our model, conflict is a game of

[1]In a seminal paper, Fearon (1995) poses this as a "central puzzle" that rationalist explanations fail to solve (see also Reiter, 2003). The key word 'rationalist' rules out explanations where the parties to the conflict are entirely or largely irrational.

[2]Henry Kissinger, *Daily Telegraph*, 28 June 1999.

[3]George Kenney, "Rolling Thunder: the Rerun," *The Nation*, 27 May 1999.

two-sided incomplete information—with an arbitrarily small probability either party to the dispute is *tough*—does not experience the disutility of conflict. This small uncertainty is magnified by equilibrium play into a significant force that protracts conflict.

The same idea of intimidation also explains why negotiations may not succeed with certainty: even if Defender can make offers to Challenger before conflict begins, equilibrium offers are rejected with strictly positive probability by normal (i.e., not tough) Challenger. One might imagine that negotiations fail because Defender is afraid to reveal whether he is tough or normal. We show that this intuition is incomplete by focusing on the case of Defender being *uninformed*, in the sense of not knowing his cost of pursuing conflict when he makes an offer. Even in this case where offers do not reveal whether Defender is tough or normal, the opportunity to make an offer is a double-edged sword for Defender. On the one hand, higher offers are better for Defender because they have a higher probability of being accepted. Indeed, if beliefs were held fixed both before and after the negotiations, Defender could completely avoid conflict with normal Challenger by offering slightly more than her expected value of entering conflict. On the other hand, a generous offer that has a high probability of being accepted increases Challenger's expected payoff from conflict. Therefore, instead of deterring conflict, more generous offers encourage Challenger to seek conflict. We show that in equilibrium Defender always makes an offer that is both accepted and rejected with strictly positive probability by normal Challenger. Therefore, conflict begins with positive probability even if Challenger is normal. The key intuition can be best seen when considering an offer that Challenger should accept with certainty. Challenger knows that, if she rejects this offer, Defender would conclude that Challenger is tough for sure. Therefore, if conflict begins, Defender would concede at the first opportunity.

This detrimental effect of generous offers is particularly evident when Defender has multiple chances to make offers. In equilibrium, all offers, except the one made in the last round, must be unacceptable (no Challenger would ever accept them). Indeed, suppose that before the last round Defender makes an offer that normal Challenger accepts with positive probability. Then Challenger strictly prefers to reject it, as this signals she is tough, thus intimidating Defender into making an even more generous offer in the last round. In other words, offers that could be successful (but weren't) make subsequent offers more costly. Therefore, long negotiations resolve in *brinkmanship*: the parties make no progress towards a peaceful solution up until the last opportunity to negotiate.

This brinkmanship result provides an explanation for negotiation in the shadow of conflict not making any progress until the very last chance. A possible example are the negotiations leading up to the Treaty of Porthsmouth in the Russo-Japanese War (1904–1905) (which subsequently gained President Theodore Roosevelt the Nobel Peace Prize). Negotiations for the treaty were held while Russia brought further troops to Manchuria, a move that would have given Russia an advantage in case of conflict. Fredrik (2010) notes how the Japanese delegation demanded the southern part of the island of Sakhalin *and* war reparations throughout the negotiations. Only upon the arrival of further four Russian divisions, at what was conceivably their last chance to do so, did the Japanese drop their claim for reparations and avert conflict.

While in our benchmark model negotiations can only happen before conflict begins, in reality the parties to a conflict may wish to negotiate also once conflict has begun. In Section 7 we allow Defender to make offers in any period (and therefore after being informed about whether he is tough). We show that our main result holds even in this variation of the model: negotiations fail to avoid conflict with certainty.

Our framework also allows us to explicitly derive the probability that conflict extends to the next chance of attacking. Crucially, the probability that conflict extends to the next period does not depend either on the probabilities that the players are tough or for how long the conflict has already been fought. Thus, our model predicts that conflict is non-duration dependent: the past duration of conflict does not predict its probability of ending. From a theoretical perspective, our result that conflict is non-duration dependent is equivalent to the constant concession rate in reputational models of bargaining (Abreu and Gul, 2000) and international crises in the shadow of conflict (Özyurt, 2014). In our model, armed conflict is itself a reputational game and therefore the parties capitulate at a constant rate.[4] Empirically, non-duration dependence characterizes both interstate (Bennet and Stam, 1996) and civil (Collier, Hoeffler, and Soderbom, 2004) conflicts.

Our main contribution is to provide an integrated model of negotiation and conflict. Without such a model it would be impossible to study the two-way feedback relation by which negotiations affect conflict and conflict affects negotiations. Essentially, we connect two literatures: the crisis bargaining literature and the bargaining and reputation literature. In the crisis bargaining literature, conflict is an exogenously given outside option for the negotiating parties. Once conflict begins, the parties' relative military strengths determine the final outcome. This literature focuses on explaining why parties would delay reaching a settlement and ultimately reach conflict with positive probability. Since Gul, Sonnenschein, and Wilson (1986) clarified that private information alone cannot lead to

[4] In contrast, Reiter (2003) notes that most existing models of conflict based on informational asymmetries fail to capture that conflict is in reality non-duration dependent, as these asymmetries should fade in time.

delays in bargaining, a large literature has arisen. Fearon (1994) and Özyurt (2014) model "audience costs": further delays increase the cost of conceding to one's opponent. Therefore, delaying helps one commit to fighting. In our model, delay happens in spite of this motive being absent. Furthermore, in contrast with our result, flexible or strategic types in Özyurt (2014) always concede before war begins.

A different source of delay in bargaining is the one driven by reputation *à la* Kreps and Wilson (1982) and Milgrom and Roberts (1982).[5] Acharya and Grillo (2015) explicitly model this reputational motive in a model of crisis bargaining with irrational types. The option to engage in total war is never exercised in equilibrium, unlike in our model. Nevertheless, the same reputational motives drive the conflict part of our model. An approach similar to ours is the one by Lapan and Sandler (1988), who model terrorism as a repeated game between players who are tough with some probability. In their model, absent a concession, the public belief that a player is tough jumps up to an arbitrary and exogenously given quantity. Hodler and Rohner (2012) make this endogenous, but they have only two periods, which in turn means that they predict attacks only when the probability of the terrorist being tough is very large. Our model endogenously determines both the termination of the war and the evolution of beliefs about the degree of irrationality of one's opponent, and shows that prolonged conflict is compatible with very small degrees of irrationality.

Closer to us, Heifetz and Sagev (2005) study a model of negotiation during war. In their model, conflict is already happening, but one of the players, the aggressor, can choose to escalate—essentially increasing the flow cost of conflict. They show that there exist conditions such that, restricting attention to separating equilibria, the aggressor chooses to escalate. This is because, if escalation is more costly to the defender than to the aggressor, escalation both shortens conflict and induces further negotiations more in the aggressor's favor. The choice to escalate is similar to our choice to go to conflict, but in our model there is a unique equilibrium and the parties go to conflict with strictly positive probability even if this decision limits their ability to negotiate further (and for any relative cost of conflict for Challenger and Defender).

Our logic of brinkmanship shares a common feature with Sechser (2010) who shows that if conflict is potentially repeated, a player may incur the cost of rejecting an offer to avoid revealing its weakness in view of future negotiations. In Brito and Intriligator (1985) time is needed to screen among various types of opponents; in Sánchez-Páges (2009) time is needed to convey credible information to an uninformed party about the eventual out-

[5] Delay also arises in models with a non-common prior so that each player could be overly optimistic about her chances of being selected as the proposer (Yildiz 2004; Bénabou and Tirole, 2009).

come of rejecting agreements and triggering conflict. In contrast to these three papers, in our model brinkmanship arises because the uninformed party chooses to avoid giving the informed party a chance to signal she is tough.

Schelling (1956, 1960, 1966)[6] first developed the idea that bargaining parties can benefit if they convince their opponent that they are committed to their threat—hence the argument that governments should appear committed to hawkish positions when facing a terrorist threat. But in our model, as well as in Abreu and Gul (2000) and Özyurt (2014), once conflict begins, the expected payoff for (normal) Defender is independent of his probability of being tough. In fact, the entire advantage of being perceived as tough comes from the ability to induce a normal Challenger to attack with very low probability. But if the Challenger attacks nonetheless, then Defender must update his beliefs to assign a very high probability to Challenger being tough.

Our idea of intimidation is also related to Silverman (2004), a random-matching model where violence is instrumental in deterring future violence against oneself. If the fraction of agents who directly gain from violence is sufficiently large, then other agents can also engage in it to acquire a reputation for toughness. Yared (2009) considers a defender with private knowledge of his cost of conceding the flow resource in each period; in equilibrium the challenger attacks with positive probability when no concession is made, so that the defender has an incentive to concede often enough. Since costs are drawn independently across periods, there is no reputation at play, unlike in our model.

2 Benchmark model

In our benchmark model, the parties have a single chance to reach an agreement before conflict begins. In Section 5 we discuss how our results generalize to multiple rounds of negotiation.

Time is continuous and indexed by $\tau \geq 0$. There are two players: Challenger and Defender. Both players discount future payoffs with rate $r > 0$. They contest a resource, which is initially held by Defender. Holding the resource gives a flow rent normalized to 1.

The game played by Challenger and Defender is best described by dividing it into two phases: negotiation and conflict.

[6]See Crawford (1982) for a formal treatment of this idea.

Negotiation. At time $\tau = 0$, Defender can offer a fraction x of the resource to Challenger. Upon observing the offer, Challenger decides to accept or reject the offer. If Challenger accepts the offer, then the game ends and Challenger and Defender enjoy flow rents x and $(1 - x)$, respectively, thereafter. Otherwise, the game immediately enters the conflict phase.

Conflict. In the conflict phase, the following two-stage game is played out at each time $\tau_t = (t-1)\,\Delta$, $t \in \{1, 2, \ldots\}$.

Stage 1: Challenger chooses whether to attack or concede;

Stage 2: Defender chooses whether to resist or concede.

We refer as the interval of time from τ_t to τ_{t+1} as period t. Notice that there is no time interval between the two stages.

As soon as one party concedes, the other party gets to enjoy the entire resource in that period and forever afterwards; thus conflict is less flexible than negotiation.

Types and payoffs. Each player can be of two types: *tough* or *normal*. Challenger (Defender) is tough with non-zero probability π^C (π^D). Normal types dislike conflict in the sense that in each period in which Challenger attacks, normal Challenger and Defender suffer losses

$$L^C \equiv \left(1 - e^{-r\Delta}\right)\ell^C, \text{ with } \ell^C > 0$$
$$L^D \equiv \left(1 - e^{-r\Delta}\right)\ell^D, \text{ with } \ell^D > 0$$

respectively. Therefore, in any period in which neither Challenger nor Defender concede, normal Challenger's payoff is given by $-L^C$, while normal Defender's payoff is given by $V - L^D$, where

$$V \equiv 1 - e^{-r\Delta}$$

is the value of holding the resource for one period. As soon as Defender (Challenger) concedes, Challenger (Defender) enjoys a payoff equal to V from that period onward. Notice that the cost of the attack is suffered by both players whenever Challenger attacks: if Challenger attacks and Defender concedes, normal Challenger's payoff is $V - L^C$ and normal Defender's payoff is $-L^D$.

Instead, both tough Challenger and tough Defender suffer no loss from conflict. Tough Challenger does not accept any offer short of $x = 1$ and attacks until Defender concedes. Tough Defender never makes an offer greater than $x = 0$ and never concedes.

Challenger privately observes her type at the beginning of the game. Our aim is to show that negotiation fail even if Defender has no incentive to conceal his type. Therefore, in our benchmark model we assume that Defender privately observes his type only if and when conflict begins. While we do this because it better illustrates the logic behind our main results, we also note that this may be a realistic assumption in many conflicts. For example, when Challenger attacks using new technologies that have yet to be proven, both players may be uncertain about Challenger's ability to overcome Defender's defenses and cause him harm. In a different example, Defender may be a democratic government who does not know a priori whether the electorate will be able to withstand Challenger's attacks. In both cases, Defender will be able to evaluate the losses caused by Challenger's first attack and therefore discover his type at the beginning of period 1. The timing at which players privately observe their types is common knowledge.

Obviously, if L^C is too large, then normal Challenger would never carry out an attack, even if Defender concedes for sure in period 1. Also, if L^D is too small, then normal Defender would never concede, even if Challenger attacks for sure at every period $t = 1, 2, \ldots$. We are interested in those cases in which it is at least conceivable for normal Challenger to attack or normal Defender to concede. Therefore, we restrict our attention to those conflicts in which L^C is sufficiently small and L^D is sufficiently large.

Assumption 1. *Let $\delta \equiv e^{-r\Delta}$ be the discount factor between periods. We assume $\delta L^D > V$; $L^C < V(1-\delta)^{-1}$.*

Our solution concept is perfect Bayesian equilibrium (henceforth *equilibrium*). In what follows we derive the entire equilibrium of the game when attacks are frequent, i.e., when Δ is sufficiently small. Nonetheless, we show the complete solution of the conflict phase for any $\Delta > 0$, as this allows us to derive precise comparative statics on the probability that conflict ends in any period t.

2.1 Strategies and public beliefs

At each period t of conflict, the complete history is summarized by the vector of public beliefs. In stage 1, the state of the game is the vector $\left(\pi^C_{t-1}, \pi^D_{t-1}\right)$, where π^C_{t-1} is Defender's belief that Challenger is tough, and π^D_{t-1} is Challenger's beliefs that Defender is tough. In stage 2, the state vector is $\left(\pi_t^C, \pi_{t-1}^D\right)$, where Defender's belief about Challenger's type has been updated from π_{t-1}^C to π_t^C in light of Challenger's action at stage 1. Notice that $\left(\pi_0^C, \pi_0^D\right)$, the public beliefs with which conflict begins, is not necessarily equal to $\left(\pi^C, \pi^D\right)$. In fact, actions during negotiation may affect the post-negotiation public belief π_0^C that Challenger is tough.

Negotiation. In the negotiation phase, Defender's strategy is an offer in $[0, 1]$. Normal Challenger's strategy maps from the current beliefs and the offer to a probability of accepting the offer in $[0, 1]$.

Conflict. In the conflict phase, a (behavior) strategy for normal Challenger is a sequence of mappings $\sigma_t^C : [0, 1]^2 \to [0, 1]$, $t \in \mathbb{N}$, where $\sigma_t^C\left(\pi_{t-1}^C, \pi_{t-1}^D\right)$ is the probability that normal Challenger concedes in period t as a function of the public beliefs. A strategy for normal Defender is a sequence of mappings $\sigma_t^D : [0, 1]^2 \to [0, 1]$, one for each $t \in \mathbb{N}$, where $\sigma_t^D\left(\pi_t^C, \pi_{t-1}^D\right)$ is the probability that normal Defender concedes in period t as a function of the public beliefs.[7,8] Notice that σ_t^C and σ_t^D are conditional probabilities of concession. I.e., they are probabilities with which players concede in period t, conditional on no previous concession.

Since tough players never concede, the *average* probabilities of concession by Challenger and Defender respectively are obtained by multiplying the probability of the normal type by the probability that the (respective) normal type concedes:

$$\begin{aligned} \bar{\sigma}_t^C\left(\pi_{t-1}^C, \pi_{t-1}^D\right) &= \left(1 - \pi_{t-1}^C\right)\sigma_t^C\left(\pi_{t-1}^C, \pi_{t-1}^D\right); \\ \bar{\sigma}_t^D\left(\pi_t^C, \pi_{t-1}^D\right) &= \left(1 - \pi_{t-1}^D\right)\sigma_t^D\left(\pi_t^C, \pi_{t-1}^D\right). \end{aligned} \tag{1}$$

Obviously, $\pi_t^C = 0$ at any history where Challenger has conceded. If Challenger does not concede before or at period t, the updated belief π_t^C that Challenger is tough is recursively

[7] Note that Challenger's action in period t depends on the beliefs at the end of period $t-1$, as is standard. In contrast, Defender observes Challenger's move at t, updates her belief about Challenger's type to π_t^C, and only then chooses an action.

[8] We define the players' strategies as functions of the current public beliefs instead of the entire history of play. This is without loss of strategic flexibility. In particular, at any period t in which the players can move, the history of play is (i) Defender made an offer x, (ii) Challenger rejected the offer, (iii) no player has conceded in any period $t' < t$. Thus, the only loss of information is given by the entity of the offer x. As we shall show, at any period t, the optimal strategy for each player is uniquely pinned down by the expected payoff of continuing the conflict, which in turn is uniquely determined by the current public beliefs. That is, offers that were made during negotiations affect the players' expected payoffs only because they change the public belief that Challenger is tough if rejected.

derived by Bayes' rule from π^C_{t-1} and σ^C_t:

$$\pi^C_t = \frac{\pi^C_{t-1}}{1-\bar{\sigma}^C_t\left(\pi^C_{t-1}, \pi^D_{t-1}\right)}. \tag{2}$$

Similarly, $\pi^D_t = 0$ at any history where Defender has conceded. If Defender does not concede before or at period t, the updated belief π^D_t that Defender is tough is recursively derived by Bayes' rule from π^D_{t-1} and σ^D_t:

$$\pi^D_t = \frac{\pi^D_{t-1}}{1-\bar{\sigma}^D_t\left(\pi^C_t, \pi^D_{t-1}\right)}. \tag{3}$$

3 Negotiation failure and conflict

In this section, we preview our main results. Before doing so, it is convenient to define two threshold values. In our model, each normal player concedes if he or she believes the enemy to be tough with probability 1. Since payoffs are continuous in beliefs, so are the optimal strategies. Therefore, in any period t, if Defender believes Challenger to be tough with sufficiently high probability, then he will concede immediately, even if he knows that normal Challenger will concede in stage 1 of period $t+1$ (at her next chance to concede). Similarly, if Challenger believes Defender to be tough with sufficiently high probability, then she will prefer to concede immediately, even if she knows that normal Defender will concede with certainty in stage 2 of period t itself (at his next chance to concede). The following lemma finds these thresholds exactly.

Lemma 1 (Threshold beliefs)**.** *Let $\bar{\pi}^C$ and $\bar{\pi}^D$ be given by*

$$\bar{\pi}^C := \frac{1}{\delta}\left[1+L^D\right]^{-1}, \text{ and} \tag{4}$$

$$\bar{\pi}^D := 1-L^C. \tag{5}$$

In any equilibrium,

(i) *Assume that normal Defender concedes immediately after Challenger attacks in period $t+1$. Then normal Challenger strictly prefers to attack in period $t+1$ if $\pi^D_t < \bar{\pi}^D$, is just indifferent at $\bar{\pi}^D$, and strictly prefers to concede otherwise;*

(ii) *Assume that normal Challenger concedes in period $t+1$. Then normal Defender strictly prefers to resist in period t if $\pi^C_t < \bar{\pi}^C$, is just indifferent at $\bar{\pi}^C$, and strictly prefers to concede otherwise.*

Proof. All proofs are in Appendix. □

It is useful to notice that $\bar{\pi}^C$ and $\bar{\pi}^D$ do not depend on either the players' beliefs or the period t.

We can now describe our main results.

Negotiations fail. In the equilibrium, Defender makes an offer that is both accepted and rejected by normal Challenger with strictly positive probability. Therefore, conflict begins with probability strictly greater than π^C, but strictly smaller than 1.

Even when negotiations fail, they are not without consequences. In fact, the strategies played during negotiation affect the beliefs $\left(\pi_0^C, \pi^D\right)$ with which conflict begins, in turn affecting how conflict is played. In particular, let β be the equilibrium probability that negotiations succeed (i.e., the total probability that Challenger accepts the equilibrium offer), then

$$\pi_0^C = \frac{\pi^C}{1-\beta}.$$

Therefore, a higher probability of success leads, in case of failure, to a higher belief that Challenger is tough.

Conflict. After negotiations fail, conflict begins with post-negotiation beliefs $\left(\pi_0^C, \pi^D\right)$. In equilibrium, these beliefs are such that there exists $n \in \mathbb{N}$ with $\left(\bar{\pi}^C\right)^{n+1} \leq \pi_0^C < \left(\bar{\pi}^C\right)^n$ and $\pi^D < \left(\bar{\pi}^D\right)^{n+1}$. We call n the conflict *horizon*.

Definition 1. A conflict has horizon n if n is the largest non-negative integer such that $\pi_0^C < \left(\bar{\pi}^C\right)^n$ and $\pi^D < \left(\bar{\pi}^D\right)^n$.

Along the equilibrium path, in period 1 Challenger attacks with probability 1 and Defender resists this first attack with total (i.e., averaged over normal and tough Defender) probability

$$\frac{\pi^D}{\left(\bar{\pi}^D\right)^n}.$$

From period 2 onward, Challenger attacks with probability $\bar{\pi}^C$ and Defender resists with probability $\bar{\pi}^D$, as long either $\pi_t^C \leq \bar{\pi}^C$ or $\pi_t^D \leq \bar{\pi}^D$. Therefore, from period 2 and until period n, public beliefs evolve in time according to

$$\begin{cases} \pi_t^D = \frac{\pi_{t-1}^D}{\bar{\pi}^D}; \\ \pi_t^C = \frac{\pi_{t-1}^C}{\bar{\pi}^C}. \end{cases} \tag{6}$$

That is, after each attack, Defender updates his belief that Challenger is tough upwards. Similarly, after each time Defender resists, Challenger updates her belief that Defender is tough upwards. This process of intimidation terminates after n periods. Indeed, following the law of motion in (6), after Challenger attacks in period $n+1$, $\pi^C_{n+1} > \bar{\pi}^C$. Therefore, in period $n+1$, if Challanger attacks, Defender concedes with probability 1. Thereafter, both normal types concede with probability 1. Thus conflict goes beyond $n+1$ periods (indeed, forever) if and only if both players are tough.

The horizon n naturally relates to the post-negotiation beliefs that the players are tough (Abreu and Gul, 2000). When both players are believed to be normal with very high probability, then conflict has a longer horizon, as it takes a long time to intimidate one's opponent. Conversely, if at least one player is believed to be tough with high probability, then conflict has a shorter horizon, as few periods of conflict are sufficient to intimidate one's opponent.

We further note that the hazard rate of conflict—that is, the probability that conflict ends in period $t > 1$, conditional on not having ended before—equals $\bar{\pi}^C\bar{\pi}^D$ from period 2 to n. Since neither threshold depends on the public beliefs or on the period t, this implies that conflict in our model is non-duration dependent.

In the following sections, we develop a more precise analysis of the equilibrium play and offer some useful comparative statics. For ease of exposition, we proceed backwards, beginning with a discussion of how conflict is played.

4 Equilibrium in conflict

Propositions 1 and 2 below characterize the unique equilibrium of the conflict phase, conditional on the post-negotiation public beliefs $\left(\pi^C_0, \pi^D_0\right) = \left(\pi^C_0, \pi^D\right)$.[9] We give a complete characterization of the equilibrium for any $\Delta > 0$, but under an additional assumption. We show in Appendix A.5 that when this assumption fails the set of equilibria can be identified using Propositions 1 and 2. As Δ becomes small, the expected payoff of all these equilibria converge to the expected payoff of the equilibrium in Proposition 1. Therefore, as Δ becomes small, this assumption plays no role in the determination of the expected payoffs in the conflict game and the unique equilibrium of the whole model.

Assumption 2. *The quantities* $\ln \pi^D / \ln \bar{\pi}^D$ *and* $\ln \pi^C_0 / \ln \bar{\pi}^C$ *are not integers.*

[9] In a model where negotiations are not possible, conflict begins with public beliefs equal to the prior beliefs $\left(\pi^C, \pi^D\right)$. Since in our model negotiations only affect the public belief that Challenger is tough, then the post-negotiation belief π^D_0 equals π^D.

The continuation equilibrium once negotiations have failed is unique, but it can be of one of two types, depending on the post-negotiation public beliefs π_0^C and π^D. More precisely, what matters is how these beliefs compare to their respective thresholds $\bar{\pi}^C$ and $\bar{\pi}^D$. That is, what matters is whether Challenger or Defender are perceived to be more or less tough, compared to what would be needed to induce the other player to concede immediately. If π_0^C is sufficiently closer to $\bar{\pi}^C$ then π^D is to $\bar{\pi}^D$ (Proposition 1), then we say that Challenger is more *intimidating*. Otherwise (Proposition 2), we say that Defender is more intimidating. As we shall see, when negotiations fail, then conflict always begins with Challenger being more intimidating. (Notice that Points 2 and 3 in Propositions 1 and 2 are identical.)

Proposition 1 (When Challenger is more intimidating than Defender.)**.** *Let* $\left(\bar{\pi}^C\right)^{n+1} < \pi_0^C < \left(\bar{\pi}^C\right)^n$ *and* $\pi^D < \left(\bar{\pi}^D\right)^{n+1}$ *for some* $n \in \mathbb{N}$*. In the unique equilibrium*

1. *in period* 1*, Challenger attacks with probability* 1 *and Defender concedes with total probability*

$$1 - \frac{\pi^D}{\left(\bar{\pi}^D\right)^n}; \tag{7}$$

2. *subsequently, as long as both* $\pi_t^C \leq \bar{\pi}^C$ *and* $\pi_t^D \leq \bar{\pi}^D$*, Challenger and Defender concede with total probabilities* $1-\bar{\pi}^C$ *and* $1-\bar{\pi}^D$ *respectively, and the public beliefs* $\left(\pi_t^C, \pi_t^D\right)$ *evolve according to (6);*

3. *once* $\pi_t^D > \bar{\pi}^D$ *(*$\pi_t^C > \bar{\pi}^C$*), normal Challenger (Defender) concedes with probability* 1*.*

Proposition 2 (When Defender is more intimidating than Challenger.)**.** *Let* $\left(\bar{\pi}^D\right)^{n+1} < \pi^D < \left(\bar{\pi}^D\right)^n$ *and* $\pi_0^C < \left(\bar{\pi}^C\right)^n$ *for some* $n \in \mathbb{N}$*. In the unique equilibrium*

1. *in period* 1*, Challenger concedes with total probability*

$$1 - \frac{\pi_0^C}{\left(\bar{\pi}^C\right)^n}; \tag{8}$$

2. *subsequently, as long as both* $\pi_t^C \leq \bar{\pi}^C$ *and* $\pi_t^D \leq \bar{\pi}^D$*, Challenger and Defender concede with total probabilities* $1-\bar{\pi}^C$ *and* $1-\bar{\pi}^D$ *respectively, and the public beliefs* $\left(\pi_t^C, \pi_t^D\right)$ *evolve according to (6);*

3. *once* $\pi_t^D > \bar{\pi}^D$ *(*$\pi_t^C > \bar{\pi}^C$*), normal Challenger (Defender) concedes with probability* 1*.*

We give the full proof of Propositions 1 and 2 and further details about its intuition in Appendix A.2. Here we offer only the basic thrust behind its logic necessary to understand how negotiation and conflict are linked.

First, from period 1, stage 2 onward, both players concede with strictly positive probability. Challenger's (Defender's) probability of concession is such that Defender (Challenger) is indifferent between conceding and resisting (attacking). The fact that normal players are expected to concede with positive probability is what fuels intimidation. Upon observing an attack, Defender must conclude that Challenger is more likely to be tough than he previously believed. Similarly, when Defender does not concede, Challenger must conclude that Defender is more tough than she previously believed. In particular, since the conditional probabilities of concession are constant from period 1, stage 2 onward, public beliefs move according to (6). As conflict continues, these beliefs get closer and closer to the threshold beliefs $\bar{\pi}^C$ and $\bar{\pi}^D$, eventually passing them after a finite number of periods.

Second, once one of the public beliefs passes its threshold, conflict must end if at least one of the players is normal. Lemma 4 in Appendix A.2 says that if both beliefs are strictly below their threshold, no belief leaps over the corresponding threshold at the next step without touching the corresponding threshold exactly. Intuitively, suppose that $\pi_{t-1}^C < \bar{\pi}^C$ and that Challenger knows that attacking in period t will induce a public belief π_t^C above the threshold $\bar{\pi}^C$. Then, by Lemma ,1, she will strictly prefer to attack unless $\pi_{t-1}^D = \bar{\pi}^D$. But if he attacks with probability 1, then $\pi_t^C = \pi_{t-1}^C < \bar{\pi}^C$. The implication is that when a normal player concedes with probability 1 for the first time, the public belief that he or she is tough is exactly equal to its threshold value. In particular, when Challenger is more intimidating, we have $\pi_n^D = \bar{\pi}^D$; when Defender is more intimidating, we have $\pi_n^C = \bar{\pi}^C$. Thereafter, both normal players concede with probability 1. Therefore, the horizon of conflict n is naturally linked to the maximum number of periods for which conflict protracts between normal players.

Third, one key difference between the two cases in Propositions 1 and 2 is that when Challenger is more intimidating, she attacks with probability 1 at the start, whereas she mixes when Defender is more intimidating. That is, when Challenger is perceived to be more likely to be tough, her expected equilibrium payoff from conflict is greater (see Abreu and Gul, 2000; and Özyurt, 2014 for a continuous time version of this result).

Remark 1. If at the beginning of conflict Challenger is more intimidating than Defender, then normal Challenger's expected payoff $u^C\left(\pi_0^C,\pi^D\right)$ is given by

$$u^C\left(\pi_0^C,\pi^D\right)=\left(1-\frac{\pi^D}{\left(\bar{\pi}^D\right)^n}\right)-L^C \tag{9}$$

and normal Defender's expected payoff is $-L^D$. The unconditional probability of an attack in period $t\in\{2,\ldots,n\}$ is given by

$$\Pr\left(\text{attack at } t\right)=\frac{\pi^D}{\left(\bar{\pi}^D\right)^n}\left(\bar{\pi}^C\right)^{t-1}\left(\bar{\pi}^D\right)^{t-2}.$$

Remark 2. If at the beginning of conflict Defender is more intimidating than Challenger, then normal Challenger's expected payoff is 0 and normal Defender's expected payoff $u^D\left(\pi_0^C,\pi^D\right)$ is given by

$$u^D\left(\pi_0^C,\pi^D\right)=\left(1-\frac{\pi_0^C}{\left(\bar{\pi}^C\right)^n}\right)-\frac{\pi_0^C}{\left(\bar{\pi}^C\right)^n}L^D. \tag{10}$$

The unconditional probability of an attack in period $t\in\{1,\ldots,n+1\}$ is given by

$$\Pr\left(\text{attack at } t\right)=\frac{\pi_0^C}{\left(\bar{\pi}^C\right)^n}\left(\bar{\pi}^C\bar{\pi}^D\right)^{t-1}.$$

We now provide some comparative statics regarding the length of conflict and link them to some empirical regularities.

As in Abreu and Gul (2000), the conflict horizon is finite and it is shorter when players are believed to be tough with greater probability.

Corollary 1. *Unless both players are tough, the maximum length of a conflict is determined by the conflict horizon n. If Challenger is more intimidating, normal Challenger does not attack after period $n+1$. If Defender is more intimidating, normal Challenger does not attack after period n.*

The next result determines the probability that a conflict that has lasted until period t, $1<t<n-1$ survives to period $t+1$. This probability is independent of which player is more intimidating, how much the players are likely to be tough, or the period t. That is, the hazard rate of the conflict depends only on the threshold values $\bar{\pi}^C$ and $\bar{\pi}^D$, which do not depend on the priors π_0^C and π_0^D or on t. Thus, until period $t=n$ is reached, conflict is non-duration dependent.

Corollary 2. *In each period $t:1<t<n$, if conflict has not yet ended, Challenger attacks with constant probability $\bar{\pi}^C$ and Defender resists with constant probability $\bar{\pi}^D$. Therefore, if conflict*

has not yet ended by the end of period $t-1$, the probability that conflict will not end by the end of period t equals $\bar{\pi}^C \bar{\pi}^D$.

The non-duration dependence of conflict is driven by the reputational nature of our model of conflict. In fact, as we noted in the introduction, in models of bargaining with reputation (e.g., Abreu and Gul, 2000; Özyurt, 2014), the conditional probability that bargaining continues is constant in time. Notice that as π_0^C and π^D become small, n grows and the length of conflict (conditional on there being a conflict at the end of period 1) is therefore approximated by a geometric distribution with hazard rate $\bar{\pi}^C \bar{\pi}^D$.

Although the probability that conflict extends to the next period does not depend on time, it nonetheless depend in intuitive ways on other primitives of the model, in particular on the players' cost of fighting. Thus conflict duration is determined by the opponents' military capabilities, rather than on their intimidation abilities.

Corollary 3. *Conditional on there being a conflict at time $t > 1$, the probability of an attack in period $t' > t$ is decreasing in the cost of fighting (L^C and L^D).*

For π_0^C and π^D sufficiently small, such that conflict length is approximated by a geometric distribution, the same comparative statics apply to the expected length of conflict. Furthermore, our normalization of the value of the resource to 1 does not allow us to derive explicit comparative statics with respect to it. Nonetheless, it is easy to show that if the value of the resource is given by V', then the probability of an attack in period t' is increasing in V'.

The evidence concerning the effect of institutional characteristics on the duration of conflict is ambiguous. Bennet and Stam (2009), Langlois and Langlois (2009), and Henderson and Bayer (2013) find that the relation between democracy and conflict duration is not significant once the military capabilities of the parties and the physical characteristics of conflict (common boundaries, terrain, etc.) are taken into account. Our results suggest that that the probability of continuation of conflict depends indeed only on the costs and benefits of war, and only to a lesser extent on the probability of being tough. Thus, in our model as in the real world, physical and technological characteristics matter more than political ones.

We now turn to the question of when an armed conflict is more likely to begin, i.e. there is a first attack. The following corollary describes how this probability depends to the public beliefs that the players are tough.

Corollary 4. *Fix the likelihood π^D that Defender is tough. The probability that Challenger begins* *to attack is increasing in the post-negotiation belief that Challenger is tough π_0^C. It is* strictly *increasing if and only if Defender is more intimidating than Challenger.*

For Defender, an image of toughness can pay: if Defender is more intimidating than Challenger, then the probability of a first attack is strictly less than 1. In this case, the probability of a conflict is $\pi_0^C / \left(\bar{\pi}^C\right)^n$, where n is the largest natural number such that $\pi^D \leq \left(\bar{\pi}^D\right)^n$. Thus, if π^D increases, the probability of a first attack decreases.

Corollary 5. *Let Defender be more intimidating than Challenger. Then, the probability that Challenger begins to attack is decreasing in π^D.*

Nonetheless, the advantage of being perceived as tough should not be overstated. After the first attack is carried out, Challenger levels the playing field with Defender and the expected payoff for Defender is $-L^C$, independently of π^D. Indeed, in equilibrium, Defender is indifferent between conceding and resisting whenever he plays.

One of the few empirical regularities about conflict is that pairs of democracies are less likely to fight each other.[10] Our results suggest that conflict begins when there is greater imbalance between the parties' probability of being tough. We argue that democratic leaders are kept in check by their citizens and that therefore democracies tend to have similar probabilities of being tough or irrational. On the contrary, autocrats of the like of Kim Jung-un are known for their unpredictable behavior. Thus, a pair of autocracies or one democracy and one autocracy are more likely to have unbalanced probabilities of being tough and are therefore more likely to engage in conflict.

5 Why negotiations fail to eliminate conflict

We now turn to negotiations and show why in our model conflict may not be avoided even if Defender and Challenger can commit to a peaceful division of the resource. An important feature of our model is that conflict and negotiation are linked: when choosing an action during negotiations, each player knows that her action will affect public beliefs at the onset of conflict. In turn, these beliefs determine the expected payoff from conflict, affecting the relative appeal of negotiating peace. In particular, it is useful to remember that our analysis of the equilibrium of the conflict phase shows that conflict can be of two types. First, by Remark 1, when at the beginning of conflict (and therefore after negotiations have failed) public beliefs are such that $\left(\bar{\pi}^C\right)^{n+1} \leq \pi_0^C < \left(\bar{\pi}^C\right)^n$ and $\pi^D < \left(\bar{\pi}^D\right)^{n+1}$

[10] See Bueno de Mesquita and Smith (2012) for a survey.

for some $n \in \mathbb{N}$, then Challenger's expected payoff from conflict is strictly positive. We labeled this situation as Challenger being more intimidating than Defender. Otherwise, by Remark 2, Challenger's expected payoff from conflict is 0.

We now show that in equilibrium Defender makes a strictly positive offer that normal Challenger then rejects with strictly positive probability. Obviously, normal Challenger would not reject a strictly positive offer unless her expected payoff from conflict is also strictly positive. Therefore, in equilibrium rejecting an offer must induce a belief π_0^C such that Challenger is more intimidating than Defender.

Proposition 3. *For Δ sufficiently small, in equilibrium Defender makes an offer that is accepted with strictly positive probability, but that normal Challenger rejects with strictly positive probability. Therefore, the probability that negotiations fail to avoid conflict is strictly greater than $1-\pi^C$. Furthermore, conflict always begins with Challenger being more intimidating than Defender, i.e., along the equilibrium path, there exists n such that $\left(\bar{\pi}^C\right)^{n+1} \leq \pi_0^C < \left(\bar{\pi}^C\right)^n$ and $\pi^D < \left(\bar{\pi}^D\right)^{n+1}$.*

In Appendix A.3 we show that a continuity argument guarantees that this result holds for any Δ sufficiently small. We now give all the elements to prove this result in the limit as Δ goes to zero. We begin by highlighting the logic for why more generous offers increase Challenger's expected payoff from conflict.

When negotiations begin, Defender knows he has a chance to avoid conflict with normal Challenger for sure. Indeed, he could make to Challenger an offer so generous that normal Challenger would not reject. Yet, this offer is very large. In fact, it is not sufficient to offer Challenger the expected value of conflict given the initial public belief $\left(\pi^C, \pi^D\right)$. To see this, suppose that Challenger is expected to accept that offer with probability 1. If she then rejects, Defender would have to conclude that Challenger is tough for sure. Thus, normal Defender would concede at the first occasion: stage 2, $t = 1$. But then Challenger would strictly prefer to reject the offer unless it is at least equal to the expected payoff of going to conflict against a Defender who is so intimidated that, if normal, would concede at the first occasion: $\left(1-\pi^D\right) V - L^C$.

This logic can be easily extended to offers that are accepted with positive probability. A more generous offer that is accepted with greater probability is a double-edged sword. On the one hand, it increases the chances that conflict is avoided. On the other hand, it fosters Challenger's desire to reject, as a rejection boosts Challanger's expected payoff from conflict. To see this, let $\beta(x)$ be the total probability that Challenger accepts the offer x. That is, if normal challenger accepts with probability $\alpha(x)$, then $\beta(x) = \left(1-\pi^C\right)\alpha(x)$.

Then

$$\pi_0^C = \frac{\pi_C}{1-\beta(x)}. \tag{11}$$

Therefore, an offer x that is accepted with a larger probability $\beta(x)$ implies that, if it is rejected, conflict begins with a higher public belief that Challenger is tough. It follows that Defender is more likely to concede after the first attack. In particular, since the conflict horizon n is decreasing in π_0^C, the probability that Defender concedes in period 1,

$$1 - \frac{\pi^D}{\left(\bar{\pi}^D\right)^n}$$

and normal Challenger's expected payoff from conflict $u^C\left(\pi_0^C, \pi^D\right)$ (see Remark 1) increase with $\beta(x)$.

As Δ becomes small, we can make this key intuition more precise. Suppose first that pre-negotiation beliefs are such that Challenger is more intimidating. Therefore, since negotiations can only increase Defender's belief that Challenger is tough, if negotiations fail then conflict will also begin with Challenger being more intimidating then Defender. Notice that

$$n \approx \frac{\ln \pi_0^C}{\ln \bar{\pi}^C}.$$

Furthermore,

$$\lim_{\Delta \downarrow 0}\left(-\frac{\ln \bar{\pi}^D}{\ln \bar{\pi}^C}\right) = \lim_{\Delta \downarrow 0}\left(\frac{\ln\left(1-\left(1-e^{-r\Delta}\right)\ell^C\right)}{\ln\left(e^{-r\Delta}\left[1+\left(1-e^{-r\Delta}\right)\ell^D\right]\right)}\right)$$

which, after using de L'Hôpital's rule and taking the limit as $\Delta \downarrow 0$, yields

$$\lim_{\Delta \downarrow 0}\left(-\frac{\ln \bar{\pi}^D}{\ln \bar{\pi}^C}\right) = \frac{\ell^C}{1-\ell^D} < 0$$

where the last inequality follows from $\ln(a) < 0$ for all $a \in (0,1)$ and $\bar{\pi}^D, \bar{\pi}^C \in (0,1)$.[11] Then Defender concedes in period 1 with probability

$$1 - \frac{\pi^D}{\left(\bar{\pi}^D\right)^n} \approx 1 - \pi^D\left(\pi_0^C\right)^{\frac{\ell^C}{1-\ell^D}} \equiv P^D\left(\pi_0^C, \pi^D\right).$$

Therefore, Challenger accepts offer x only if

$$x \geq P^D\left(\pi_0^C, \pi^D\right). \tag{12}$$

[11] Alternatively, from Assumption 1.

When choosing the offer, Defender maximizes his expected payoff, given by

$$U^D = \beta(x)(1-x) + (1-\beta(x))\pi^D.$$

Since this expression is greater than 0, evidently Defender prefers to make an offer that tough Challenger would not accept (i.e., an offer less than 1). Yet, he may choose to make an offer that normal Challenger would accept with positive probability. If so, then (12) must hold, i.e., Defender maximizes U^D subject to (12). Furthermore, there is no point for Defender to offer more than what normal Challenger would accept anyway. Therefore, (12) must hold with equality, giving

$$x = P^D\left(\pi_0^C, \pi^D\right). \tag{13}$$

This indifference condition for Challenger highlights one side of the logic behind our main result: a higher offer corresponds to to a higher probability that Defender will concede after the first attack.

Using (11) and (13), we can therefore write x as a function of β and, after substituting this expression in Defender's objective function we have the following first order condition:

$$\frac{dU^D}{d\beta} = \pi^D\left(\frac{\pi^C}{1-\beta}\right)^{\frac{\ell^C}{1-\ell^D}} - \beta\frac{dx}{d\beta} - \pi^D = 0 \tag{14}$$

where

$$\frac{dx}{d\beta} = -\pi^D\left(\frac{\pi^C}{1-\beta}\right)^{\frac{\ell^C}{1-\ell^D}}\frac{1}{1-\beta}\frac{\ell^C}{1-\ell^D}.$$

After substituting for $dx/d\beta$, the first order condition in condition (14) yields

$$\frac{1}{\pi^D}\frac{dU^D}{d\beta} = \left(\frac{\pi^C}{1-\beta}\right)^{\frac{\ell^C}{1-\ell^D}}\left(1+\frac{\beta}{1-\beta}\frac{\ell^C}{1-\ell^D}\right) - 1 = 0. \tag{15}$$

It is easy to show that $dU^D/d\beta$ is strictly positive at $\beta = 0$, strictly negative at $\beta = 1 - \pi^C$, and U^D is single-peaked in $\beta \in (0,1)$. Therefore, when before negotiations begin Challenger is more intimidating, the equilibrium offer is both accepted and rejected with strictly positive probability by normal Challenger.

We can now turn to the case when pre-negotiation beliefs are such that Defender is more intimidating. That is, there exists $m \in \mathbb{N}$ such that $\left(\bar{\pi}^D\right)^{m+1} < \pi^D < \left(\bar{\pi}^D\right)^m$ and $\pi_0^C < \left(\bar{\pi}^C\right)^m$. In this case Defender has one more possibility: making an offer so small, and accepted with such a small probability, that even if Challenger rejects it the resulting post-negotiation beliefs still induce a conflict in which Defender is more intimidating. That is, Defender can make an offer x which is accepted with probability $\beta(x)$ such that

$$\frac{\pi^C}{1-\beta(x)} = \pi_0^C < \left(\bar{\pi}^C\right)^m. \tag{16}$$

We now show that this never happens in equilibrium.

To see why Defender does not make such a small offer in equilibrium, notice that (16) implies $\beta(x) < 1 - \pi^C$. Indeed, as we have argued before, if normal Challenger accepts with certainty ($\beta(x) = 1 - \pi^C$) then Defender would have to conclude that if Challenger rejects she is tough with certainty ($\pi_0^C = 1$). From the point of view of normal Challenger, though, accepting such an offer gives a payoff of $x > 0$. Instead, rejecting it, induces a conflict with Defender more intimidating than Challenger—an expected payoff of 0. Therefore, normal Challenger would strictly prefer to accept this offer with certainty, contradicting the hypothesis that $\beta(x) < 1 - \pi^C$.

Therefore, even if Defender is more intimidating before negotiations begin, after negotiations fail Challenger must be more intimidating. The optimal offer must then satisfy the first order condition in (14).

6 Brinkmanship (multiple offers)

In the previous sections, we established why a single round of negotiations may fail to avoid conflict. But with multiple rounds, Defender could potentially 'screen' Challenger, i.e., take out a proportion of normal types at each round, lowering the probability of conflict, which could conceivably go to 0 as the number of rounds increases. We now show that this is not in fact the case: our results on the failure of negotiation are robust to an arbitrary number of offers.

Suppose that there are K rounds of negotiations. At each round $k = 1, \ldots, K$, Defender can offer any fraction of the resource to Challenger. The conflict phase is triggered if all K offers are rejected.

Notice that Defender can always afford the same expected payoff he would get if there was only one round of negotiations. In fact, he can choose to make offers that no Challenger would accept until the K-th round, and then make the same offer he would make if there was only one round. Therefore, the question is whether Defender can do better than this by making an acceptable offer before the K-th and last round of negotiation. Proposition 4 says that the answer to this question is negative and that in equilibrium offers are accepted only on the brink of conflict.

Proposition 4. *Let β_k^*, $k = 1, \ldots, K$ be the equilibrium probability that Challenger accepts the offer in round k and β^* be the equilibrium probability that Challenger accepts if there is only one round of negotiation. For Δ sufficiently small, all equilibria must satisfy*

$$\beta_1^* = \ldots = \beta_{K-1}^* = 0 < \beta_K^* = \beta^*.$$

The proof in Appendix A.4 explicitly allows for positive intervals of time between rounds of negotiation. Therefore the result holds verbatim also when the different rounds of negotiation are separated by time.

Once again, the intuition behind our brinkmanship result lies in the ability of Challenger to intimidate Defender—in this case, into making larger offers in the future. To illustrate this intuition, we focus here on the simple case of $K = 2$. Let x^* be the optimal offer in our benchmark model. Notice that if an agreement is not reached in the first round, then the continuation game is identical to our benchmark model. Yet, Defender's belief that Challenger is tough when round 2 comes along, may have been affected by actions taken in round 1. Suppose that, in the first round, Defender makes an offer that normal Challenger accepts with probability $\beta_1' > 0$. If Challenger rejects, then Defender's belief that Challenger is tough rises to

$$\pi^{C\prime} = \frac{\pi^C}{1 - \beta_k'} > \pi^C.$$

But then, when round 2 comes along, Defender would make an offer larger than x^*. Since Challenger must be indifferent between accepting the first and the second offer, then the earlier offer too must be larger than x^*. Thus, the cost for Defender of buying Challenger's agreement is the same in the two rounds and it is greater than x^*, the cost it pays when there is only one round of negotiation. Since we noted that Defender can always attain the same expected payoff he would get if there was only one round of negotiations, making an acceptable offer in round 1 is not optimal for him.

7 Negotiating during conflict and informed Defender

Our benchmark model makes a stark distinction between before and after conflict begins. Before conflict begins, the parties can negotiate a peace that involves sharing the resource between them. After conflict begins, the parties can only concede the whole resource. As noted by Langlois and Langlois (2009) negotiations *during* conflict between states are uncommon. For example, Pillar (1983) finds that in only nineteen of one-hundred and forty two interstate wars parties negotiated during conflict and before an armistice. However, there are situations in which negotiations may happen after conflict begins. We now discuss a variant of the model in which Defender can make any offer to share the resource to Challenger in any period $t \in \{1, 2, \ldots\}$. Since Defender learns his type once conflict begins, this variant of the model also features offers from an informed Defender. The main message is that neither the possibility of negotiating during conflict nor Defender being informed at the time he makes an offer change our main result: initial negotiations fail with strictly positive probability and the ensuing conflict is prolonged.

Consider a variation of our model in which, after the initial negotiation fails, the following three-stage game is played out at each time $\tau_t = (t-1)\,\Delta$, $t \in \{1, 2, \ldots\}$.

Stage 1: Challenger attacks with certainty unless she has previously accepted an offer;

Stage 2: Defender makes an offer $x_t \in [0, 1]$;

Stage 3: Challenger chooses whether to accept the offer.

Notice that we assume that Challenger's decision to reject an offer and continue the conflict is taken at the end of each period, rather then at the beginning of the next.[12] We choose this option because it makes it easier to compare this model to our benchmark case. There, after Challenger attacks, Defender can choose to concede the whole resource, in which case the transfer happens immediately. Similarly here, after Challenger attacks, Defender can choose to concede a fraction of the resource. If Challenger accepts, then the transfer happens immediately.

Since Defender is informed about her type at the moment she makes an offer in period $t \geq 1$, then making a strictly positive offer reveals that he is normal. Here we focus on an equilibrium in which, if at the end of period t Defender is publicly known to be normal, then he concedes the whole resource in period $t + 1$.[13] In particular:

[12] The assumption that Challenger attacks with certainty after rejecting an offer is without loss of generality: accepting an offer strictly dominates rejecting the same offer and then concede in the next period.

[13] This may not be the unique equilibrium of the model. Nevertheless, in all equilibria (i) initial negotiations fail and (ii) conflict protacts for multiple periods with strictly positive probability.

(Initial) negotiations fail. In the equilibrium, Defender makes an The choice to escalate is similar to our choice to go to conflict, but in our model the parties go to conflict with strictly positive probability in the unique equilibrium, and even if this decision limits their ability to negotiate further, and for any relative cost of conflict for Challenger and Defenderoffer that is both accepted and rejected by normal Challenger with strictly positive probability. Therefore, conflict begins with probability strictly greater than π^C, but strictly smaller than 1.

Conflict. After negotiations fail, conflict begins with post-negotiation beliefs $\left(\pi_0^C, \pi^D\right)$. In equilibrium, these beliefs are such that Challenger strictly prefers to attack. From then on, Defender mixes between offering nothing and $x_t = \underline{x} = 1/\bar{\pi}^C$ as long as π_{t-1}^C is less than a threshold level $\pi^{C*} < 1$ (see below). Normal challenger accepts the offer $\underline{x}$ is with certainty and rejects the offer of nothing. If π_{t-1}^C is greater than π^{C*} but less than $\bar{\pi}^C$, then Defender mixes between offering nothing and $x_t = 1$. Once π_{t-1}^C is greater than $\bar{\pi}^C$, if Challanger attacks, Defender concedes with probability 1. Thereafter, both normal types concede with probability 1.

The logic of the equilibrium play during conflict is akin to the one underlying our results for our benchmark model. In equilibrium, in any period $t \geq 1$ and as long as $\pi_t^C \leq \pi^{C*}$, normal Defender makes the offer $\underline{x}$ with a probability that makes normal Challenger indifferent between accepting and rejecting the offer. Similarly, if Defender offers nothing, normal Challenger concedes with a probability that makes normal Defender indifferent between offering nothing and $\underline{x}$. Along the equilibrium path, in any period in which Defender does not make a positive offer, his reputation for being tough increases. Similarly, after any period in which Challenger does not concede, her reputation for being tough increases. While the whole process is similar to the one in our benchmark model, the probability with which normal Challenger concedes is greater than in the benchmark model. In fact, while the cost of attacking is the same in the two models, normal Challenger's gains from forcing Defender to concede are now smaller, as in equilibrium Defender will only concede a fraction $\underline{x}$ to normal Challenger. Similarly, normal Defender makes positive offers with higher probability than he would concede in the benchmark model. In fact, while the cost of a further attack from Challenger is the same, the loss from conceding to normal Challenger is smaller as Defender only needs to concede a fraction $\underline{x}$ of the resource. Therefore, in this variation of the model, reputation grows faster and conflict is on average shorter. This reinforces our general message that while negotiations cannot avoid conflict for sure, they nonetheless make conflict shorter.

We can determine $\underline{x}$ as follows. If Challenger rejects $x_t = \underline{x}$, then she can extract the whole resource in period $t+1$ at the cost of one attack. Therefore, the payoff of rejecting $x_t = \underline{x}$ equals $\delta\left(1 - L^C\right)$.[14] Therefore, the minimum offer $x_{\min}$ that would be accepted by Challenger is given

$$x_{\min} = \delta\left(1 - L^C\right)$$
$$x_{\min} = \delta\left[1 - L^C\right] = \frac{1}{\bar{\pi}^C}.$$

Obviously, Defender would not offer more than $x_{\min}$ and therefore $\underline{x} = x_{\min}$.

Furthermore, we can explicitly calculate π^{C*}, the minimum probability that Defender puts on Challenger being tough before he decides that he is going to concede the entire resource, rather than get an agreement from the normal type at a slightly lower concession $\underline{x}$ and then concede (to the tough type) if $\underline{x}$ isn't enough. Let us denote by $u^D(\tau)$ Defender's expected payoff along an equilibrium path where it takes exactly τ periods before normal Defender concedes with certainty (so that absent a concession after τ periods the probability of Defender being tough reaches 1). The following two options are always open to Defender. (i) $u^D(0) = 0$ from conceding immediately; (ii) $u^D(1)$ from conceding in the next period. The second option amounts to making an offer $\underline{x}$ that is accepted by normal Challenger in the current period and leave tough Challenger to attack next period and extract full concession:

$$u^D(1) = \left(1 - \pi_t^C\right)(1 - \underline{x}) + \pi_t^C\left(V - \delta L^D\right).$$

Therefore, $u^D(1) > u^D(0)$ if and only if

$$\pi_t^C < \frac{(1 - \underline{x})}{(1 - \underline{x}) + (\delta L^D - V)} =: \pi^{C*} < 1.$$

We can now turn to the question of why the original negotiations at time $\tau = 0$ fail with strictly positive probability. Defender's expected payoff if negotiations fail is given by the loss $-L^D$ of one period of conflict, exactly as in our benchmark model. Instead, Challenger's expected payoff from conflict is now different. Yet, it still depends only on the probability that Defender concedes (offers $\underline{x}$) in period 1 and the value for Challenger of such a concession (the present discounted value of a $1 - \underline{x}$ share of the resource).

In the limit as Δ tends to 0, the offer $\underline{x}$ approaches 1. That is, as Δ approaches 0, the

[14] Recall that $V = \left(1 - e^{-r\Delta}\right) = 1 - \delta$. Therefore $\delta\left(\frac{V}{1-\delta} - L^C\right) = \delta\left(1 - L^C\right)$.

equilibrium in this version of the model approaches the the equilibrium in our benchmark model. Therefore, Challenger's expected payoff from conflict also approaches its value in our benchmark model. It follows that, in the limit as Δ approaches 0, the equilibrium in pre-conflict negotiations is identical to the one in our benchmark model. As for Proposition 3, a continuity argument guarantees that this result holds for any Δ sufficiently small.

8 Discussion

We have seen how negotiations and conflict are linked by a two-way feedback. On the one hand, when choosing how to negotiate, the parties to a dispute take into account how they expect an eventual conflict to unfold. On the other hand, the way conflict unfolds is also determined by how and why negotiations fail. In modeling this two-way feedback between negotiations and conflict, we uncovered a novel reason for why negotiations may fail to avoid conflict: offers that have a higher probability of being accepted also increase the incentive for the aggressor to initiate conflict. Thus, negotiations can mitigate but not prevent conflict. This result should not be seen as a claim that negotiations are useless: a neutral observer who seeks to reduce conflict will gain from bringing the parties to the negotiation table. If negotiations succeed, then conflict is avoided; if they fail, conflict will be shorter. In fact, in our model, the result of failed negotiations is to increase the belief π_0^C that Challenger is tough, and therefore decrease the conflict horizon n. Furthermore, the ability to negotiate increases the expected payoff of both players.

One possible caveat of our approach is that we do not give to Defender a chance to afford peace with a tough Challenger unless he offers the entire resource to Challenger. In many realistic applications, even a committed Challenger would admit that before conflict ensues she is willing to accept a smaller offer $X < 1$. Nevertheless, such a scenario would not change the main insight from our model unless X is sufficiently small. In fact, suppose that $X > \left(1-\pi^D\right)V$. As we discussed in Section 5, an offer equal or greater to $\left(1-\pi^D\right)V$ would convince normal Challenger to accept for sure. We can then see whether Defender would prefer to strike a deal only with normal Challenger or with both normal and tough Challenger. In this case, as $\Delta \downarrow 0$, Defender's expected payoff of offering N equals

$$\frac{1}{r}\left[\left(1-\pi^C\right)(1-N)+\pi^C\pi^D\right]$$

where $N = \lim_{\Delta\downarrow 0}\left(1-\pi^D\right)V$. Instead, Defender's expected payoff of offering X equals

$(1-X)/r$. Therefore, Defender prefers to strike a deal with only normal Challenger as long as

$$X > \left(1-\pi^C\right) N + \pi^C \left(1-\pi^D\right).$$

From Section 5, we know that if Defender prefers to strike a deal with only normal Challenger, then it strictly prefers to make an offer that normal Challenger both accepts and rejects with strictly positive probability.

In our model, we also do not allow Challenger to offer peace deals. But our key insight about the failure of negotiations is also true if both Defender and Challenger can make offers. In fact, if normal Defender accepts the equilibrium offer for sure, then post-negotiation beliefs would be such that normal Challenger would never attack (that is, $\pi_0^D = 1$). But then normal Defender would strictly prefer to reject such an offer. Similarly, consider a variant of the model in which there are K rounds of two-step negotiations. At each stage, Challenger makes a demand first. If Defender accepts the offer, the game ends; otherwise, Defender makes an offer. If Challenger accepts the offer, the game ends; otherwise, the game moves to the next round of negotiations (or to conflict, if the last round of negotiations has already been reached). If Challenger ever demands less than unity, she reveals herself to be the normal type and, in the ensuing game, Defender has no incentive to concede anything to her and Challenger would never attack. Thus Challenger will simply demand 1 at each round and Defender will refuse and offer zero until the last round—our brinkmanship result.

Our brinkmanship result shows that there is little point in insisting on multiple rounds of negotiations. If acceptable offers are to be made, they will only be made at the last opportunity to avert conflict. In a sense, neutral observers should take advantage of any ultimatum imposed by the parties, rather than pressuring them to have softer deadlines. An interesting extension of our model would be to allow for there to be multiple commitment types of Challenger, from softer ones that would accept all offers above a threshold less than 1 to the toughest ones who would accept only $x = 1$. Even in this model, the forces that lead to delay would be present, but an additional force would be at work—an uninformed Defender could potentially screen some of the softest types in the early rounds of negotiations. Furthermore, the main insights from our model may still drive the analysis if we were to allow softer types to strategically mimic the toughest ones.

References

[1] **Abreu, Dilip and Faruk Gul (2000).** Bargaining and Reputation, *Econometrica*, 68(1): 85–117.

[2] **Acharya, Avidit and Edoardo Grillo (2015).** War with Crazy Types. *Political Science Research and Methods*, 3(2): 281–307.

[3] **Bénabou, Roland and Jean Tirole (2009).** Over my Dead Body: Bargaining and the price of dignity. *The American Economic Review* 99(2): 459–465.

[4] **Bennet, Scott and Allan C. Stam (1996).** The Duration of Interstate Wars, 1816-1985, *The American Political Science Review*, 90(2): 239–257.

[5] **Bennet, Scott and Allan C. Stam (2009).** Revisiting Predictions of War Duration, *Conflict Management and Peace Science*, 26(3): 256–267.

[6] **Brito, Dagobert L. and Michael D. Intriligator (1985).** Conflict, War and Redistribution, *The American Political Science Review*, 79(4): 943–957.

[7] **Bueno de Mesquita, Bruce and Alastair Smith (2012).** Domestic Explanations of International Relations, *Annual Review of Political Science*, 14: 1–21.

[8] **Chatterjee Kalyan and Larry Samuelson (1988).** Bargaining Under Two-sided Incomplete Information: The Unrestricted Offers Case. *Operations Research*, 36(4): 605–618.

[9] **Collier, Paul, Anke Hoeffler, and Mans Soderbom (2004).** On the Duration of Civil War, *Journal of Peace Research*, 41(3): 253–273.

[10] **Crawford, Vincent P. (1982).** A Theory of Disagreement in Bargaining, *Econometrica*, 50(3): 607–637.

[11] **Esteban, Joan and József Sákovics (2008).** A Theory of Agreements in the Shadow of Conflict: The Genesis of Bargaining Power, *Theory and Decision*, 65(3): 227–252.

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

[13] **Gul, Faruk, Hugo Sonnenschein and Robert Wilson (1986).** Foundations of Dynamic Monopoly and the Coase Conjecture, *Journal of Economic Theory* 39: 155–190.

[14] **Heifetz, Aviad and Ella Segev (2005).** Escalation and Delay in Protracted International Conflicts, *Mathematical Social Sciences*, 49: 17–37.

[15] **Henderson, Errol A. and Reşat Bayer (2013).** Wallets, Ballots, or Bullets: Does Wealth, Democracy, or Military Capabilities Determine War Outcomes?, *International Studies Quarterly*, 57(2): 303–317.

[16] **Hodler, Roland and Dominic Rohner (2012).** Electoral Terms and Terrorism, *Public Choice*, 150(1): 181–193.

[17] **Kreps, David and Robert Wilson (1982).** Reputation and Imperfect Information. *Journal of Economic Theory*, 27: 253–79.

[18] **Langlois, Catherine C. and Jean-Pierre P. Langlois (2009).** Does Attrition Behavior Help Explain the Duration of Interstate Wars? A Game Theoretic and Empirical Analysis, *International Studies Quarterly*, 53(4): 1051–1073.

[19] **Lapan, Harvey E. and Todd Sandler (1988).** To Bargain or Not to Bargain: That Is the Question, *The American Economic Review*, 78(2): 16–21.

[20] **Milgrom, Paul and John Roberts (1982).** Predation, Reputation, and Entry Deterrence, *Journal of Economic Theory*, 27: 280–312.

[21] **Özyurt, Selçuk (2014).** Audience Costs and Reputation in Crisis Bargaining, *Games and Economic Behavior*, 88: 250–259.

[22] **Pillar, Paul R. (1983).** *Negotiating Peace: War Termination as a Bargaining Process*, Princeton: Princeton University Press.

[23] **Ponsatì Clara and József Sákovics (1995).** The War of Attrition with Incomplete Information, *Mathematical Social Sciences*, 29(3): 239-254.

[24] **Powell, Robert (1996).** Bargaining in the Shadow of Power, *Games and Economic Behaviour*, 15: 255–289.

[25] **Powell, Robert (2004).** Bargaining and Learning While Fighting, *American Political Science Review*, 48(2): 344–361.

[26] **Reiter, Dan (2003).** Exploring the Bargaining Model of War, *Perspectives on Politics*, 1(1): 27–43.

[27] **Sanchez-Pages, Santiago (2009).** Conflict as a Part of the Bargaining Process, *Economic Journal*, 119: 1189–1207.

[28] **Sechser, Todd S. (2010).** Goliath's Curse: Coercive Threats and Asymmetric Power, *International Organization*, 64(4): 627–660.

[29] **Schelling, Thomas C. (1956).** An Essay on Bargaining, *American Economic Review*, 6: 281–306.

[30] **Schelling, Thomas C. (1960).** *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.

[31] **Schelling, Thomas C. (1966).** *Arms and Influence*, New Haven, CT: Yale University Press.

[32] **Silverman, Dan (2004).** Street Crime and Street Culture, *International Economic Review*, 45(3): 761–786.

[33] **Stanton, Fredrik (2010).** *Great Negotiations: Agreements that Changed the Modern World*, Yardley, Pennsylvania: Westholme Publishing.

[34] **Yared, Pierre (2009).** A Dynamic Theory of War and Peace, *Journal of Economic Theory*, 145: 1921–1950.

[35] **Yildiz, Muhamet (2005).** Waiting to Persuade. *The Quarterly Journal of Economics*, 119(1): 223–248.

A Omitted proofs

A.1 Proof of Lemma 1

Proof. (i) Assume that normal Defender concedes immediately after Challenger attacks in period $t+1$. That is $\sigma^D_{t+1}\left(\pi^C_{t+1}, \pi^D_t\right) = 1$. By Bayes's rule, if Defender does not concede, then $\pi^D_{t+1} = 1$. Therefore, normal Challenger would concede at $t+2$. Then the expected payoff of attacking in period $t+1$ equals

$$\left(1-\pi^D_t\right)\frac{V}{1-\delta} - L^C = \left(1-\pi^D_t\right) - L^C.$$

Since not attacking yields a payoff of 0, Challenger strictly prefers to attack if $\pi^D_t < \bar{\pi}^D$, is indifferent if $\pi^D_t = \bar{\pi}^D$, and strictly prefers to concede if $\pi^D_t > \bar{\pi}^D$.

(ii) Assume that normal Challenger concedes in period $t+1$. That is, $\sigma_{t+1}\left(\pi^C_t, \pi^D_t\right) = 1$. By Bayes's rule, if Challenger does not concede, then $\pi^C_{t+1} = 1$. Therefore, normal

Defender would concede at $t+1$. Then the expected payoff of resisting in period t equals

$$V + \delta\left[\left(1-\pi_t^C\right)\frac{V}{1-\delta} - \pi_t^C L^D\right] = (1-\delta) + \delta\left[\left(1-\pi_t^C\right) - \pi_t^C L^D\right].$$

Since resisting yields a payoff of 0, Defender strictly prefers to resist if $\pi_t^C < \bar{\pi}^C$, is indifferent if $\pi_t^C = \bar{\pi}^C$, and strictly prefers to concede if $\pi_t^C > \bar{\pi}^C$. □

A.2 Equilibrium in the game of conflict (Propositions 1 and 2)

We now present a few lemmas that identify necessary conditions for equilibrium and thereby pin down the unique one in the game of conflict.

We first ask if a normal player ever mimics the tough type. Lemma 2 says that only Challenger in period 1 can mimic the tough type (i.e. attack) with probability 1.

Lemma 2. *In any equilibrium, normal types of both players concede with strictly positive probability in all periods, except possibly Challenger in period* 1*.*

Proof. The *concession sequence* $\langle\kappa_i\rangle_{i\in\mathbb{N}}$ of any strategy profile is a sequence in $[0,1]$, where each odd (even) term is the probability that Challenger (respectively, Defender) concedes at that time conditional on no player having conceded yet. A concession sequence arising from an equilibrium profile is called an *equilibrium concession sequence*.

Lemma 2 then says that in any equilibrium concession sequence, all terms (except possibly the first) must be strictly positive.

Step 1. The proof is based on the key idea that if the string $(\kappa_i, 0, \kappa_{i+2})$ appears in an equilibrium concession sequence and $\kappa_{i+2} > 0$, then $\kappa_i = 1$: if the opponent is not conceding in the interim the value of concession can only go down because the positive cost to fighting strictly exceeds the flow utility derived from the resource; therefore concession should have been strictly better at the step before.

Step 2. We now show that, along any concession sequence, adjacent terms cannot be 0. Let $\kappa_i = 0 = \kappa_{i+1}$; if $\kappa_{i+2} > 0$, it would contradict Step 1. Induction implies that if two adjacent terms of the concession sequence are 0, all subsequent terms are 0 too. But since there is a positive probability of the tough type, it cannot be an equilibrium to never concede, knowing that your opponent will not. Therefore, no equilibrium concession sequence contains adjacent 0's.

Step 3. Suppose $\kappa_i = 0$ for some $i > 1$. By Step 2, we must have $\kappa_{i+1} > 0$; from Step 1 it means that $\kappa_{i-1} = 1$. If the player who is supposed to concede with probability 1 does not do so, his/her reputation immediately jumps to 1 and the normal opponent must concede immediately thereafter, i.e. $\kappa_i = 1$—a contradiction! □

Lemma 3 characterizes the players' strategies if they are indifferent for two consecutive periods.

Lemma 3. *If Challenger is indifferent between conceding at times t and $t+1$ in any equilibrium, then normal Defender's equilibrium concession probability and the public beliefs about him are*

$$\tilde{\sigma}^D\left(\pi^D_{t-1}\right) := \frac{1-\bar{\pi}^D}{1-\pi^D_{t-1}}; \textit{ and } \pi^D_t = \frac{\pi^D_{t-1}}{\bar{\pi}^D} \tag{17}$$

respectively. Similarly, if Defender is indifferent between conceding at times t and $t+1$ in any equilibrium, then normal Challenger's probability of conceding and the public beliefs about her type are

$$\tilde{\sigma}^C\left(\pi^C_t\right) := \frac{1-\bar{\pi}^C}{1-\pi^C_t}, \textit{ and } \pi^C_{t+1} = \frac{\pi^C_t}{\bar{\pi}^C}. \tag{18}$$

Proof. Challenger is indifferent if and only if $V^C_t\left(\pi^C_{t-1}, \pi^D_{t-1}\right) = 0$, which is the payoff of Challenger from conceding. Suppose that Challenger is indifferent for two consecutive periods: $V^C_t\left(\pi^C_{t-1}, \pi^D_{t-1}\right) = V^C_{t+1}\left(\pi^C_t, \pi^D_t\right) = 0$. By Lemma 2, $\sigma^C_{t+1}\left(\pi^C_t, \pi^D_t\right) \neq 1$. Therefore

$$\bar{\sigma}^D_t\left(\pi^C_t, \pi^D_{t-1}\right) = 1 - \bar{\pi}^D.$$

This corresponds to a strategy for Defender such that $\sigma^D_t\left(\pi^C_t, \pi^D_{t-1}\right) = \tilde{\sigma}^D\left(\pi^D_{t-1}\right)$ in (17) if $\pi^D_{t-1} \leq \bar{\pi}^C$.

Defender is indifferent if and only if $V^D_t\left(\pi^C_t, \pi^D_{t-1}\right) = -L^D$, the payoff Defender gets if he concedes.[15] Suppose that Defender is indifferent for two consecutive periods (or is indifferent at time t and concedes at time $t+1$): $V^D_{t+1}\left(\pi^C_{t+1}, \pi^D_t\right) = -L^D = V^D_t\left(\pi^C_t, \pi^D_{t-1}\right)$. By Lemma 2, $\sigma^D_t\left(\pi^C_t, \pi^D_{t-1}\right) \neq 1$. Therefore

$$\bar{\sigma}^C_{t+1}\left(\pi^C_t, \pi^D_t\right) = 1 - \bar{\pi}^C.$$

This corresponds to a strategy for Challenger such that $\sigma^C_{t+1}\left(\pi^C_t, \pi^D_t\right) = \tilde{\sigma}^C\left(\bar{\pi}^C_t\right)$ in (18) if $\pi^C_t \leq \bar{\pi}^C$. □

[15] This payoff is not 0 but $-L^D$ because when it is Defender's turn to decide if he wants to concede or prolong the fight, Challenger has already attacked and the loss will be experienced by Defender in the current period regardless of his choice of move.

Remark 3. Challenger's mixing probability at $t = 1$ need not equal $\tilde{\sigma}^C$; Defender's mixing at $t = 1$ can be different from $\tilde{\sigma}^D$ only if Challenger strictly prefers to attack at $t = 1$.

Combining Lemmas 2 and 3, in equilibrium both players concede with the probabilities in Lemma 3 above, except possibly at $t = 1$. Beliefs evolve according to the above lemma, except possibly at $t = 1$ and until they hit $\bar{\pi}^C$ or $\bar{\pi}^D$.

Conflict continues as long as no player has conceded. If players mix, then beliefs about their type increase until a threshold is crossed.

Lemma 4 says that if both beliefs are strictly below their threshold, no belief leaps over the corresponding threshold at the next step without touching the corresponding threshold exactly.

Lemma 4. *In equilibrium (i) $\pi_t^D < \bar{\pi}^D$ and $\pi_{t+1}^C < \bar{\pi}^C$ implies $\pi_{t+1}^D \leq \bar{\pi}^D$; (ii) $\pi_t^C < \bar{\pi}^C$ and $\pi_t^D < \bar{\pi}^D$ implies $\pi_{t+1}^C \leq \bar{\pi}^D$.*

Proof. Suppose not. Let $\pi_t^D < \bar{\pi}^D$, $\pi_{t+1}^C < \bar{\pi}^C$ but $\pi_{t+1}^D > \bar{\pi}^D$. Lemma 1 implies that normal Challenger will concede with probability 1 at time $t + 2$ if Defender does not concede at $t + 1$. So if Defender does not concede at time $t + 1$ he gets a continuation payoff of 1 from $t + 2$ onwards if Challenger is the normal type. Since Challenger is normal with probability $1 - \pi_{t+1}^C$, Defender's payoff from $t + 1$ (the current period) onwards is

$$(1-\delta)\left(V - L^D\right) + \delta\left[\left(1-\pi_{t+1}^C\right)V - \pi_{t+1}^C L^D (1-\delta)\right].$$

Defender strictly prefers to not concede if the above exceeds the payoff $-(1-\delta) L^D$ from conceding immediately at $t + 1$:

$$(1-\delta) V + \delta\left[\left(1-\pi_{t+1}^C\right)V - \pi_{t+1}^C L^D (1-\delta)\right] > 0. \tag{19}$$

Inequality (19) reduces to $\pi_{t+1}^C < \bar{\pi}^C$, which is true by assumption. Therefore Defender strictly prefers to fight at $t + 1$, i.e. $\sigma_{t+1}^D\left(\pi_{t+1}^C, \pi_t^D\right) = 0$—which contradicts Lemma 2, implying that $\pi_t^D < \bar{\pi}^D$ and $\pi_{t+1}^C < \bar{\pi}^C$ cannot lead to $\pi_{t+1}^D > \bar{\pi}^D$.

Now let $\pi_t^C < \bar{\pi}^C$ and $\pi_t^D < \bar{\pi}^D$, but $\pi_{t+1}^C > \bar{\pi}^C$. By a similar logic Challenger strictly prefers to fight at $t + 1$ if

$$-(1-\delta) L^C + \left(1-\pi_t^D\right)V > 0.$$

The expression above reduces to $\pi_t^D < \bar{\pi}^D$. So Challenger strictly prefers to fight at $t + 1$, i.e. $\sigma_{t+1}^C\left(\pi_t^C, \pi_t^D\right) = 0$—which contradicts Lemma 2. □

What do our previous results imply about period 1's probability of attack? By Lemma

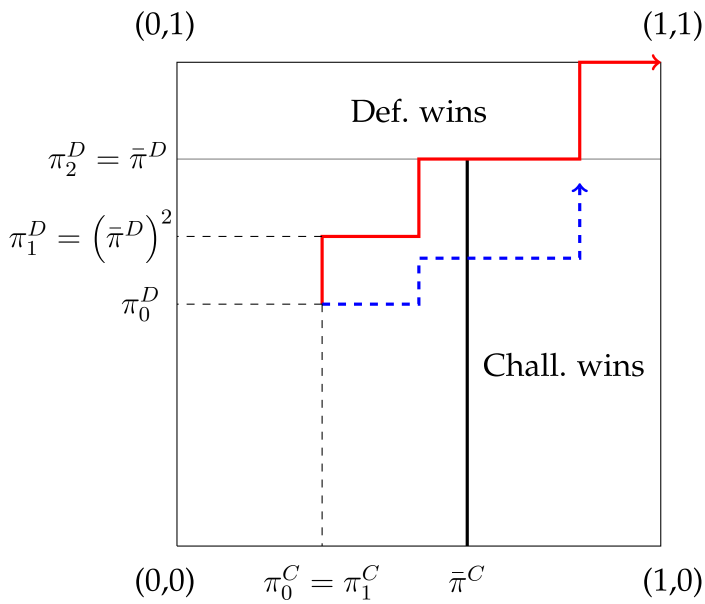

Figure 1: How Beliefs Evolve

3, from period 2 onward beliefs must grow by a factor $\left(\bar{\pi}^C\right)^{-1}$ and $\left(\bar{\pi}^D\right)^{-1}$, respectively. The solid line in Figure 1 depicts the equilibrium evolution of beliefs in a conflict of horizon 2 with Challenger more intimidating than Defender. The dashed line represents the evolution of beliefs if it is common knowledge that both Defender and Challenger play the strategies in Lemma 3 from period 1 onward. In this case, $\pi_1^D < \bar{\pi}^D$ and $\pi_2^C > \bar{\pi}^C$, violating Lemma 4. In equilibrium, Defender must concede with sufficiently large probability in period 1 so as to 'level the playing field' with Challenger and guarantee $\pi_2^D = \bar{\pi}^D$. Since Defender is conceding with a higher probability than what would make Challenger indifferent, in period 1 Challenger strictly prefers to attack.

Remark 4. If Challenger is more intimidating than Defender, in period 1 Challenger attacks with probability 1 and Defender concedes with probability $1 - \pi^D / \left(\bar{\pi}^D\right)^n > 1 - \bar{\pi}^D$.

A similar logic applies to the case when Defender is at least as committed as Challenger. In this case, Challenger must concede with sufficiently high probability in period 1 so as to 'level the playing field' with Defender and guarantee $\pi_n^C = \bar{\pi}^C$.

Remark 5. If Defender is more intimidating than Challenger, in period 1 Challenger attacks with probability $\pi_0^C / \left(\bar{\pi}^C\right)^n$.

The next lemma shows that along the equilibrium path, provided no one concedes, both reputations grow according to equations (2) and (3) from period 2 onwards until a time t when either $\pi_t^C = \bar{\pi}^C$ or $\pi_t^D = \bar{\pi}^D$.

Lemma 5. *For any period $t \geq 2$, if $\pi_{t-1}^D \leq \bar{\pi}^D$ and $\pi_t^C \leq \bar{\pi}^C$, then Challenger plays $\tilde{\sigma}^C\left(\pi_t^C\right)$ and Defender plays $\tilde{\sigma}^D\left(\pi_{t-1}^D\right)$.*

Proof. We show the result for Defender. The result for Challenger follows a symmetric argument.

Proceed by contradiction. If $\sigma_t^D\left(\pi_t^D\right) \neq \tilde{\sigma}^D\left(\pi_t^D\right)$, by Lemma 3, Challenger is not indifferent at either t or at $t+1$. There are two possibilities. First, she strictly prefers to concede. But then Defender would concede with probability 0 in the previous period, contradicting Lemma 2. Second, she strictly prefers to fight. But then by Lemma 2 she is Challenger in period 0 and $t = 1 < 2$. □

The lemma above is useful in proving Propositions 1 and 2, which apply, respectively, to the cases where Challenger is more committed than Defender and where Defender is at least as committed as Challenger.

A.2.1 Proof of Proposition 1

Proof. Existence. We first show that the stategies σ^* defined in Proposition 1 constitute an equilibrium. From Lemma 3 it is clear that after the first move by Challenger in period 1 players are indifferent and therefore willing to mix. Since normal players concede in σ^* once the thresholds are crossed, this is consistent with Lemma 1. Since Defender concedes with a *larger* probability than $\tilde{\sigma}^D$ in the first period, Lemma 3 implies that Challenger strictly prefers to fight at $t = 1$. Also note that by Bayes' rule the equilibrium belief about Challenger's type after non-concession at $t = 1$ is given by $\left(\bar{\pi}^C\right)^n$.

Uniqueness. If $\pi_0^C \geq \bar{\pi}^C$, then Lemma 1 implies that the above is the only equilibrium; similarly for the case $\pi_0 \geq \bar{\pi}$. Therefore let $\left(\pi_0^C, \pi_0^D\right) < \left(\bar{\pi}^C, \bar{\pi}^D\right)$, so that $n \geq 1$. If normal types follow $\tilde{\sigma}^C, \tilde{\sigma}^D$ defined in equations (18) and (17) up to and including time n, there will be a jump since $\pi_0^C / \left(\bar{\pi}^C\right)^n > \bar{\pi}^C$; but jumps are ruled out by Lemma 4. By Lemmas 3 and 5, the only freedom we have is in choosing different strategies for $t = 1$.

By contradiction, suppose that Challenger concedes with positive probability in period 1. This implies she expects Defender to concede with probability at least $\tilde{\sigma}^D$. But this implies that there is $m \leq n$ such that beliefs are $\left(\pi_{m+1}^C, \pi_m^D\right)$ with $\pi_{m+1}^C > \bar{\pi}^C$ and $\pi_m^D < \bar{\pi}^D$, contradicting Lemma 4.

Last, since Challenger cannot concede with probability less than 0, we have that $\pi_{n+1}^C > \bar{\pi}^C$. Thus, by Lemma 4, Defender must concede in period 1 with probability exactly σ_t^{D*}. □

A.2.2 Proof of Proposition 2

Proof. Existence. Lemmas 1 and 3 imply that the above is an equilibrium. In particular, σ_1^{C*} and Bayes's rule imply that the equilibrium belief about Challenger's type after non-concession at $t = 1$ is given by $\left(\bar{\pi}^C\right)^n$.

Uniqueness. If $\pi_0^C \geq \bar{\pi}^C$, then Lemma 1 implies that the above is the only equilibrium; similarly for the case $\pi_0^D \geq \bar{\pi}^D$. Therefore let $\left(\pi_0^C, \pi_0^D\right) < \left(\bar{\pi}^C, \bar{\pi}^D\right)$, so that $n \geq 1$. If normal types follow $\tilde{\sigma}^C, \tilde{\sigma}^D$ defined in equations (18) and (17) up to and including time n, there will be a jump since $\pi_0^D / \left(\bar{\pi}^D\right)^n > \bar{\pi}^D$; but jumps are ruled out by Lemma 4. By Lemmas 3 and 5, the only freedom we have is in choosing different strategies for $t = 1$.

Case 1: $\sigma_1^C < \sigma_1^{C*}$. Suppose that $\sigma_1^C < \sigma_1^{C*}$. The inequality $\sigma_1^C < \sigma_1^{C*}$ implies that Challenger's reputation increases at a slower rate such that $\pi_n^C < \bar{\pi}^C$. If $\sigma_1^D < \tilde{\sigma}_1^D$, then Challenger prefers to concede immediately ($\sigma_1^C = 1$) since Challenger is just indifferent at $\tilde{\sigma}^D$; this contradiction implies that $\sigma_1^D \geq \tilde{\sigma}^D$, which in turn gives $\pi_1^D \geq \pi_0^D / \bar{\pi}^D$ and therefore $\pi_n^D > \bar{\pi}^D$ i.e. there exists $m \leq n$ such that belief profile is $\left(\pi_m^C, \pi_m^D\right)$ with $\pi_m^C < \bar{\pi}^C$ and $\pi_m^D > \bar{\pi}^D$, contradicting Lemma 4. Therefore, $\sigma_1^C \geq \sigma_1^{C*}$ is the only possibility in equilibrium.

Case 2: $\sigma_1^C > \sigma_1^{C*}$. Suppose that $\sigma_1^C > \sigma_1^{C*}$. Now $\pi_1^C > \pi_0^C / \bar{\pi}^C$, $\pi_2^C > \pi_0^C / \left(\bar{\pi}^C\right)^2$, etc. Since Proposition 1 implies that Defender's reputation is growing as the same rate $1/\bar{\pi}^D$ it follows from the Defender being more intimidating and $\left(\bar{\pi}^C\right)^{n+1} < \pi_0^C$ that $\pi_n^C > \bar{\pi}^C$, i.e., a jump occurs by time n. Therefore, $\sigma_1^C \leq \sigma_1^{C*}$ is the only possibility in equilibrium.

Last, since Challenger must be indifferent at $t = 0$ to play σ_1^{C*}, then $\sigma_1^D = \tilde{\sigma}^D = \sigma_1^{D*}$. □

A.3 Proof of Proposition 3

Proof. We now show that that the solution to the game with discrete (i.e. $\Delta > 0$) but frequent enough opportunities to concede approaches the solution of the limiting model in the body of the paper. Let $U_\Delta^D(x, \beta)$ denote the utility of the uninformed Defender when his offer of x is accepted with probability β:

$$U_\Delta^D(x, \beta) := \beta(1 - x) + (1 - \beta)\left\{\pi^D - \left(1 - \pi^D\right)\ell^D\left(1 - e^{-r\Delta}\right)\right\}. \tag{20}$$

Defender's problem is

$$\mathcal{P}_{\Delta}: \quad \max\ U_{\Delta}^{D}(x,\beta) \text{ s.t. } (x,\beta) \in F(\Delta), \tag{21}$$

where $F : [0,1] \to [0,1] \times \left[0, 1-\pi_0^C\right]$ gives the set of feasible pairs of (β, x) at Δ. A value of Δ determines $\delta = e^{-r\Delta}$ and thus the thresholds $\bar{\pi}^C$ and $\bar{\pi}^D$ according to (4) and (5), which in turn determine the conflict horizon $n(\Delta)$ as

$$\frac{\pi^C/(1-\beta)}{\left(\bar{\pi}^C\right)^{n(\Delta)-1}} < \bar{\pi}^C \leq \frac{\pi^C/(1-\beta)}{\left(\bar{\pi}^C\right)^{n(\Delta)}}. \tag{22}$$

This value of $n(\Delta)$ in turn determines the probability $P^D(\Delta,\beta)$ with which Defender concedes at the start. P^D is given by

$$1 - \frac{\pi^D}{\left(\bar{\pi}^D\right)^{n(\Delta)+1}} \leq P^D(\Delta,\beta) \leq 1 - \frac{\pi^D}{\left(\bar{\pi}^D\right)^{n(\Delta)}} \tag{23}$$

if the inequality in (22) holds as an equality, and it is given by

$$P^D(\Delta,\beta) = 1 - \frac{\pi^D}{\left(\bar{\pi}^D\right)^{n(\Delta)}} \tag{24}$$

if both inequalities in (22) are strict.

Each value of P^D determines a unique value of the offer $x(\Delta,\beta)$ that leaves Challenger indifferent between accepting and rejecting the offer, which is given by the following indifference condition:

$$P^D(\Delta,\beta) - \ell^C\left(1 - e^{-r\Delta}\right) = x(\Delta,\beta). \tag{25}$$

We assert that F is a compact-valued correspondence continuous at $\Delta = 0$. The image of Δ under F can be written as a correspondence $\phi : \beta \twoheadrightarrow \phi(\beta)$ where $\phi(\beta)$ is a singleton given by (24) if (Δ,β) satisfies Assumption 2, and is otherwise a closed interval given by (23). Fix any β. Suppose $\left(\Delta^k\right)_{k\geq 1}$ is any sequence decreasing to 0; let $\left(\beta, x^k\right)_k$ be a sequence of points in $F\left(\Delta^k\right)$. Then from the construction of P^D we see that

$$P^D(\Delta,\beta) \to P^D(0,\beta),$$

which using (25) implies that $x^k \to x$. In other words, $(\beta, x) \in F(0)$. Since ϕ is uniformly bounded this means that if $\left(\beta^k, x^k\right)_k$ is a sequence of points in $F\left(\Delta^k\right)$ converging to (β, x)

then we have $(\beta, x) \in F(0)$. Therefore, F is upper hemicontinuous.

Suppose $(\beta, x) \in F(0)$ and $\left(\Delta^k\right)_{k\geq 1}$ is any sequence decreasing to 0. Then pick any sequence $\left(\beta^k, x^k\right)_k$ such that $\beta^k = \beta$ for all k and $\left(\beta, x^k\right) \in F\left(\Delta^k\right)$. By construction the bounds in (23) and the expression in (24) reduce to $P^D(0, \beta)$ as $\Delta^k \to 0$. Thus it must be the case that $P^D\left(\Delta^k, \beta\right) \to P^D(0, \beta)$, and therefore, by the continuity of (25), we have $x^k \to x$; hence F is lower hemicontinuous.

The objective function $U^D_\Delta(x, \beta)$ is jointly continuous in the variables (x, β) and the parameter Δ. Since F is both upper and lower hemicontinuous, it is continuous at $\Delta = 0$. Then the maximum theorem immediately implies that the optimal solutions are also upper hemicontinuous at $\Delta = 0$. In other words, as Δ goes to zero all optimal solutions of the constrained maximization problem $\mathcal{P}_\Delta$ approach the unique solution of the problem $\mathcal{P}_0$, which is the limiting problem we solve in the body of the paper. □

A.4 Proofs of Section 6

Proof of Proposition 4

Proof. Let the posterior probability of the tough type at the end of round k be $\hat{\pi}^C_k$. A strategy of Defender comprises finitely many functions x_k for $k = 0, 1, \ldots, K$, mapping from $[0, 1]$ to $[0, 1]$ such that the k^{th} offer is $x_k\left(\hat{\pi}^C_{k-1}\right)$. Let $(\beta_k \mid k = 0, 1, \ldots, K)$ be the corresponding acceptance probability mappings from $[0, 1]$ to $[0, 1]$, such that the offer made in round $k \in \{0, 1, \ldots, K\}$ is accepted with probability $\beta_k\left(\hat{\pi}^C_{k-1}\right)$. The strategy of Challenger decides which offer to accept and with what probability, one for each round; we omit the notation for this.

We prove this by induction on the number of rounds. Let $K = 2$ henceforth and assume that the offer in round 1 is accepted with strictly positive probability β_1. The optimal offer at each state and each history is (i) Markovian, i.e. it depends only on beliefs about the type of Challenger; and (ii) deterministic. If B is the function such that $\beta^* = B\left(\pi^C\right)$ is the unique solution to the equation (15), then it is clear that B is increasing in π^C (strictly increasing unless it has hit 1), and so is the posterior probability $\pi^C / \left(1 - B\left(\pi^C\right)\right)$. Take any candidate equilibrium of the 2-offer game with the equilibrium acceptance functions (β_1, β_2). Clearly, sequential rationality requires that $\beta_2\left(\hat{\pi}^C_1\right) = B\left(\hat{\pi}^C_1\right)$, where $\hat{\pi}^C_1$ is given by $\hat{\pi}^C_1 = \pi^C / (1 - \beta_1) > \pi^C$. The total probability (summed over rounds and types) that conflict will not start is then given by $\bar{\beta} = \beta_1 + (1 - \beta_1)\beta_2$. The posterior at the end of round 2 is

$$\pi_0^C = \frac{\hat{\pi}_1^C}{1-\beta_2} = \frac{\pi^C}{(1-\beta_1)(1-\beta_2)} = \frac{\pi^C}{1-\bar{\beta}}.$$

By the above and using (13), the sequentially rational offer in round 2 is

$$\bar{x} = e^{-r\Delta_{1,2}} \left\{ 1 - \pi^D \left(\frac{\pi^C}{1-\bar{\beta}} \right)^{\frac{\ell^C}{1-\ell^D}} \right\}, \tag{26}$$

where $\Delta_{1,2} \geq 0$ is the time interval between the first and the second offer. Incentive compatibility of normal Challenger then requires that acceptance in any round give the same utility to Challenger. This, together with the fact that $\left(\bar{x}, \bar{\beta}\right)$ satisfies (26), means that the utility of Defender is the utility of a game in which he takes out a mass $\bar{\beta}$ of Challengers (at a cost of $\bar{x}$ per unit mass); this is denoted by $U_0^D\left(\bar{x}, \bar{\beta}; \pi^C, \pi^D\right)$. Let $\left(x^*\left(B\left(\pi^C\right)\right), B\left(\pi^C\right)\right)$ be the optimal pair when we have a single round with the prior π^C. Since $\bar{x} > x^*$ and $\bar{\beta} > B\left(\pi^C\right)$, it follows that

$$U_0^D\left(\bar{x}, \bar{\beta}; \pi^C, \pi^D\right) < U_0^D\left(x^*\left(B\left(\pi^C\right)\right), B\left(\pi^C\right); \pi^C, \pi^D\right).$$

If Defender deviates in round 1 from β_1 and chooses $\beta_1' = 0$ instead, it would have been sequentially rational to make the optimal offer x^* in round 2 and have it accepted with probability β^*. Therefore, if two offers are accepted with positive probability in any equilibrium, it is better to deviate and change the earlier offer to 0, so that the first offer is rejected. Hence there is no such equilibrium: the only possibility is $\beta_1^* = \ldots = \beta_{K-1}^* = 0$. $\square$

A.5 The role of Assumption 2 as $\Delta \to 0$

We solve the conflict part of our model for any positive Δ subject to Assumption 2. Suppose that, contrary to Assumption 2, $\ln \pi_0^C / \ln \bar{\pi}^C = m \in \mathbb{N} \setminus \{1\}$ while $\ln \pi^D / \ln \bar{\pi}^D < m$ (the case $\ln \pi^D / \ln \bar{\pi}^D \in \mathbb{N} \setminus \{1\}$ is symmetric). Lemmas 3 to 5 do not rely on Assumption 2. Therefore, in all equilibria, beliefs move according to (6) from period 2, stage 1 onward. The first part of the proof of Proposition 1, then guarantees that there exists an equilibrium as in Proposition 1: at $t = 1$, Challenger attacks with probability 1 and Defender concedes with probability $1 - \pi^D / \left(\bar{\pi}^D\right)^m$; $\pi_{m-1}^D = \bar{\pi}^D$ and $\pi_m^C = \bar{\pi}^C$. Yet, there exist also other equilibria. In all other equilibria, at $t = 1$, Challenger attacks with probability 1 and Defender concedes with probability p:

$$p \in \left[1 - \frac{\pi^D}{(\bar{\pi}^D)^m}, 1 - \frac{\pi^D}{(\bar{\pi}^D)^{m-1}}\right].$$

Therefore, $\bar{\pi}^C_m = \bar{\pi}^C$ and $\pi^D_{m-1} < \bar{\pi}^D \leq \pi^D_m$. Lemmas 3 to 5 guarantee that these are all the possible equilibria. While this means there are multiple equilibria, it is easy to see from Lemma 1 that

$$\left(1 - \frac{\pi^D}{(\bar{\pi}^D)^{m-1}}\right) - \left(1 - \frac{\pi^D}{(\bar{\pi}^D)^m}\right) \downarrow 0 \text{ as } \Delta \downarrow 0.$$

Therefore in the limit as $\Delta \downarrow 0$, the equilibrium strategies are uniquely determined by those in Proposition 1 and Challenger's continuation payoff from conflict converges to the one given by Remark 1.