ESCALATION AND DELAY IN LONG INTERNATIONAL CONFLICTS*

Aviad Heifetz† Ella Segev‡

June 2002

Abstract

Why do escalations in long international conflicts sometimes hasten the pace of negotiations? And why is it sometimes the case that the resulting terms of agreement were deemed unacceptable to one or both sides before the escalation? We analyze these issues in a game-theoretic setting with asymmetric information, in which the delay a party exercises before it makes an acceptable offer is served to signal credibly its true stand, of which the other side is initially uncertain. The implications of escalation are twofold. First, escalation makes both sides more eager to settle than before, as an agreement would end the increased level of hostilities. Thus escalation increases the "peace pie" – the overall gains from agreement, and one side or both may eventually yield to terms they wouldn't accept at the outset, in order to put an end to the increased level of damages that they suffer. Second, escalation shortens the delay each side needs to exercise in order to credibly portray its genuine stand to the opponent. The combination of these "constructive" effects may tempt one or both sides to escalate. However, it turns out that the larger is the overall increase in violence implied by escalation, the higher are also the chances that the aggressor will eventually regret its decision to escalate.

JEL Classification Code.: C78, P16

*We benefited from insightful comments of Anat Admati, Zeev Maoz, Bruce Bueno de Mesquita, Jim Fearon, Gilat Levy and Motty Perry.

†The School of Economics, Tel Aviv University, heifetz@post.tau.ac.il. Support from the Hammer fund for Economic Cooperation is gratefully acknowledged.

‡The Faculty of Management, Tel Aviv University, segeve@post.tau.ac.il

1 Introduction

Clausowitz's celebrated insight that "war is simply a continuation of political intercourse with the addition of other means"[1] is compatible with several views about the onset of armed conflicts. By some approaches the threat to resort to violence, possibly accompanied by costly escalating measures, is launched in the attempt of improving one's bargaining position over some asset at stake. War erupts when the opponent is in fact relatively strong and decides to challenge the threat, to the dismay of the aggressor. By these accounts war continues but ends diplomacy.[2] In Wagner (2000), the outcome of a limited war may narrow the gap between the belligerents' initially-misaligned beliefs about their chances in a hypothetical all-out war. In particular, the defeated side may update its belief, and consequently yield to a deal it refused at the outset. Here, violence is interwoven in the diplomatic process of bargaining.

Oftentimes, however, asymmetry of information concerns mostly the opponents' reservation values regarding some issue in dispute, and not necessarily their military capabilities in an ultimate war. Still, military escalations sometimes do hasten agreement in prolonged conflicts, *in terms which were deemed unacceptable for one or both sides before the upsurge of violence.*

For example, the 1973 war between Israel and Egypt was not only intensive by itself, but led also to an accelerated arms-race in subsequent years. Nevertheless, the war was shortly followed by a series of disengagement agreements and finally a peace treaty in 1979. In this treaty Israel conceded the entire Sinai peninsula in exchange for peace[3]. These very terms were proposed by Egypt already in 1971[4], but Israel rejected them with mistrust[5].

Similarly, the 1987-1990 Palestinian *Intifada* (upheaval) was soon followed by the Israeli-Palestinian 1993 Oslo Accords. These accords should have led, in a gradual process, to establishing a Palestinian state – an intolerable idea to most of the Israeli public before the *intifada.* However, Israel's initiative in 2000-2001 to *hasten* the process and terminate the conflict right away – suggesting a far-reaching territorial compromise and a *de-facto* division of Jerusalem[6] – was rejected with deep mistrust by the Palestinians[7],

[1] Clausewitz (1976, p. 605)

[2] Banks (1990), Bueno de Mesquita and Lalman (1989), Bueno de Mesquita et al. (1997), Fearon (1994), Morrow (1989), Powell (1987).

[3] Israel conceded even though it had the superior military position at the end of the 1973 war (its troops crossed the Suez canal and were only 101 Kilometers from Cairo, while the Egyptian 3rd Army in Sinai was under Israeli siege (Hertzog 1975)). Moreover, Israel's chances in an absolute war (in which its alleged nuclear superiority would be central) were not undermined by the war. Therefore, it is probably not a change in these assessments *per se* that have led to the breakthrough.

[4] In an official document presented to the UN mediator Jarring on February 15, 1971 (Riad 1981, p. 188)

[5] See e.g. Cohen (1990, pp. 156-159).

[6] Shavit (2001).

[7] Agha and Mallie (2001), Shavit (2001).

and led to the outburst of a second, fierce *Intifada*.

In Ireland, the IRA made it clear during the 1990's that though it favored talks, it was also determined to sustain violence until its demands were met, and most of the negotiators at that period admit that violence had indeed made them more determined to keep the peace process going.[8] The 1998 "Good Friday" agreement eventually granted self-ruling to Northern Ireland in exchange for an appeal to decommission the paramilitary groups. However, when the IRA announced a cease-fire already in August 1994, it was encountered by confusion on the part of the British government, who demanded a unilateral complete decommission of the IRA, and thus forestalled the process.[9]

What is the internal logic of such tragic chains of events? What is the "merit" of mutually-painful escalations in the course of long international conflicts? And why are proposals for mutual concessions confronted with mistrust when they are not preceded by long-enough suffering? In this essay we propose a game-theoretic analysis of this conundrum.

One important implication of escalation is that it *increases the peace pie,* i.e. the overall gains from agreement: Both sides have to endure the increased level of hostilities as long as the conflict is not over, and are therefore willing to give up more than before in order to bring an end to violence. This may create room for agreement where it did not exist before. Moreover, we will show how the extra suppleness of the parties *loosens their incentives* to show resolve by sustaining the hardship of war.

Specifically, at each given level of escalation, delay may be used by one of the parties to *signal credibly* its reservation stand[10]: If it is relatively 'soft' – in the sense that it conceives large gains from agreement and is therefore eager to strike a deal – it will put on the table an acceptable, generous offer *sooner* than if it were relatively 'tough'. In the latter case, in which it believes that the gains from agreement are small anyhow, it would prefer to wait and settle at better terms rather than settle quickly at worse terms.[11] With asymmetric information on both sides, delay may be used both to screen the potential stands of the opponent, as well as to signal one's own resolve.[12] At equilibrium, the parties are initially engaged in a "war of attrition". The tougher one's stand is, the longer it takes until it approaches the negotiation table or makes an offer which reveals its stand. The other party then delays further its offer in order to signal its own stand in a credible way, as above.

A particularly remarkable feature of these models is that the resulting delay tends

[8] Mallie and McKittrick (2001, pp. 83, 224-225). Many violent episodes gave the peace process a push forward. In October 1993, for example, after 23 people were killed in one week, Britain prime minister, John Major said: "The process was on a knife-edge. I think it would have broken down had not the Shankill and Greysteel tragedies intervened" (Mallie and McKittrick (2001, p.128).

[9] The cease-fire ended in February 1996 with an IRA bomb in London and the continuation of violence (Mallie and McKittrick 2001, p.175-202).

[10] For instance, a reservation stand of Israel may consist of the maximal territorial concessions it is in fact willing to yield in a given front in exchange for peace (of some well-defined characteristics).

[11] Admati and Perry (1987), Cramton (1992).

[12] Cramton (1992), Wang (2000).

to infinity as the actual gains from agreement vanish.[13] In other words, different potential 'types' of the same side, with very close degrees of resolve, may exercise extremely different delays until they make an acceptable offer.

Consequently, a small softening in a party's stand may lead to a much shorter delay to agreement. Escalation will therefore hasten the pace of bargaining, because the stands of both opponents will be softened relative to what they used to be before the escalation. As for the terms of the eventual agreement, escalation will naturally favor the aggressor the smaller are the perils of escalation for itself, and the larger they are for its rival.

Summing up the pro-s and con-s, escalation may very well be deemed as promising an expected gain to the aggressor. However, escalation is in fact a gamble which always carries with it the risk of turning sour. This happens when the opponent's true stand is revealed to be rather tough, leading to a long delay and increased suffering before the agreement is reached. If the opponent is particularly tough, the aggressor might even find itself yielding to terms it would not have accepted at the outset, in order to terminate the enduring burden of escalation (caused by its own initiative). The same considerations will also lead some types of the other side to compromise at worse terms than they would initially accept – just as in the above historical examples.

While the equilibrium behavior of the parties in the model imply the timing of resolution, their off-equilibrium behavior in the model is no less revealing. The 1971 peace initiative of Egypt towards Israel, as well as Israel's initiative towards the Palestinians in 2000-2001, each proposed that *both* sides to the conflict make substantial concessions relative to their true stands[14]. The idea behind these initiatives was that it is rather likely that similar concessions will be made soon after the conflict spirals with further violence (as the model predicts), and so it is better for both sides to concede right away.

Unfortunately, such proposals are pre-mature in terms of the model, and constitute off-equilibrium offers which are bound to be rejected. This is because the other side reasons as follows: "If the proposer made us an offer this soon, it must mean that it is rather eager to settle; but in such a case it must also be willing to settle at terms which are even better for us.[15] Alternatively, if the proposer did make us a fair offer (given its true reservation stand, of which we are uncertain), it should have waited longer, and prove to us its resolve by its willingness to suffer longer the misfortunes of the conflict."

[13] This is because in the limit case where the sides have nothing to gain from an agreement, no agreement is ever reached, and so the delay is infinite.

[14] In 1971, Egypt's suggestion implied giving up her role as leading the Arab world in its struggle with Israel. Israel was asked to give up her strategic geopgraphical posture in Sinai.

In 2000-2001, Israel suggested to withdraw from most of the West Bank and Gaza strip, to compensate the Palestinians further with territory in the Negev, and to hand to Palestinian control key parts of old Jerusalem, including the temple mount. In return, the Palestinians were asked to declare an end to the Israeli-Palestinian conflict (i.e. give up their claims over the entirety of Palestine) and to give up their demand for the right of return of all Palestinian refugees to their pre-1948 homes.

[15] Indeed, during the 2000 negotiations with the Palestinians, Israel gradually made more and more territorial concessions in her offers to the Palestinians (Agaha and Mallie 2001, Shavit 2001). However, these concessions never matched what would lead the Palestinians to declare an ultimate end to the conflict, as Israel demanded in return.

In short, a misalignment between the content of the proposal and its timing is bound to entail confusion and mistrust, as was indeed the case in these historical episodes.[16]

In the following section we present our basic model. A potential Aggressor with a known reservation stand is uncertain of the reservation stand of its rival, the Proposer, who should choose how long to wait before it comes up with an acceptable offer to resolve the dispute. As explained above, the delay time will reveal the Proposer's stand in a credible way. The potential Aggressor can choose whether to escalate the conflict at the outset, by causing a flow of damages both to its rival and to itself at known, fixed extents.[17] We show that when the damages inflicted upon the Proposer are large enough relative to those suffered by the Aggressor, the latter may indeed find it worthwhile to escalate. This, however, implies a positive probability that the Aggressor will eventually regret the escalation – when the Proposer reveals its toughness by enduring the hostilities for a long time before it puts a serious offer on the table. Moreover, there is a positive probability that the Aggressor eventually concedes beyond its original red lines, in order to put an end to the increased level of violence. We find that the larger is the overall extent of the hostilities implied by the escalation, the higher is also the probability that the Aggressor regrets it ex post.

Next, we explore a completely symmetric model, in which both sides can choose whether or not to escalate at the outset, while they are each uncertain about the reservation stand of the other. At equilibrium, the tougher types of each side choose not to escalate, while the remaining, softer types escalate. This initial decision therefore conveys some information to the other side. While enduring the resulting level of violence, both sides engage in a "war of attrition," each waiting for the other to approach the negotiation table. The party who is in fact relatively "softer" does so first, and the time it took it do so reveals its stand. The other party then waits further until it comes up with a serious offer which terminates the conflict. This full-fledged model delivers the same insights as does the basic model. In particular, escalation hastens the pace of bargaining, but both sides might end up wishing they would not have escalated at the outset.

[16] In an interview to *The Times* on March 13, 1971 (cited in Cohen 1990, p. 158), the Israeli prime minister Golda Meir affirmed that president Sadat was the first Egyptian leader to say that he was "prepared to make peace." "At least, he said it. But does he mean it?"

Agha and Malley (2001) report that Arafat's main aim in the 2000 Camp David summit was to evade coercion, since he viewed the summit as a coordinated Israeli-American trap, and therefore did not believe that the Israelis actually intend to abide by their own proposals.

[17] In real life, the aggressor cannot be certain in advance regarding the exact extent of these damages, which moreover tend to vary with time. In the current treatise we abstract from these extra dimensions of uncertainty, in order to isolate the interplay between the extent of the damages – whatever these turn to be – and the issue of timing in bargaining with asymmetric information about fundamental stands.

2 The Basic Model

2.1 Signaling Resolve by Delay

Our basic model adds an escalation stage to a bargaining model of Admati and Perry (1987) and Cramton (1992), which we now turn to describe in our context.

Side A has a known stand O. The quantity O represents the flow of welfare for side A at the status quo level of the conflict. In other words, in order for A to accept the terms of an agreement to end the conflict, the level of welfare flow V that side A gets at the agreement should be at least O. In such a case, $V - O$ is the increase of welfare flow for A once the agreement is implemented. We assume that V stems from the value of some tangible assets (like territory or self-ruling rights) that side B yields to A in the agreement.

The exact value P, which side B attaches to ending the conflict, is not known to side A. The quantity P represents the *increase* in the flow of welfare that B would enjoy if the conflict were simply to end, without any concessions on the part of B. Thus, the value V of the tangible assets that B would be willing to yield at an agreement to end the conflict cannot exceed P. In such an agreement, $P - V$ is the increase of welfare flow for side B[18] once the agreement takes effect.[19]

Future welfare flows are discounted at the rate $r > 0$, and we define the *payoff* of each party (from a given path of welfare flows over time) to be *the increase of its present-value of welfare relative to the present-value of indefinite continuation of the status quo.* Thus, if no agreement is ever reached (and the conflict continues at its present level), the payoffs for the sides are zero. If an agreement is signed at time t at the terms V, the payoff for A is

$$U_A = \int_0^t 0re^{-r\tau}d\tau + \int_t^\infty (V - O)\, re^{-r\tau}d\tau = e^{-rt}(V - O) \tag{2.1}$$

while B's payoff is

$$U_B = \int_0^t 0re^{-r\tau}d\tau + \int_t^\infty (P - V)re^{-r\tau}d\tau = e^{-rt}(P - V) \tag{2.2}$$

This is because for both sides the increase in welfare occurs only from time t and on, and from present-day perspective a welfare increase at the future instant τ is discounted by $re^{-r\tau}$.

[18] More generally, we could have assumed that the welfare increase for B is continuously increasing with $P - V$. This would not change the qualitative outcome of the analysis.

[19] To take a simplistic example, suppose the Palestinians (side A) prefer the status quo (the continuation of the Israeli-Palestinian conflict) over establishing an independent state unless they get to do so over at least O percent of the West Bank and Gaza strip. Similarly, let P stand for the maximal territorial concessions Israel (side B) is in fact willing to yield in a peace treaty with the Palestinians. Then if a peace treaty is eventually signed, in which a Palestinian state is established over V percent of the territories, then $V - O$ represents the Palestinian's 'gain' beyond their reservation stand, while $P - V$ represents Israel's 'gain' beyond its actual red lines. In other words, these are the improvements in welfare for the parties once the treaty is implemented.

Side A believes that B's stand, P, is distributed with a strictly positive density function g and a cumulative distribution function G in the interval $[\underline{P}, \overline{P}]$ containing O. Thus, when $P > O$, any agreement $V \in (O, P)$ would be mutually beneficial relative to the status quo. In contrast, there are no acceptable terms for an agreement when $P < O$. We normalize to have $O = 0$ and assume that $[\underline{P}, \overline{P}] = [-1, 1]$. We also assume for now that P is uniformly distributed in $[-1, 1]$, i.e. $G(P) = \frac{P+1}{2}$ and $g(P) = \frac{1}{2}$.

The sides alternate in making offers for an agreement, with a minimal possible elapse t_0 between the offers. We denote by $\delta = e^{-rt_0}$ the discount associated with this minimal elapse.

The first side to offer is B, which we therefore call the Proposer. It has to decide when to make its offer V. If the offer is made at time t and A accepts the offer, the payoffs of the sides are given by (2.1) and (2.2) above. If A turns down the offer, A can make a counter-offer V' to B after the minimal delay t_0. A can delay the counter-offer as long as it likes (possibly indefinitely) beyond t_0. If A made its counter-offer at time t' and it was accepted by B , the payoffs are as above with V replaced by V' and t by t'. The alternating offers scheme repeats itself.

On the separating equilibrium path the Proposer's delay will fully reveal its stand, i.e. the delay is in one-to-one correspondence with the Proposer's stand P. Thus, when the delay is over, the remaining game essentially becomes one of complete information as in Rubinstein (1982), in which the Proposer will offer at equilibrium

$$V(P, O = 0) = \frac{\delta P}{1 + \delta} \tag{2.3}$$

and A will accept immediately.

Denote by $\beta(P)$ the equilibrium delay of a Proposer with stand P. If the Proposer delays its offer by $\beta(P)$ but makes an offer less generous than $V(P, 0)$, A will reject the offer.[20]

To compute the equilibrium delay $\beta(P)$, denote by $P(\Delta)$ the inverse function, which specifies the stand of the Proposer who delays its offer by Δ at equilibrium. The Proposer's problem is therefore to balance between the terms of trade and the losses from delay, by finding a delay $\beta(P)$ which maximizes its payoff

$$U_B(\beta(P)) = e^{-r\beta(P)}(P - V) = \max_{0 \leq \Delta} \left(e^{-r\Delta}(P - V(P(\Delta), 0)) \right) \tag{2.4}$$

where $V(P(\Delta), 0)$ is the offer made by a Proposer who waits Δ – the Rubinstein (1982) subgame-perfect offer (2.3) in the complete-information game between A and B. A longer delay will signal a tougher stand and will be accompanied with a less generous offer, as if saying: "I have less to gain from an agreement, therefore I do not mind waiting longer before offering my terms, if this will convince you I will not give up further."

[20] A could then simply delay its counter-offer indefinitely in order to deter the Proposer from such a deviation, but Cramton (1992) specifies also a "milder" off-equilibrium reaction which still achieves this deterring effect.

We will restrict our analysis to the limit case where $\delta \to 1$, i.e. the minimal time between consecutive offers, t_0, goes to zero.[21] In the limit $V(P(\Delta), 0) \to \frac{P(\Delta)}{2}$, and the gains from agreement are equally shared.

Proposition 1 *The equilibrium delay time (Cramton 1992): On the separating equilibrium path, when $\delta \to 1$ the delay time of a Proposer with stand $P > 0$ tends to*

$$\beta(P) = \begin{cases} -\frac{1}{r}\ln P & \text{For } P \in (0,1] \\ \infty & \text{For } P \in [-1,0] \end{cases} \tag{2.5}$$

Proof. In the appendix. ■

We can see that the delay is logarithmic in the Proposer's type. The type who has the most to gain from an agreement, $P = 1$, will propose immediately, while other types will wait longer the less they have to gain. Ultimately, the types $P \in [-1, 0]$, with whom there is no wedge for a mutually beneficial agreement, will never come up with an offer.

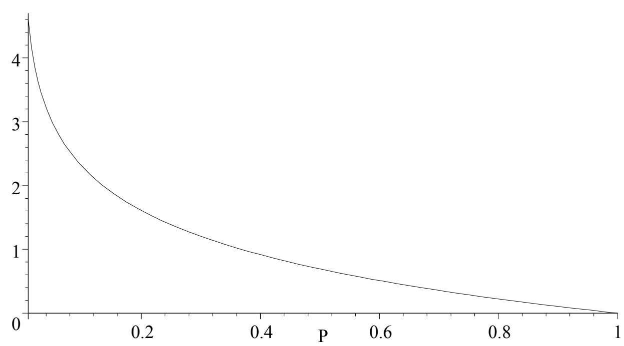

FIG 1: THE DELAY TIME $\beta(P)$ MULTIPLIED BY r

2.2 Escalation

We now add to this bargaining model of Admati-Perry (1987) and Cramton (1992) the possibility of escalation by side A, which we therefore call the Aggressor.[22] Just before the onset of bargaining, A can choose whether or not to escalate, where escalation implies that sides A and B incur the flows of damages $C, D > 0$, respectively, as long as the dispute is not resolved with an agreement.[23] We will assume throughout that $C < D -$

[21] Thus, the assumption $\delta \to 1$ does *not* mean that the proponents tend to be infinitely patient.

The restriction to the limit case $\delta \to 1$ is only in order to simplify the exposition and to avoid cumbersome computations. The qualitative results remain intact also when $\delta < 1$.

[22] Our model is related but different than that of Cramton and Tracy (1992), which dealt with strikes.

[23] Indeed, the Palestinian *intifada* or the ongoing violence in North Ireland are examples of such protracted escalations taking place *in parallel* to the bargaining process.

the Aggressor A suffers less than the damage it inflicts upon B, and that $C + D < 1$, meaning that the overall level of violence is not excessively large.

Again side B is the first to propose. If B makes a proposal of terms V at time t and A accepts it, V is implemented immediately and A's payoff is (recall that we normalize to have $O = 0$):

$$U_A = \int_0^t (-C) r e^{-r\tau} d\tau + \int_t^\infty V r e^{-r\tau} d\tau = (-C)\left(1 - e^{-rt}\right) + V e^{-rt} \tag{2.8}$$

While B's payoff is

$$U_B = \int_0^t (-D) r e^{-r\tau} d\tau + \int_t^\infty (P - V) r e^{-r\tau} d\tau = (-D)\left(1 - e^{-rt}\right) + (P - V) e^{-rt} \tag{2.9}$$

Otherwise, if A turns down the offer, it continues to inflict the damage D upon B while bearing the cost C. After a minimal delay t_0, A gets a chance to choose when and what counter-offer V' to make to B, and so on, as before.

2.2.1 The Implications of Escalation

Escalation turns to have two important effects on the process of bargaining. First of all, the new status quo is obviously worse than before for both sides, and hence they are willing to settle for less than what they would agree before the escalation. The violence makes it more worthwhile for the parties to settle, since it will not only give them at least part of their original aspirations, but it will also put an end to the increased level of hostilities. This is the sense in which escalation increases the "peace pie."

Formally, side A would now prefer any resolution at which it gets $V > -C$ rather than continuing to endure the flow of damages C. Similarly, side B with an original stand P would now prefer any settlement at which it "pays" $V < P + D$ rather than continuing to sustain the flow of damages D (that is, the maximum it is willing to give in exchange for an agreement increases). This means that when the Proposer's stand eventually gets revealed, the Rubinstein (1982) subgame-perfect equilibrium offer that P will make is given by

$$V(P + D, -C) = \frac{\delta\,(P + D) - C}{1 + \delta} \tag{2.10}$$

and this offer will be accepted immediately by the Aggressor.

The second, less obvious effect of escalation is its repercussions for the timing of resolution:

Proposition 2 *(Equilibrium delay given escalation). On the separating equilibrium path, as $\delta \to 1$ the delay of the Proposer as a function of its original stand P tends to*

$$\beta(P) = \begin{cases} -\frac{1}{r}\ln\left(\frac{P+C+D}{1+C+D}\right) & \text{For } P \in (-C - D, 1] \\ \infty & \text{For } P \in [-1, -C - D] \end{cases} \tag{2.11}$$

Proof. In the appendix ■

The resulting delay time is, once again, logarithmic in the difference between the sides' stands, but this time normalized to the new, larger range of B's types who will come up with an acceptable offer in a finite time. While the type $P = 1$ is still the one to make an immediate offer, it is now the case that any type with original stand $P > -(C + D)$ will eventually make an acceptable offer. As explained above, this is because the agreement will put an end to the suffering of the damage flows C, D, and thus any terms between $-C$ and $P + D$ (and in particular the terms (2.10)) are preferred by both sides over the on-going level of escalation.

The important insight of proposition 2 is that once escalation has occurred, the delay time (2.11) of all the Proposer types $P \in (-C - D, 1)$ will be shorter than the delay time without escalation (2.5). *With the increased level of hostilities, it takes less time for the Proposer to signal credibly its true stand to the other party.*

To grasp the intuition for this result, consider any two types $P' > P$. At equilibrium, the tougher-stand type P signals its position by waiting longer than P', but eventually offers to settle at better terms for itself than the terms that P' offers earlier. Given the eventual Rubinstein (1982) subgame-perfect offers once the positions are revealed, how much longer should P wait in order to differentiate itself credibly from P'? In other words, what should be the extra waiting time $\beta(P) - \beta(P')$ so that P' would not be tempted to imitate P's waiting time and offer? The crucial observation is that *this necessary extra waiting-time decreases with the overall level of hostilities* $C + D$. The harsher is the on-going violence, the less tempting it becomes for P' to postpone its offer further just in order to improve the eventual terms of agreement. Consequently, the smaller also becomes the gap in waiting time $\beta(P) - \beta(P')$ which the genuinely-tougher P has to exercise, in order to differentiate its own behavior from one which P' might like to imitate.

Importantly, proposition 2 stems directly from the incentive compatibility constraints of the various Proposer types P, and is therefore *independent of the distribution* G *of the Proposer's reservation stands.*

The following graph depicts the delay time with and without escalation for the case $C + D = \frac{1}{2}$ and $r = 1$

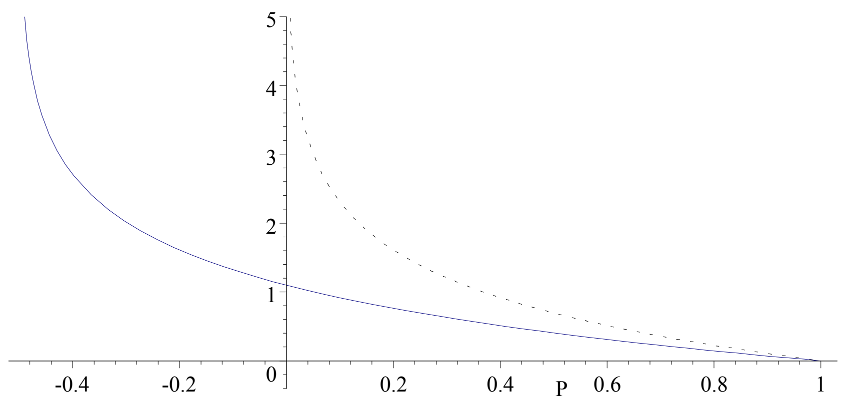

FIG 2: DELAY TIME WITH ESCALATION (solid line) AND WITHOUT (dashed line)

2.2.2 The Aggressor's Gamble

From the point of view of the aggressor, escalation has two conflicting effects. On one hand, escalation will shorten the time to agreement with all the types $P \in (-C - D, 1]$. Moreover, since $C < D$ (the aggressor A suffers less than the damage it inflicts upon B), the terms of the agreement (2.10) will eventually be better for the aggressor than the terms (2.3) without escalation whenever agreement is reached. On the other hand, the Aggressor will have to endure the cost C until the Proposer comes up with the offer (2.11). Unfortunately, this might never happen – in case the Proposer's reservation stand is so tough, $P \in [-1, -C - D]$, that it prefers to endure the increased level of violence indefinitely rather than settle at any acceptable terms.

When will A prefer to escalate?

Proposition 3 *If A maximizes its expected payoff, it will prefer to escalate if*

$$C < \frac{D}{5}$$

in the limit as $\delta \to 1$.

Proof. In the Appendix. ■

However, ex post A might very well regret its decision to escalate. This can happen in one of the following cases:

1. The dispute continues indefinitely at its escalated level.
2. The Aggressor eventually yields to an offer $V < 0$ it would not have accepted at the outset.
3. The eventual terms of the agreement are better for the Aggressor than the terms without escalation, but the damage it has to endure until agreement is reached overshadows the improvement in the final outcome.

How is the overall probability of regret related to the damage levels C, D?

Proposition 4 *The probability that the Aggressor will eventually regret the escalation – carried out by its own initiative – is increasing in the overall level of violence $C + D$. It is also decreasing with the difference $D - C$ of escalation damages that B and A have to endure.*

Proof. *In the Appendix* ■

Escalation is thus a two-edged sword. The larger is the overall level of damages it entails, the higher is also the probability that its initiator will eventually regret it, even if a priori it conceived it to be a worthwhile gamble.

3 A Symmetric Two-Sided Model

In this section we extend the above results to a symmetric, two-sided model. Both sides will now be uncertain about the other's reservation stand, and both would be able to induce escalation at the outset. Our purpose here is to show how the qualitative results above do not depend on the asymmetric roles of Aggressor and Proposer, by presenting an equilibrium with similar properties in the symmetric setting.

The model will be identical to the one above, with the following changes:

1. We will assume that A's position, which we now denote by a, is not known to side B, but rather distributed in the interval $[0, 2]$. To ease the computations we will assume that a is uniformly distributed in $[0, 2]$. We stick with the assumption that B's position, that we denote by b, is uniformly distributed in $[-1, 1]$.
2. Before the bargaining starts, each side (knowing its own reservation stand but not that of the other) can decide whether or not to initiate an escalation, with the flow

of damages C to itself and D to its rival. If both sides escalate, they each endure the damage flows $C + D$ until the conflict is settled.[24]

3. Observing each other's decision whether or not to escalate, each of the opponents should then decide by when to approach the negotiation table if the other party will not yet have done so earlier. Once one side is seated at the negotiation table, the other one has to decide when to approach it as well and make the inaugurating offer. If this offer is accepted, it is implemented immediately. Otherwise the alternating offers scheme repeats itself as in the previous section

We will look for a symmetric equilibrium in which the softer types of each side escalate, while the remaining, tougher ones do not. Namely, we will look for an equilibrium characterized by a threshold $x > \frac{1}{2}$, such that the types of side A in the interval $[0, 1-x)$ escalate, those in $(1-x, 2]$ do not, and the type $a = 1 - x$ escalates with probability $\frac{1}{2}$. Similarly the types of B in $(x, 1]$ escalate, those in $[-1, x)$ do not, and the type $b = x$ escalates with probability $\frac{1}{2}$.

This equilibrium structure is plausible, because the toughest types of each side see little or no prospect for resolution. They would rather stick with the current character of the dispute, and see no point in enduring further damages indefinitely. In contrast, it is the softer types who see mild or even substantial chances for resolution. That's why they are willing to take their chances and escalate, hoping this will reduce the bargaining time and improve their position at the negotiation table.

Our intended equilibrium structure implies that, given the observed decisions about escalation, both sides will know that the *effective* reservation stands of A, B come from some intervals of the form $[\underline{a}, \bar{a}]$, $[\underline{b}, \bar{b}]$, respectively. These intervals are determined by the decisions regarding escalation and the resulting flows of damages. More precisely, these will be

	B escalated	B did not escalate
A escalated	$[\underline{a}, \bar{a}] = [-(C+D), 1-x-(C+D)]$ $[\underline{b}, \bar{b}] = [x+(C+D), 1+(C+D)]$	$[\underline{a}, \bar{a}] = [-C, 1-x-C]$ $[\underline{b}, \bar{b}] = [-1+D, x+D]$
A did not escalate	$[\underline{a}, \bar{a}] = [1-x-D, 2-D]$ $[\underline{b}, \bar{b}] = [x+C, 1+C]$	$(\underline{a}, \bar{a}] = [1-x, 2]$ $[\underline{b}, \bar{b}] = [-1, x]$

(3.1)

As in the previous section, we first turn to describe the timing behavior given the decisions to escalate or not, and then solve backwards to find the equilibrium choices regarding escalation.

[24] In principle, we could have allowed for several consecutive stages in which escalation can be initiated or reciprocated. This, however, would create a preliminary stage of mutual signaling about resolve by the escalation decisions themselves, while the focus of the current paper is on *delay* as a signaling device. This, of course, is not to say that escalation is not useful or relevant as a signaling measure. Rather, our aim here is to disentangle the effect of signaling by delay from other signaling procedures.

3.1 War-of-Attrition Bargaining

Given any prior combination of decisions by the rivals to escalate or not, we will adopt the equilibrium behavior suggested by Wang (2000) for the bargaining stage. Given a pair of intervals $[\underline{a}, \bar{a}]$, $[\underline{b}, \bar{b}]$ of reservation stands, Wang (2000) analyses a host of equilibria for the bargaining stage, one of which is symmetric.

The equilibrium starts with a "war of attrition," in which each side waits for the other to approach the negotiation table first. This is because, at equilibrium, the second to approach is the one who will make the acceptable offer, and at the eventual Rubinstein (1982) subgame-perfect equilibrium the offeror gets a larger share of the gains from agreement.

At equilibrium, the types $\underline{a}$ and $\bar{b}$ approach the table without delay. In the sequel, the softer of the types of each side do so before the tougher types. The timing of $a \in [\underline{a}, \bar{a}]$ and $b \in [\underline{b}, \bar{b}]$ is

$$t_A(a) = -\frac{1}{r} \ln \frac{((\bar{b} + \underline{a}) - 2a)}{(\bar{b} - \underline{a})} \tag{3.2}$$

$$t_B(b) = -\frac{1}{r} \ln \frac{(2b - (\bar{b} + \underline{a}))}{(\bar{b} - \underline{a})} \tag{3.3}$$

with the reverse functions

$$A(t) = \frac{1}{2}(1 - e^{-rt})(\bar{b} - \underline{a}) + \underline{a} \tag{3.4}$$

$$B(t) = \frac{1}{2}(1 + e^{-rt})(\bar{b} - \underline{a}) + \underline{a} \tag{3.5}$$

This staggering implies that for every given period of time by which nobody is yet at the table, each side understands that the type of the other comes from a smaller interval of tougher types. This induces the next-in-line types to approach the table, because by then they know that further delay, and the chance it brings with it to be the second to approach, would not compensate for the losses due to further delay.

Once one side has approached, its position becomes common knowledge by the delay it exercised. The other side then delays further its approach, as a function of its position, by

$$t_A(a)|_{B=b} = -\frac{1}{r} \ln \frac{b - a}{2b - (\bar{b} + \underline{a})} \tag{3.6}$$

$$t_B(b)|_{A=a} = -\frac{1}{r} \ln \frac{b - a}{(\bar{b} + \underline{a}) - 2a} \tag{3.7}$$

Thus, the overall time until both sides are at the table is given by

$$t(a, b) = \begin{cases} -\frac{1}{r} \ln \frac{(b-a)}{(\bar{b}-\underline{a})} & b > a \\ \infty & b \leq a \end{cases} \tag{3.8}$$

(Remarkably, this total delay time results also in all the other, non-symmetric equilibria with a similar structure that Wang (2000) introduces.)

Thus, when both sides are at the table, the delays they exercised have already revealed at equilibrium their effective positions. At this point Wang (2000) assumes, by definition, that the Rubinstein (1982) offer is then implemented. This tends to be

$$V(a,b) = \frac{a+b}{2} \tag{3.9}$$

as $\delta \to 1$. However, the more elaborate equilibrium structure in Cramton (1992) can be applied here, to show how the second to approach may be induced to make this offer at equilibrium by its beliefs about its rival's reaction in case of deviation.[25]

The Wang (2000) equilibrium implicitly assumes, that if there is a finite time by which a party should have approached[26] but it had failed to do so, no agreement is ever implemented. We will adopt this convention here. More precisely, we will assume that, at equilibrium, the preliminary decisions about escalation make the parties believe that the ranges of stands are given by table (3.1) above. Thus, any excessive delays, that could have corresponded to stands that protrude these intervals, would result in the lack of any settlement.

3.2 Escalation

Given the above behavior at the bargaining stage, we now look for conditions under which we will have a unique equilibrium characterized by the threshold x, as detailed at the beginning of this section. In such an equilibrium, each type should carry out the escalatory behavior meant for it – escalate (E) or not escalate (NE). First of all, this equilibrium should have the property that

The type $a = 1 - x$ of A is just indifferent between escalating or not. (Indifference)

By the symmetry of the construction, this condition will be mathematically equivalent to the condition that the type $b = x$ is indifferent between the two options. This equality constraint will pin down the candidate $x(C, D)$ for the threshold as a function of the damage levels C, D.

The other incentive compatibility constraints say that

Types $a \in [0, 1 - x)$ prefer to escalate rather than not (IC1)

and

Types $a \in (1 - x, 2)$ prefer not to escalate rather than to escalate (IC2)

[25] In fact, the setting and the equilibrium of Wang (2000) is a simplification of that of Cramton (1992), who assumes that also the first to approach makes an offer. That asumption makes the equilibrium structure more intricate.

[26] This is the case when $\bar{a} < \underline{b}$

Analogous conditions should obtain for side B, but here again the mathematical conditions will be identical to those for (IC1) and (IC2), due to the symmetry of the equilibrium.

Notice, first, that if a type $a \in [0, 1-x)$ ever decides not to escalate (contrary to the equilibrium behavior), B will believe it comes, in fact, from the interval $[1-x, 2]$. Hence, a should better imitate the delay time of one of the types in $[1-x, 2]$, as otherwise no agreement will ever be reached, by assumption.

Lemma 1 *If $a \in [0, 1-x)$ were not to escalate, it would fare best imitating the delay time of $1-x$ given that side B does adhere to its own equilibrium strategy.*

Proof. In the appendix ■

Similarly, if $a \in (1-x, 2]$ contemplates to disregard the equilibrium recommendation and escalate, it should better plan what type in $[0, 1-x]$ to imitate regarding the delay time in the bargaining stage

Lemma 2 *If $a \in (1-x, 2]$ were to escalate, it would fare best imitating the delay time of $1-x$ given that side B adheres to its own equilibrium strategy.*

Proof. In the appendix ■

These two lemmata pin down the best off-equilibrium behavior given a deviation at the escalation stage. The incentive-compatibility constraints are thus well defined. They consist of two inequalities which should obtain at equilibrium, and relate C, D, and $x(C, D)$.

Unfortunately, the three conditions cannot be solved explicitly so as to isolate $x(C, D)$ or to isolate the restrictions on C as a function of D, because they involve high-degree polynomials. Nevertheless, the equilibrium threshold $x(C, D)$ can be solved numerically given particular values of (C, D). We have done so for a wide range of values of (C, D), and verified that the incentive compatibility constraints (IC1) and (IC2) obtain. For example, it turns out that for $C = 0.1$ and $D = 0.5$, the unique equilibrium of the above structure is with the threshold $x = 0.832\,53$.

Since all three constraints vary smoothly with C, D, existence and uniqueness will typically be preserved at some open neighborhood of parameters for which the equilibrium exists and is unique.

Theorem 1 *There is an open range of parameter values (C, D) for which there is a unique threshold equilibrium of the above structure.*

Proof. In the Appendix ■

Here again the parties will sometimes regret ex post their decision to escalate. Most notably, when both sides escalate, the eventual terms of agreement are the same as when they do not escalate (due to the symmetry) but until agreement is reached they suffer the cost of escalation and its damages. For example, the types $a = 1 - x$ and $b = x$, if they escalate, will trade at a price that tends to $\frac{1}{2}$ as $\delta \to 1$, and their payoffs will tend to be

$$U_{a=1-x} = U_{b=x} = -(C+D)(1-e^{-rt(1-x-C-D,x+C+D)}) + \left(\frac{1}{2} - (1-x)\right) e^{-rt(1-x-C-D,x+C+D)}$$

$$= -(C+D)\left(1 - \frac{2x-1+2(C+D)}{1+2(C+D)}\right) + \left(\frac{1}{2} - (1-x)\right)\left(\frac{2x-1+2(C+D)}{1+2(C+D)}\right)$$

$$= -(1-x)\left(1 - \frac{2(1-x)}{1+2(C+D)}\right) + x - \frac{1}{2}$$

which is *decreasing* in the overall level of hostilities $C + D$.

4 Conclusion

We have demonstrated how escalation may speed up the path to resolution on account of two factors. First, the flow of damages that follows escalation makes both sides more eager to settle than before, and the "peace pie" – the range and overall gains of acceptable agreements – is enlarged. Second, it loosens the incentives of the sides to sustain further the burden of the conflict, when they try convey to each other their true positions in a credible way. We hope that these insights may provide a relevant perspective on the interplay between violence and diplomacy.

Notably, the same effects would follow if an external mediator were to offer in advance to both sides some tangible compensation if and when they reach an agreement. The compensation may have the form of financial aid, acceptance of refugees as immigrants, deployment of a peace-keeping force etc.. Here again the "peace pie" would be enlarged – this time in a constructive fashion, and the incentives of the opponents to show resolve by delay will be weakened. Unfortunately, not too often does a third party have a large enough interest in resolving the conflict so as to offer sufficient such compensation. Moreover, the extent of the promised aid might itself become the issue of tacit bargaining. Still, the analysis we presented suggests that a long dispute, in which the sides seem to be locked in their positions, may get a positive turn when such a conditional aid is offered.

5 Appendix

Proof of Proposition 1: The Proposer B has to choose the delay $\beta(P)$ which maximizes its payoff:

$$U_B(\beta(P)) = e^{-r\beta(P)}(P - V) = \max_{0 \leq \Delta} e^{-r\Delta}(P - V(P(\Delta), 0))$$

Where in the limit when $\delta \to 1$:

$$V(P(\Delta), O = 0) = \frac{P(\Delta)}{2}$$

Thus

$$U_B(\Delta) = e^{-r\Delta}(P - V(P(\Delta), 0)) = e^{-r\Delta}\left(P - \frac{P(\Delta)}{2}\right)$$

We take the derivative with respect to Δ and solve

$$\frac{d}{d\Delta}U_B(\Delta) = -re^{-r\Delta}\left(P - \frac{P(\Delta)}{2}\right) - \frac{1}{2}e^{-r\Delta}\frac{dP(\Delta)}{d\Delta} = 0$$

In equilibrium the maximum is achieved for $\Delta = \beta(P)$ and $P(\beta(P)) = P$ thus we get

$$\frac{dP(\Delta)}{d\Delta} = -rP$$

and

$$\frac{d\Delta}{dP} = -\frac{1}{rP}$$

With the initial condition that $\beta(P = 1) = 0$ i.e. that the most impatient type will offer immediately. This condition holds since if in equilibrium this type, who offers first, chooses a positive delay $t > 0$ then offering the same offer at $t = 0$ off the equilibrium path will give him a higher payoff.

Integrating then gives

$$\beta(P) = -\int \frac{1}{rP}dP = -\frac{1}{r}\ln P$$

■

Proof of Proposition 2: Suppose that the Aggressor escalated. An agreement is feasible iff $P + D > 0 - C \Leftrightarrow P > -C - D$. Side A's flow of welfare until an agreement is reached is $(-C)$ and side B's maximum terms in which it is willing to settle are now $P + D$. If $P \leq -C - D$ there are no possible gains from an agreement and the sides will continue to endure the damages forever.

A Proposer B of type $P > -C - D$ has to choose the delay $\beta(P)$ which maximizes its payoff.

$$U_B = \max_{0 \leq \Delta}\left(\int_0^\Delta (-D)\, re^{-r\tau}d\tau + \int_\Delta^\infty (P - V(P(\Delta), 0))re^{-r\tau}d\tau\right)$$

where in the limit when $\delta \to 1$:

$$V(P(\Delta), O = 0) = \left(\frac{P(\Delta) + D - C}{2}\right)$$

We take the derivative of $U_B(\Delta)$ with respect to Δ and solve

$$\frac{d}{d\Delta}U_B(\Delta) = -Dre^{-r\Delta} - re^{-r\Delta}\left(P - \frac{P(\Delta) + D - C}{2}\right) - \frac{1}{2}\frac{dP(\Delta)}{d\Delta}e^{-r\Delta} = 0$$

In equilibrium the maximum is achieved for $\Delta = \beta(P)$ and $P(\beta(P)) = P$. Thus we get

$$\frac{dP(\Delta)}{d\Delta} = -r\left(P + D + C\right)$$

and

$$\frac{d\Delta}{dP} = -\frac{1}{r\left(P + D + C\right)}$$

Again we have the initial condition that $\beta(P = 1) = 0$, i.e. that the most impatient type will offer immediately. Integrating gives

$$\beta(P) = -\int \frac{1}{r\left(P + D + C\right)} dP = -\frac{1}{r} \ln\left(\frac{P + D + C}{1 + D + C}\right)$$

■

Proof of Proposition 3: Without escalation, the expected ex-ante payoff for A as $\delta \to 1$ tends to (recall that P is uniformly distributed in the interval $[-1, 1]$)

$$U_A = \int_0^1 V(P, 0) e^{-r\beta(P)} g(P) dP = \int_0^1 \frac{P}{2} P \frac{1}{2} dP = \frac{1}{12}$$

.

Substituting (2.10) and (2.11) into (2.8), and averaging over the Proposer's types $P \in [-1, 1]$, we find that if A chooses to escalate, its expected ex ante payoff tends to

$$U_A^e = \int_{-D-C}^1 \left(\int_0^{-\frac{1}{r}\ln\left(\frac{P+C+D}{1+C+D}\right)} (-C) r e^{-r\tau} d\tau + \int_{-\frac{1}{r}\ln\left(\frac{P+C+D}{1+C+D}\right)}^{\infty} \frac{(P + D) - C}{2} r e^{-r\tau} d\tau \right) \frac{1}{2} dP$$
$$+ \int_{-1}^{-D-C} \left(\int_0^{\infty} (-C) r e^{-r\tau} d\tau \right) \frac{1}{2} dP = \frac{1}{12} \left(1 + D + C\right)^2 - C$$

as $\delta \to 1$. Side A will therefore choose to escalate if $U_A^e > U_A$ i.e. if

$$\frac{1}{12} \left(1 + D + C\right)^2 - C > \frac{1}{12}$$

or equivalently

$$C < 5 - D - \sqrt{(25 - 12D)}$$

In particular, A will prefer to escalate if

$$C < \frac{D}{5} \leq 5 - D - \sqrt{(25 - 12D)}.$$

■

Proof of Proposition 4:

1. With probability $\frac{1-C-D}{2}$, P turns to be in the range $[-1, -C-D]$, in which case the dispute continues indefinitely at its escalated level.

2. With probability C, P turns to be in the range $(-C-D, C-D)$. In such a case an agreement is reached after a finite delay but the Aggressor yields to an offer it would not have accepted at the outset, because then

$$V = \frac{(P+D)-C}{2} < 0$$

3. With probability $\frac{1}{2}\left(1-\sqrt{(D-C)\left(1+\frac{1}{C+D}\right)}+D-C\right)$, P turns to be in the range $\left[C-D, 1-\sqrt{(D-C)\left(1+\frac{1}{C+D}\right)}\right)$. In this case the eventual terms of the agreement are better for the Aggressor than the terms without escalation, but the damage it has to endure until agreement is reached overshadows the improvement in the final outcome:

$$\int_0^{-\frac{1}{r}\ln\left(\frac{P+C+D}{1+C+D}\right)}(-C)re^{-r\tau}d\tau+\int_{-\frac{1}{r}\ln\left(\frac{P+C+D}{1+C+D}\right)}^{\infty}\frac{(P+D)-C}{2}re^{-r\tau}d\tau<\int_{-\frac{1}{r}\ln P}^{\infty}\frac{P}{2}re^{-r\tau}d\tau$$

i.e.

$$\left(1-\frac{P+C+D}{1+C+D}\right)(-C)+\frac{(P+D)-C}{2}\left(\frac{P+C+D}{1+C+D}\right)<\frac{P^2}{2}$$

Overall, the Aggressor regrets the escalation ex post when P turns to be in the range $\left[-1, 1-\sqrt{(D-C)\left(1+\frac{1}{C+D}\right)}\right)$, i.e. with probability

$$1-\frac{1}{2}\sqrt{(D-C)\left(1+\frac{1}{C+D}\right)}$$

This probability is *increasing* in the total level of violence $C+D$, though, as could be expected, it decreases with the difference $D-C$ between the damages B and A have to sustain due to the escalation. ■

Proof of Lemma 1: Suppose that a type $a \in [0, 1-x)$ decides not to escalate (contrary to its equilibrium behavior). We know that if a chooses to wait less than what type $1-x$ would have waited, B will not negotiate with it at all (it follows from B's off the equilibrium path beliefs). Thus a should better imitate the delay time of one of the types $k \in [1-x, 2]$ (actually $k \in [1-x, 1+C+D]$ since a would never choose to imitate a type k for which $k \in [1+C+D, 2]$ because then it would never reach an agreement with side B). If a sees that side B has escalated it believes that $b \in [x, 1]$ and that its own expected payoff is (the *effective* reservation stand of a is $k-D$ and of b is $b+C$):

$U^{NE}_{a\in[0,1-x)}(k)|_{b\in[x,1]} = \left(\frac{2}{1-x}\right)\left(\int_x^{\max(k-C-D,x)}(-D)f(b)db\right.$
$\left.+\int_{\max(k-C-D,x)}^1\left(\left(\frac{(b+C)+(k-D)}{2}-a\right)\frac{(b+C)-(k-D)}{C+D+x}-D\left(1-\frac{(b+C)-(k-D)}{C+D+x}\right)\right)f(b)db\right)$
This is a continuous non increasing function of k and its maximum in the interval $[1-x, 1+C+D]$ is achieved at $k=1-x$. Thus in this case, if a saw B escalating, the best it can do is imitate $k=1-x$, and then its expected payoff would be

$$U^{NE}_{a\in[0,1-x)}(1-x)|_{b\in[x,1]}$$
$$= \frac{1}{6}\frac{-2x^2+7x-2+3(C+D)(1-2a)+3a(1-3x)}{x+C+D} - \frac{1}{2}(D-C)$$

If on the other hand, a sees that side B did not escalate, a believes that $b\in[-1,x]$ and that its own expected payoff is (the *effective* reservation stand of a is k and of b is b):

$$U^{NE}_{a\in[0,1-x)}(k)|_{b\in[-1,x]} = \left(\frac{2}{1+x}\right)\left(\int_{-1}^{\min(k,x)}0f(b)db + \int_{\min(k,x)}^{x}\left(\frac{b+k}{2}-a\right)\frac{b-k}{2x-1}f(b)db\right)$$

This is also a continuous non increasing function of k and its maximum in the interval $[1-x, 1+C+D]$ is achieved at $k=1-x$. Thus in this case also the best a can do is imitate $k=1-x$, and then its expected payoff would be

$$U^{NE}_{a\in[0,1-x)}(1-x)|_{b\in[-1,x]} = \left(\frac{2}{1+x}\right)\left(-\frac{1}{2}\left(x-\frac{1}{2}\right)a-\frac{1}{6}x^2+\frac{5}{12}x-\frac{1}{6}\right)$$

Then we see that no matter what B has done, a chooses to imitate $k=1-x$ and its expected payoff is therefore:

$$U^{NE}_{a\in[0,1-x)}(1-x)$$
$$= U^{NE}_{a\in[0,1-x)}(1-x)|_{b\in[x,1]}\times\Pr(b\in[x,1]) + U^{NE}_{a\in[0,1-x)}(1-x)|_{b\in[-1,x]}\times\Pr(b\in[-1,x])$$
$$= \frac{1}{12}(1-x)^2\frac{3a-2(1-x)}{C+D+x}+\frac{1}{4}(1-x)(1-(D-C))-\frac{1}{4}a+\frac{1}{12}(2x-1)(2-x)$$

■

Proof of Lemma 2: Suppose that $a\in(1-x,2]$ chooses to escalate (contrary to its equilibrium behavior). Then side B believes that $a\in[0,1-x]$. If a escalates but waits longer than what $a=1-x$ would have waited, B would never settle at equilibrium. Thus the best a can do is imitate also the delay of some $k\in[0,1-x]$. Thus if a sees that B has escalated its expected payoff is (the *effective* reservation stand of a is $k-C-D$ and of b is $b+C+D$):
$U^{E}_{a\in(1-x,2]}(k)|_{b\in[x,1]} = \left(\frac{2}{1-x}\right)\times\left(\int_x^{\max(k-2(C+D),x)}(-(C+D))f(b)db\right.$
$\left.+\int_{\max(k-2(C+D),x)}^1\left(\left(\frac{b+k}{2}-a\right)\frac{(b+C+D)-(k-C-D)}{1+2(C+D)}-(C+D)\left(1-\frac{(b+C+D)-(k-C-D)}{1+2(C+D)}\right)\right)f(b)db\right)$

This is a continuous non decreasing function of k and its maximum in the interval $[0,1-x]$ is achieved when $k=1-x$. Thus in this case, if a saw B escalating, the best it can do is imitate $k=1-x$, and then its expected payoff would be

$$U^E_{a\in(1-x,2]}(1-x)|_{b\in[x,1]} = -\frac{1}{6}\frac{2x^2+9xa-7x+6(C+D)(2a-x)+2-3a}{1+2C+2D}$$

If, on the other hand, it sees that B did not escalate its expected payoff is (the *effective* reservation stand of a is $k-C$ and of b is $b+D$):

$U^E_{a\in(1-x,2]}(k)|_{b\in[-1,x]} = \left(\frac{2}{1+x}\right)\left(\int_{-1}^{\min(k-C-D,x)}(-C)\,f(b)db\right.$
$\left.+\int_{\min(k-C-D,x)}^{x}\left(\left(\frac{b+D+k-C}{2}-a\right)\frac{b+D-(k-C)}{x+C+D}-C\left(1-\frac{b+D-(k-C)}{x+C+D}\right)\right)f(b)db\right)$

This function is also a continuous non decreasing function of k and its maximum in the interval $[0,1-x]$ is achieved when $k=1-x$. Thus in this case also the best a can do is imitate $k=1-x$, and then its expected payoff would be

$U^E_{a\in(1-x,2]}(1-x)|_{b\in[-1,x]} = \frac{1}{6(x+C+D)(1+x)}((2-x)(2x-1)^2+12ax(1-x-C-D)$
$+3x((D-C)+(C+D)(1-x))-3a(1-2C-2D)$
$+(C+D)((C+D)(C+D-3)+3(D-C)-3)-3a(C+D)^2)+\frac{1}{2(1+x)}x(D-C)$

Therefore a will imitate $k=1-x$ and its expected payoff will be

$U^E_{a\in(1-x,2]}(1-x)=$
$U^E_{a\in(1-x,2]}(1-x)|_{b\in[x,1]}\times\Pr(b\in[x,1])+U^E_{a\in(1-x,2]}(1-x)|_{b\in[-1,x]}\times\Pr(b\in[-1,x])=$
$\frac{1}{12}\frac{1}{(1+2S)(x+S)}(2-11x-13x^3+21x^2+2x^4+2S^4-3S^2(3+2S)+S^3-S+3aS(1-2S)(1+S)$
$-3xaS(x+4+4S)-6xS(1-x)(1-2x-2S)-3a-3ax(1-x)(3x-5))+\frac{1}{4}(1+x)R$

where $S=C+D$ and $R=D-C$ ■

Proof of Theorem 1: We can state the three conditions on the connections between C,D and $x(C,D)$ in their implicit form (where $S=C+D$, $R=D-C$).

1. Indifference: The type $a=1-x$ of A is just indifferent between escalating or not.

$$U^E_{a=(1-x)} = U^{NE}_{a=(1-x)} \Leftrightarrow$$
$$0=-6R(x+S)(1+2S)+(1-x)\left(2-7x^3+16x^2-10x\right)-2S^4$$
$$+(-6x+11)S^3+\left(7x-4x^2+8\right)S^2+\left(-2+16x-13x^2+3x^3\right)S$$

2. IC1: Types $a\in[0,1-x)$ prefer to escalate rather than not

$$U^E_{a\in[0,1-x)} > U^{NE}_{a\in[0,1-x)} = U^{NE}_{a\in[0,1-x)}(1-x) \Leftrightarrow$$
$$0<-\frac{1}{12(x+S)}a^3+\frac{1}{4}\left(1+\frac{1-x}{1+2S}\right)a^2-\frac{1}{4}(1-x+S)\left(1+\frac{(1-x)^2}{(S+x)(1+2S)}\right)a$$
$$+\frac{1}{12}\frac{(1-x)(x^2-5x+2)-S^2(3-S)-S(1-6R)-3x(S-2R)}{x+S}$$
$$-\frac{1}{6}x(1-x)+\frac{1}{12}\frac{1-x^3+6x(1-x)S}{1+2S}$$

3. IC2: Types $a \in (1-x, 2]$ prefer not to escalate rather than escalate

$$U^{NE}_{a\in(1-x,2]} > U^{E}_{a\in(1-x,2]} = U^{E}_{a\in(1-x,2]}(1-x) \Leftrightarrow$$

$$\begin{cases} 0 < \frac{1}{12(2x-1)}(x-a)^3 + \frac{1}{4}\frac{1-x}{x+S}a^2 + W & \text{for } \quad 1-x < a \leq x \\ 0 < \frac{1}{4}\frac{1-x}{x+S}a^2 + W & \text{for } \quad x < a \leq x+S \\ 0 < -\frac{1}{12(x+S)}a^3 + \frac{1}{4}\frac{1+S}{x+S}a^2 - \frac{1}{4}(x+S)a + \frac{1}{12}(x+S)^2 + W & \text{for } \quad x+S \leq a \leq 1+S \\ 0 < -\frac{1}{4}\left(\frac{S^2+3xS-2S+x^2-1}{x+S} + R\right)a + \frac{1}{12}(x+S)^2 & \\ +\frac{1}{12}(1+S)\frac{2S^2+3xS-2S-1}{x+S} + \frac{1}{4}R(1+S) + W & \text{for } \quad 1+S \leq a \leq 2 \end{cases}$$

Where

$$W = W(a,x,S,R) = -\frac{1}{4}\frac{3x^2-9x-6xS-2S^2+3S+5}{1+2S}a - \frac{1}{12}\frac{S\left(S^2-3S-6\right)}{(x+S)}$$
$$-\frac{1}{12}\frac{1-12x^3-10x^3S-3xS+18xS^2-11x+21x^2+2x^4+18x^2S-12x^2S^2}{(1+2S)(x+S)} - \frac{1}{2}R$$

Since all three constraints involve expressions which are smooth as a function of C, D, existence and uniqueness will typically be preserved at some open neighborhood of the parameters C, D for which the equilibrium exists and is unique. For example, for $C = 0.1$ and $D = 0.5$ (that is, $S = 0.6$ and $R = 0.4$) the indifference condition **1.** above gives $x = 0.832\,53$ as the unique solution in the relevant range $[\frac{1}{2}, 1]$. At these values, the derivative of the indifference condition **1.** with respect to x is different than zero. The implicit function theorem then guarantees that there is an open neighborhood $\mathcal{O}$ of $(C, D) = (0.1, 0.5)$ with a unique, smooth solution $x(C, D) \in [\frac{1}{2}, 1]$ of the indifference condition for $(C, D) \in \mathcal{O}$. Moreover, since the strict inequalities IC1 and IC2 obtain for $(C, D) = (0.1, 0.5)$ and $x = 0.832\,53$, they will continue to obtain for (C, D) and $x(C, D)$ in some open subset of $\mathcal{O}$. ■

6 References

Admati A. R. and M. Perry (1987), "Strategic Delay in Bargaining", *Review of Economic Studies* **54**, pp. 345-364.

Agha, H. and R. Malley (2001), Camp David: The Tragedy of Errors, *The New York Review of Books*, August 9, 2001 http://www.nybooks.com/articles/14380

Banks, J.S. (1990), "Equilibrium Behavior in Crisis Bargaining Games", *American Journal of Political Science* **34**(3), pp. 599-614

Bueno de Mesquita, B. and D. Lalman (1989), War and Reason, *Yale University Press*

Bueno de Mesquita, B., J.D. Morrow and E.R. Zorick (1997), "Capabilities, Perception and Escalation", *The American Political Science Review* **91**(1), pp. 15-27 von Clausewitz, Carl (1832/1976), On War, Edited and Translated by Michael Howard and Peter Paret, *Princeton: Princeton University Press.*

Cohen, R. (1990), Culture and Conflict in Egyptian-Israeli Relations – a Dialogue of the Deaf, *Indiana University Press.*

Cramton, P.C. (1992), "Strategic Delay in Bargaining with Two Sided Uncertainty", *Review of Economic Studies* **59**, pp. 205-225.

Cramton, P.C. and J.S. Tracy (1992), "Strikes and Holdouts in Wage Bargaining: Theory and Data", *The American Economic Review* **82**(1), pp. 100-121.

Fearon, J.D. (1994), "Domestic Political Audiences and the Escalation of International Disputes", *The American Political Science Review* **88**, pp. 577-592.

Hertzog, Chaim (1975), The War of Atonement, *Little Brown & Company* (re-issued in 1998 by *Stackpole Books*).

Mallie, E. and D. Mckittric (2001), Endgame in Ireland, *Hodder & Stoughton*

Morrow, J.D. (1989), "Capabilities, Uncertainty and Resolve: A Limited Information Model of Crisis Bargaining", *American Journal of Political Science* **33**(4), pp. 941-972

Powell, R. (1987), "Crisis Bargaining, Escalation and MAD", *The American Political Science Review* **81**, pp. 717-735

Riad, M. (1981), The Struggle for Peace in the Middle East, *Quartet Books, London.*

Rubinstein A. (1982), "Perfect Equilibrium in a Bargaining Model", *Econometrica* **50**, pp. 97-110.

Shavit, A. (2001), "End of a Journey", an interview with Shlomo Ben Ami in *Haaretz Magazine*, September 14, 2001,
http://www.haaretzdaily.com/hasen/pages/ShArt.jhtml?itemNo=74353&contrassID=2&subContrassID=14&sbSubContrassID=0&listSrc=Y

Wagner, R.H. (2000), "Bargaining and War", *American Journal of Political Science* **44**(3), pp. 469-484

Wang, R. (2000), "Separating equilibria in a continuous-time bargaining model with two-sided uncertainty", *International Journal of Game Theory* **29**, pp. 229-240