

# Conflict and Deterrence under Strategic Risk\*

Sylvain Chassang

Gerard Padró i Miquel

Princeton University

London School of Economics and NBER

April 8, 2008

## Abstract

We examine the mechanics of deterrence and intervention when fear is a motive for conflict. We contrast results obtained in a complete information setting, where coordination is easy, to those obtained in a setting with strategic risk, where players have different assessments of their environment. These two strategic settings allow us to define and distinguish predatory and preemptive incentives as determinants of conflict. We show that while weapons have an unambiguous deterrent effect under complete information, this does not hold anymore under strategic risk. Rather, we find that increases in weapon stocks can have a non-monotonic effect on the sustainability of peace. We also show that under strategic risk, inequality in military strength can actually facilitate peace and that anticipated peace-keeping interventions may improve incentives for peaceful behavior.

KEYWORDS: conflict, cooperation, deterrence, intervention, global games, exit games.  
JEL classification codes: D74, C72, C73

---

\*We thank seminar participants at the AEA meetings, Sussex, Université Libre de Bruxelles and the Polarization and Conflict group meeting at LSE for many useful comments. All remaining errors are, of course, our own.

# 1 Introduction

The usual rationale for deterrence is closely related to the rationale behind grim trigger punishment in a repeated prisoners' dilemma. Imagine two neighboring groups that repeatedly decide whether to be peaceful or to launch a surprise attack on the other. A peaceful equilibrium can only be sustained if the short run gains from a surprise attack are balanced by the long run costs of triggering conflict. The logic of deterrence is that as groups accumulate weapons, the cost of conflict increases thereby improving incentives for peaceful behavior. This reflects the idea frequently highlighted in the literature on repeated games that harsher punishments should improve incentives for cooperation.<sup>1</sup>

Although the argument for deterrence is simple and convincing, evidence for the effectiveness of deterrence is less than conclusive. On the one hand, there is a general agreement on the fact that nuclear weapons largely contributed to the absence of direct confrontation in the Cold War.<sup>2</sup> On the other hand, there is an equally wide agreement that the proliferation of semi-automatic weapons is fuelling the chronic civil wars that plague Africa.<sup>3</sup> Why do the intuitions we obtain from a standard repeated prisoners' dilemma seem to hold in some settings but not in others? This paper attempts to shed some light on this mixed evidence by taking seriously the idea of strategic risk.

We model conflict as a dynamic exit game. In each period, players decide whether to be peaceful or attack. When both players choose to be peaceful, they enjoy the economic benefits of peace and the game moves to the next period. However, if one of the players attacks, conflict begins and players are assigned exogenous continuation values.<sup>4</sup> The essence of our approach is to contrast how the accumulation of weapons affects whether peace is

---

<sup>1</sup>See for instance Abreu (1988) on penal codes.

<sup>2</sup>See, for instance, Jervis (1989).

<sup>3</sup>Among others, see the Oxfam Report (2007), and Flint and de Waal (2006).

<sup>4</sup>Because the players' payoffs upon conflict are exogenously specified, this game is not a repeated game. However, trigger strategies of a repeated game are naturally mapped into an exit game in which continuation values upon conflict are those that players obtain from repeatedly playing (*Attack, Attack*). Therefore, this exit framework is sufficiently flexible to capture the insights we typically obtain from a repeated prisoners' dilemma.

sustainable or not under complete information and under strategic risk.

Our model of strategic risk follows the global games literature.<sup>5</sup> More precisely, we consider a situation in which payoffs upon peace depend on an uncertain state of the world about which players obtain very informative but noisy signals. Because players do not have the same assessment of the state of the world this creates strategic uncertainty in equilibrium, so that one player may choose peace while the other one is attacking. As a consequence, the sustainability of peace will depend both on how tempting attacking a peaceful opponent is – this is the *predatory motive* for conflict – and on how costly it is to be attacked when peaceful – this is what we call the *preemptive motive* for conflict. In fact, while only predatory motives matter under complete information, we show that even as the players' information becomes arbitrarily precise, whether peace is sustainable under incomplete information will depend significantly on the magnitudes of *both* the predatory and the preemptive incentives. The paper then contrasts comparative statics obtained with and without strategic uncertainty and highlights how taking into account the preemptive motive for attack enriches and nuances our intuitions about the determinants of conflict.

Our first result considers groups with symmetric stocks of weapons. In this setting we show that increasing weapon stocks will always have a deterrent effect under complete information but that they may very well be destabilizing under strategic risk. This happens because upon conflict, the increased destruction caused by weapons decreases the payoffs of both the attacker and the victim of the surprise attack. Because they diminish payoffs to the attacker, weapons reduce the predatory motive for conflict. However, as the victim of a surprise attack also fares worse, weapons may increase the benefits of preemptive strikes, which can result in overall destabilization. We show that under general conditions, the effect of weapons accumulation on peace is non-monotonic and that very destructive weapons (i.e. nuclear bombs) will typically be deterrent whereas intermediate weapons (i.e. semi-automatic guns) may be destabilizing.

Our second result explores how inequality in military strength may affect the sustain-

---

<sup>5</sup>See for instance Carlsson and van Damme (1993) and Morris and Shin (1997).

ability of peace. We show that unequal military strength always pushes towards conflict under complete information but that the picture becomes more nuanced once strategic risk is introduced. Unequal military strength is destabilizing under complete information because it increases the predatory temptation of the stronger player. Inequality, however, may reduce the preemptive motive for conflict. First, the stronger group knows it has little to fear from the weaker group. Second, the weaker group knows that it can only gain very little by launching a preemptive attack. As a consequence, under strategic risk, peace might be possible between unequal contenders in circumstances under which equally armed opponents would fight. This result, however should not be interpreted as making a case for complete monopoly of violence. Indeed, while inequality can help, peace is only sustainable if the weaker group keeps enough weapons to limit the stronger group's predatory incentives.

Finally, we examine the impact of peace-enforcing interventions on peace and conflict.<sup>6</sup> We first highlight that under complete information, unless intervention is immediate so that war is prevented altogether, intervention will always have a destabilizing impact. Indeed, it is precisely the perspective of a long and painful conflict that deters groups from attacking. This conclusion, however, is not robust to strategic risk. By alleviating the costs of being the victim of a surprise attack, intervention reduces the need for preemptive strikes. In that setting we show that the promise of intervention may promote peace even if it can only happen with delay.

Because we examine deterrence in a model where agents are fully rational, this paper is related to the “realist” strand of the International Relations literature.<sup>7</sup> Our model can actually be seen as formalizing and systematically exploring the impact of “reciprocal fears of surprise attack” as discussed by Schelling (1960). In that sense, the paper is also related to the spiral theories of war of Jervis (1976, 1978) and Kydd (1997).<sup>8</sup> Our model is also closely

---

<sup>6</sup>See Collier et al (2003) for a study of the causes and consequences of civil war. Doyle and Sambanis (2006) present an analysis of peace-keeping operations.

<sup>7</sup>This literature includes many non-formal theories of war. For an early formal model in this tradition see Bueno de Mesquita (1981).

<sup>8</sup>While these models were originally developed to understand interstate conflict, Posen(1993) and Snyder and Jervis(1999) have shown they can also help understand civil wars and ethnic conflict. We believe our

related to Baliga and Sjöström (2004) who analyze the role of cheap talk in a model where incomplete information about the players' types triggers conflict via a contagion process.

Our goal in this paper is to highlight the importance of strategic risk when analyzing the impact of weapons on peace. As a result, we choose to abstract from a number of other realistic dimensions of conflict already emphasized in the literature, such as bargaining failures (see Fearon (1995)), leader bias (see Jackson and Morelli (2007)), commitment problems (see Powell (2004)), or renegotiation issues (in the context of nuclear deterrence see Schelling (1966), Jervis (1979, 1989) or Powell (1990)).<sup>9</sup> Also, unlike Garfinkel (1990), Grossman (1991) or Skaperdas (1992) we do not consider the question of endogenous investment in weapons. Rather, our purpose here is to revisit a more primitive question: how does the accumulation of weapons affect the sustainability of peace?

The paper is organized as follows. Section 2 describes the framework and provides necessary and sufficient conditions for the sustainability of peace under complete and incomplete information. Section 3 contrasts the mechanics of deterrence with and without strategic risk. Section 4 studies how inequality in military strength affects conflict. Section 5 explores the impact of intervention on peace. Section 6 concludes. Proofs are contained in the appendix.

## 2 The model

### 2.1 A Simple Model of Peace and Conflict

We consider two groups  $i \in \{1, 2\}$  that play an infinite horizon trust game, with discrete time  $t \in \mathbb{N}$ , and share a common discount factor  $\delta$ . Each period  $t$ , the players simultaneously decide whether to be peaceful (P) or to attack (A). If both players are peaceful at time  $t$ , they obtain a flow payoff  $\pi$  and the game moves on to period  $t + 1$ . When any of the players attacks, conflict begins and the players receive exogenously given conflict payoffs. The stage

---

results are relevant to the analysis of both types of conflict.

<sup>9</sup>Fearon (1995) shows that in the presence of bargaining and transfers, a rational unitary model that yields war on the equilibrium path needs either private information, bargaining indivisibilities or a commitment problem. Here we have both private information and commitment problems.

payoffs are denoted as follows

	$P$	$A$
$P$	$\pi$	$S(k_i, k_{-i})$
$A$	$F(k_i, k_{-i})$	$W(k_i, k_{-i})$

where payoffs are given for row player  $i$ , and  $k_i \in \mathbb{R}^+$  is the stock of weapons held by player  $i$ . Payoff  $\pi$  represents the flow benefits of peace<sup>10</sup>, while payoffs  $F(k_i, k_{-i})$ ,  $S(k_i, k_{-i})$  and  $W(k_i, k_{-i})$  correspond to the reduced form payoffs group  $i$  obtains upon conflict.<sup>11</sup> These payoffs depend on the timing of attacks. More specifically  $F(k_i, k_{-i})$  denotes the payoff obtained by group  $i$  if it can launch a surprise attack.  $S(k_i, k_{-i})$  is the payoff if group  $i$  suffers a surprise attack and is therefore a second mover. Finally,  $W(k_i, k_{-i})$  denotes  $i$ 's payoff when groups launch simultaneous attacks. We choose to keep these payoffs in a reduced form since it allows us to remain agnostic about the specific pattern of conflict. Note that our formalism is consistent with  $F$ ,  $W$  and  $S$  being players' payoffs upon defection when they play trigger strategies of a repeated prisoners' dilemma.

For simplicity, we will denote  $F_i = F(k_i, k_{-i})$ ,  $S_i = S(k_i, k_{-i})$  and  $W_i = W(k_i, k_{-i})$ . Whenever arm stocks are symmetric ( $k_i = k_{-i} = k$ ), we will also denote  $F(k) = F(k, k)$ ,  $S(k) = S(k, k)$  and  $W(k) = W(k, k)$ . The payoff difference  $F_i - \frac{1}{1-\delta}\pi$  corresponds to the *predatory incentives* of player  $i$ , that is, how much player  $i$  would gain from attacking a consistently peaceful opponent. The payoff difference  $W_i - S_i$  corresponds to the *preemptive incentives* of player  $i$ , that is, how much player  $i$  would gain from attacking a consistently aggressive opponent. We make the following assumption.

---

<sup>10</sup>We look at a situation where the benefits of peace  $\pi$  are symmetric for the purpose of simplicity. Extending the model to a setting with asymmetric benefits presents no conceptual difficulty.

<sup>11</sup>These reduced form payoffs summarize the history of fighting that starts after (A) is chosen by a player. It might help intuition to think of these conflict payoffs as discounted sums of flow payoffs that depend on who initiated conflict. In this case, we would write

$$F(k_i, k_{-i}) = \sum_{t=0}^{+\infty} \delta^t f_t(k_i, k_{-i}); \quad S(k_i, k_{-i}) = \sum_{t=0}^{+\infty} \delta^t s_t(k_i, k_{-i}); \quad W(k_i, k_{-i}) = \sum_{t=0}^{+\infty} \delta^t w_t(k_i, k_{-i}).$$

**Assumption 1 (First Strike Advantage)** *For all weapon stocks  $k_i$  and  $k_{-i}$ , payoffs upon conflict are such that  $F(k_i, k_{-i}) > W(k_i, k_{-i}) > S(k_i, k_{-i})$ .*

Assumption 1 states that there is an advantage in attacking first and that there is a preemptive motive for war as the payoffs from a simultaneous attack dominate those from being a second mover.

Throughout the paper, we contrast a situation in which the flow benefits of peace  $\pi$  are common knowledge and a situation in which players make private but very precise assessments of the value of  $\pi$ . In the first case, common knowledge of payoffs allows players to coordinate their actions effectively. Under incomplete information however, coordination becomes difficult as players attempt to second guess one another's value for peace before making decisions.

## 2.2 Peace and Conflict under Complete Information

By Assumption 1, since  $S_i < W_i$ , attacking simultaneously is always an equilibrium of the game whether or not there is incomplete information about  $\pi$ . The question of interest, therefore, is whether or not peace is sustainable. In this section we focus on the case of complete information: payoff  $\pi$  is fixed and common knowledge between players. Let us denote this game by  $\Gamma_{CI}$ . The following result holds.

**Proposition 1 (peace under complete information)** *Peace is an equilibrium outcome of  $\Gamma_{CI}$  if and only if*

$$\forall i \in \{1, 2\}, \quad F_i - \frac{1}{1-\delta}\pi \leq 0. \quad (1)$$

*Furthermore, whenever inequality (1) holds, then permanent peace is sustainable in equilibrium.*

Recall that the difference  $F_i - \frac{1}{1-\delta}\pi$  corresponds to the predatory temptation of player  $i$ . It represents how much player  $i$  can gain by attacking a peaceful opponent. All that matters for peace to be sustainable under complete information is that the predatory incentives of

both players be low enough. Payoffs  $S_i$  and  $W_i$  do not matter in determining whether or not peace is sustainable.

Let us denote by  $\pi_{CI}$  the smallest value of  $\pi$  such that peace is sustainable under complete information. It follows from Lemma 1 that  $\pi_{CI} = (1 - \delta) \max_i F_i$ . Thus, if  $F_i$  is high, cooperation can only be sustained if the returns to peace are high as well.

## 2.3 Peace and Conflict under Strategic Risk

### 2.3.1 Framework

Under complete information, players never fear the possibility of suffering a surprise attack. Indeed, in equilibrium the likelihood of being attacked while peaceful is zero. Our model of strategic risk follows the literature on global games and introduces the possibility of miscoordination in equilibrium by considering a situation in which the returns to peace are not common knowledge.<sup>12</sup> More precisely we follow the framework of Chassang (2007) and consider a slightly perturbed exit game with flow payoffs

	$P$	$A$
$P$	$\tilde{\pi}_t$	$S_i$
$A$	$F_i$	$W_i$

where  $\tilde{\pi}_t$  is an i.i.d. random variable with finite variance, distribution  $f$  and support  $(-\infty, +\infty)$ . This peace payoff  $\tilde{\pi}_t$  is not directly observable by the players when they make their decision at time  $t$ . Instead, players observe signals of the form  $x_{i,t} = \tilde{\pi}_t + \sigma \epsilon_{i,t}$  where  $\{\epsilon_{i,t}\}_{i \in \{1,2\}, t \in \mathbb{N}}$  is an i.i.d. sequence of centered errors with support  $[-1, 1]$ . For simplicity we assume that  $\tilde{\pi}_t$  is observable in period  $t + 1$  via the flow payoffs. Let us denote this game by  $\Gamma_{\sigma, f}$ .

---

<sup>12</sup>See for instance Carlsson and van Damme (1993) or Morris and Shin (2003) for an extensive literature review. Note that for our results it does not matter much whether uncertainty concerns payoff from peace  $\pi$  or the temptation of attack  $F_i$ .

### 2.3.2 Structural properties

In this game, a history  $h_{i,t}$  for player  $i$  is a sequence of past signals and past realizations of  $\tilde{\pi}$  taking the form  $h_{i,t} = \{x_{i,1}, \tilde{\pi}_1, \dots, x_{i,t-1}, \tilde{\pi}_{t-1}, x_{i,t}\}$ . Denote by  $\mathcal{H}$  the set of all such histories. A strategy for player  $i$  is simply a mapping  $s_i : \mathcal{H} \rightarrow \{P, A\}$ . We give a few structural results before discussing the framework.

**Definition 1 (Order on strategies)** *We define a partial order  $\preceq$  on strategies as follows:*

$$s \preceq s' \iff \{a.s. \forall h \in \mathcal{H}, s(h) = P \Rightarrow s'(h) = P\}.$$

In words, one strategy is greater than another if and only if it is always more peaceful. Our first result establishes the existence of lowest and highest equilibria taking a simple form.

**Lemma 1 (extreme equilibria)** *There exists  $\bar{\sigma} > 0$  such that*

- (i) *For all  $\sigma > 0$ , attacking always is the lowest equilibrium strategy. It is associated with the lowest pair of equilibrium values  $(W_i, W_{-i})$ .*
- (ii) *For all  $\sigma \in (0, \bar{\sigma})$ , the set of perfect Bayesian equilibria admits a highest equilibrium with respect to  $\preceq$ , denoted by  $\mathbf{s}_\sigma^{\mathbf{H}} = (s_{i,\sigma}^H, s_{-i,\sigma}^H)$ . This highest equilibrium is associated with the highest pair of equilibrium values  $\mathbf{V}_\sigma^{\mathbf{H}} = (V_{i,\sigma}^H, V_{-i,\sigma}^H)$ .*
- (iii) *For all  $\sigma \in (0, \bar{\sigma})$ ,  $\mathbf{s}_\sigma^{\mathbf{H}}$  is characterized by fixed thresholds  $(x_{i,\sigma}^H, x_{-i,\sigma}^H) \in \mathbb{R}^2$  such that player  $i$  plays peace if and only if  $x_{i,t} \geq x_{i,\sigma}^H$ .*

This setup captures the idea of strategic risk in equilibrium by allowing players to have different perceptions of their environment. Although strategies are common knowledge in equilibrium, the fact that perceptions are private implies that there is no common knowledge of which actions will be taken. As a consequence, there is some risk of miscoordination (outcome  $(P, A)$ ) in equilibrium. This leads players to second guess each other's assessment of the situation, and given that a surprise attack is always possible, preemptive incentives will

play an important role in determining players' behavior. Moreover, as the next section shows, preemptive incentives continue to matter significantly even when the players have arbitrarily good information about  $\tilde{\pi}_t$  and the likelihood of actual miscoordination is vanishing. In that sense preemptive incentives have a selection effect on what equilibrium players choose to play, even when they do not enter payoffs on the equilibrium path directly.

### 2.3.3 Sustaining peace under strategic risk

We are now interested in the properties of game  $\Gamma_{\sigma, f}$  when its payoffs and information structure become arbitrarily close to those of the complete information game  $\Gamma_{CI}$  with constant benefit  $\pi$ . For this purpose, we consider a sequence of distributions  $\{f_n\}_{n \in \mathbb{N}}$  such that for all  $n \in \mathbb{N}$ ,  $f_n$  has support  $(-\infty, +\infty)$  and  $\{f_n\}_{n \in \mathbb{N}}$  converges in mean to  $d_\pi$ , the degenerate distribution that puts a unit mass at  $\pi$ . Then, we have

**Proposition 2 (sustainability of peace under strategic risk)** *For each distribution  $f_n$  and noise level  $\sigma > 0$ , denote  $\mathbf{V}_{\sigma, f_n}^H$  the highest sustainable pair of values of  $\Gamma_{\sigma, f_n}$ . We have that*

$$\prod_{i \in \{1, 2\}} \left( \frac{1}{1 - \delta} \pi - F_i \right)^+ > \prod_{i \in \{1, 2\}} (W_i - S_i) \implies \lim_{n \rightarrow \infty} \lim_{\sigma \rightarrow 0} \mathbf{V}_{\sigma, f_n}^H = \left( \frac{1}{1 - \delta} \pi, \frac{1}{1 - \delta} \pi \right)$$

$$\prod_{i \in \{1, 2\}} \left( \frac{1}{1 - \delta} \pi - F_i \right)^+ < \prod_{i \in \{1, 2\}} (W_i - S_i) \implies \lim_{n \rightarrow \infty} \lim_{\sigma \rightarrow 0} \mathbf{V}_{\sigma, f_n}^H = (W_i, W_{-i})$$

In words, peace is robust to the introduction of small amounts of incomplete information on  $\pi$  if and only if (peace, peace) is the risk-dominant equilibrium of the  $2 \times 2$  augmented one-shot game in which the continuation value upon peace is that of permanent peace. If this isn't the case and (attack, attack) is the risk-dominant equilibrium of the augmented one-shot game, then immediate conflict is the only equilibrium that's robust to the introduction of small amounts of noise. Hence we say that cooperation is *sustainable under strategic uncertainty*

if and only if

$$\prod_{i \in \{1,2\}} \left( \frac{1}{1-\delta} \pi - F_i \right)^+ > \prod_{i \in \{1,2\}} (W_i - S_i). \quad (2)$$

Let us denote by  $\pi_{SU}$  the smallest value of  $\pi$  such that cooperation is sustainable under strategic uncertainty. It is clear from (1) and (2) that  $\pi_{SU}$  is always greater than  $\pi_{CI}$ . In other words, peace is always more difficult to sustain under strategic risk. When payoffs are symmetric ( $k_i = k_{-i}$ ), condition (2) takes a particularly simple form:

$$F - \frac{1}{1-\delta} \pi + W - S < 0. \quad (3)$$

Condition (3) makes very clear how introducing strategic risk will affect comparative statics on the sustainability of peace. While only predatory incentives ( $F - \frac{1}{1-\delta} \pi$ ) matter under complete information, both predatory and preemptive incentives ( $W - S$ ) matter under strategic risk. Whenever the two incentives move in different directions, strategic risk will significantly change our intuitions about the determinants of peace and conflict.

## 2.4 A benchmark model

Most of the results given in the paper can and will be stated in terms of reduced form payoffs  $F$ ,  $W$  and  $S$ , however we find it useful for intuition to have a benchmark model of payoffs upon conflict.

**Definition 2 (benchmark model)** *Payoffs upon conflict  $F$ ,  $S$  and  $W$  are such that*

$$(i) \quad W(k_i, k_{-i}) = \frac{k_i}{k_i + k_{-i}} m - D(k_{-i}).$$

$$(ii) \quad F(k_i, k_{-i}) = W(\rho_F k_i, \rho_S k_{-i}) \text{ and } S(k_i, k_{-i}) = W(\rho_S k_i, \rho_F k_{-i})$$

where  $\rho_F > 1 > \rho_S \geq 0$ .

The first term of  $W(k_i, k_{-i})$  corresponds to a classic contest function.<sup>13</sup> It corresponds to the idea that players are competing for a prize  $m$  and that the likelihood of obtaining  $m$

---

<sup>13</sup>See for instance Hirshleifer (1995).

depends on the relative stocks of arms. The second term  $D : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is a continuously differentiable increasing function that represents the amount of destruction incurred by player  $i$  upon conflict, independently on whether she wins prize  $m$  or not. We capture the strategic advantage or disadvantage of being a first or a second mover by allowing weapon stocks to be inflated or deflated by factors  $\rho_F$  and  $\rho_S$  depending on the timing of attacks. The difference  $\rho_F - 1$  is positive and corresponds to the first mover advantage; the difference  $1 - \rho_S$  is also positive and corresponds to the second mover disadvantage.

As of now we don't specify  $D$  any further, but we think of it as bounded (i.e. in the event of a complete nuclear holocaust, the number of atomic bombs used in the process does not seem relevant for payoffs).  $D$  may also display convex parts. This is natural if as  $k$  increases, both the quantity and the nature of weapons are changing. For instance, imagine that a low capital stock  $k_0$  corresponds to the traditional weapons of a tribal society while a higher capital stock  $k_1$  corresponds to the introduction of machine guns. In this case, a marginal increase in capital stocks will have a much larger impact on damages  $D$  at capital  $k_1$  than at capital  $k_0$ . Altogether, the typical damage function we envision is bounded with  $S$ -shaped portions.

## 3 Deterrence

### 3.1 Deterrence when weapon stocks are symmetric

#### 3.1.1 General results

This section investigates how a symmetric increase in weapon stocks affects the sustainability of peace. More precisely we will be studying the comparative statics of thresholds  $\pi_{CI}$  and  $\pi_{SU}$ . These thresholds correspond to the minimum flow return to peace  $\pi$  necessary for peace to be sustainable. The lower these thresholds, the easier it is to sustain peace. We say that weapons are deterrent if and only if the symmetric accumulation of weapons reduces the minimum value of  $\pi$  required to sustain peace. The following assumption is maintained

throughout the paper.

**Assumption 2 (weapons are destructive)** *Payoffs  $F_i$ ,  $S_i$  and  $W_i$  are increasing in  $k_i$  and decreasing in  $k_{-i}$ . Furthermore,  $F(k)$ ,  $S(k)$  and  $W(k)$  are all decreasing in  $k$ .*

This is a natural assumption: conditional on conflict, player  $i$ 's payoff is increasing in her own stock of weapons and decreasing in her opponent's stock of weapons. Moreover, a symmetric increase in the amount of weapons makes conflict more painful on all sides. Note that payoffs  $F$ ,  $S$  and  $W$  corresponding to the benchmark model of Definition 2 satisfy Assumption 2.

When  $k_i = k_{-i} = k$  cooperation thresholds are  $\pi_{CI} = (1 - \delta)F$  and  $\pi_{SU} = (1 - \delta)[F + W - S]$ . As discussed above, sustaining peace is always more difficult under strategic uncertainty. More interestingly, the deterrent effect of weapons may differ across strategic settings.

**Proposition 3 (deterrence under complete and incomplete information)** *Consider a situation in which  $k_i = k_{-i} = k$ . Under Assumption 2, we have that*

(i)  $\pi_{CI}$  is always strictly decreasing in  $k$ .

(ii)  $\pi_{SU}$  is decreasing in  $k$  if and only if

$$\frac{dF}{dk} + \frac{dW}{dk} - \frac{dS}{dk} < 0. \quad (4)$$

Point (i) of Proposition 3 highlights that in a complete information setting, increasing weapon stocks unambiguously improves the sustainability of peace. This happens because under complete information, peace is sustainable if and only if the payoff  $F$  of a first mover attack is lower than the value of permanent peace  $\frac{1}{1-\delta}\pi$ . Because accumulating weapons decreases  $F$ , it also facilitates the sustainability of peace.

This prediction does not necessarily hold anymore once strategic risk is taken into account. Indeed under strategic risk the sustainability of peace depends both on the predatory and the preemptive motive for conflict. Hence, while an increase in weapons may reduce  $F$ , it may also greatly worsen the payoffs of suffering a surprise attack (starting conflict as a

second mover). This may increase the value of  $W - S$ , thereby increasing the temptation to launch preemptive strikes. Whenever the cost  $S$  of being a second mover rises more steeply than the cost  $W$  of simultaneous war and the cost  $F$  of initiating conflict, weapons will be destabilizing instead of deterring.

### 3.1.2 Deterrence in the benchmark model

To better understand the circumstances in which weapons will be destabilizing, we now examine the meaning of condition (4) when conflict payoffs are those of our benchmark model. The threshold  $\pi_{SU}$  takes the form

$$\begin{aligned}\pi_{SU} &= (1 - \delta)[W(k, k) + W(\rho_F k, \rho_S k) - W(\rho_S k, \rho_F k)] \\ &= (1 - \delta) \left[ \frac{1}{2}m + \frac{\rho_F - \rho_S}{\rho_F + \rho_S}m - D(k) - D(\rho_S k) + D(\rho_F k) \right].\end{aligned}$$

Therefore, in this case weapons are deterrent under strategic uncertainty if and only if

$$D'(k) + \rho_S D'(\rho_S k) - \rho_F D'(\rho_F k) \equiv \phi > 0.$$

Accumulating weapons is counter-productive otherwise. The question is now to understand what may affect the sign of  $\phi$ . In particular we are interested in how the first strike advantage  $\rho_F - 1$  and the second strike disadvantage  $1 - \rho_S$  may affect the deterrent impact of weapons. The effect of parameters  $\rho_F$  and  $\rho_S$  however is subtle and may depend on the shape of function  $D$ .

**Lemma 2** *If  $D$  is convex over the range  $[\rho_S k, \rho_F k]$ , then  $\phi$  is decreasing in  $\rho_F$  and increasing in  $\rho_S$ .*

Lemma 2 suggest that a large first strike advantage and a large second strike disadvantage will tend to make weapons destabilizing. A partial intuition goes as follows: if the first mover advantage and the second mover disadvantage are large, it is likely that when weapon stocks

increase the amount of destruction suffered by second movers will rise by more than the amount of destruction suffered by a first mover. Hence when the first mover advantage and second mover disadvantage are large, one should expect that weapons will be destabilizing. Lemma 2 however is more subtle than this simple reasoning suggests, and the shape of  $D$  also plays a role in determining whether weapons are destabilizing or not. Indeed, assume for instance that  $D$  is concave around  $\rho_F k$ . This means that the destruction suffered by a second mover increases at a diminishing rate when  $\rho_F k$  increases. It follows that a further increase in  $\rho_F$  will decrease the sensitivity of  $S$  to weapon stocks. Hence, if  $D$  is concave around  $\rho_F k$ , increasing the first mover advantage might improve the deterrent effect of weapons.

Because the deterrent effect of weapons depends on the local shape of the destruction function  $D$ , the predictions of our model can potentially accommodate varied patterns of deterrence. In the following section, we highlight that under reasonable assumptions our model predicts that very destructive weapons (i.e. nukes) are deterrent while intermediate weapons (i.e. guns) may be destabilizing.

### 3.2 Guns vs. Nukes

This section explores the possibility that different levels of weapons may have different deterrent effects. We start by imposing a natural assumption on conflict payoffs.

**Assumption 3 (destruction)** *As weapon stocks become large, the payoff difference between being a second mover and simultaneous conflict is minimized:*

$$\lim_{k \rightarrow +\infty} W(k) - S(k) = \inf_{k \geq 0} W(k) - S(k)$$

This assumption corresponds to the idea that when weapon stocks are very large, the gains from launching preemptive attacks are small. This is consistent with the idea of *mutually assured destruction*. Imagine for instance that  $\lim_{k \rightarrow +\infty} F(k) = \lim_{k \rightarrow +\infty} W(k) = \lim_{k \rightarrow +\infty} S(k) = \inf_{k \geq 0} S(k)$ . In that case, when weapon stocks are large, destruction is

complete and payoffs upon conflict are independent on who initiated the first attack. As a result, neither predatory nor preemptive attacks make sense and incentives for conflict are minimized when destructive capacity is high. This yields the following result.

**Proposition 4 (nukes are deterrent)** *If Assumptions 2 and 3 hold, peace is most sustainable under strategic risk when the stock of weapons becomes arbitrarily large. More formally*

$$\lim_{k \rightarrow +\infty} \pi_{SU}(k) = \inf_{k \geq 0} \pi_{SU}(k).$$

Note that our benchmark model satisfies Assumption 3 whenever the destruction function  $D$  is bounded above. Hence, when weapon stocks are symmetric, sufficiently destructive power will guarantee the highest possibly sustainable level of peace.

This result however does not imply that weapons monotonically increase stability in a world with strategic risk. In fact we now consider a very stark example highlighting how convexities in the destruction function  $D$  may cause intermediate stocks of weapons to be destabilizing.

**Assumption 4 (disruptive technology)** *There exists a weapon level  $k^*$  such that  $D'(k^*) = +\infty$  while  $D' < +\infty$  everywhere else.*

This assumption is consistent with  $D$  being  $S$ -shaped.

**Proposition 5 (disruptive weapons precipitate war)** *Whenever Assumption 4 holds, there exists an open interval  $I \subset \mathbb{R}$  containing  $k^*/\rho_F$  such that  $\pi_{SU}$  is strictly increasing in  $k$  over  $I$ .*

When the joint stock of weapons is exactly  $k^*/\rho_F$ , the destruction experienced by a second mover is equal to  $D(k^*)$ , hence a marginal increase in weapon stocks hurts a second mover much more than a first mover. As a consequence, any marginal increase in destructive capacity is destabilizing. While Assumption 4 facilitates the statement of Proposition 5, the assumption that  $D'$  be infinite for some stock of weapons  $k^*$  is by no means necessary. For

instance if function  $D$  was  $S$ -shaped with a sufficiently steep inflexion point a similar result would hold.

It follows from this that the effect of weapons on conflict may well be non-monotonic: while very destructive weapons always have a deterrent effect, intermediate levels of weapons can be destabilizing. This helps reconcile the seemingly contradictory evidence on the effect of weapons on the sustainability of peace.

## 4 Stabilizing Inequality

In the previous section we have analyzed the case of two contenders with equal weapons stocks. We now turn to the question of how inequality in military strength affects the sustainability of peace. Inequality is parameterized by a constant  $\lambda \in [1, +\infty)$  so that  $k_i = \lambda k$  and  $k_{-i} = k$ . The following result is immediate.

**Proposition 6 (inequality is bad under complete information)** *Keeping  $k$  constant, greater inequality makes peace harder to sustain. More formally,  $\pi_{CI}$  is increasing in  $\lambda$ .*

This follows simply from the fact that  $\pi_{CI} = (1 - \delta) \max_{i \in \{1,2\}} F_i$ . As player  $i$  becomes stronger, his payoff  $F_i$  from initiating conflict increases and since only predatory incentives matter under complete information, peace becomes harder to sustain.

In contrast, Proposition 7 below shows that in a setting with strategic uncertainty, inequality in military strength can facilitate peace rather than generate war. This comes from the fact that while military inequality increases the stronger player's incentives to launch predatory attacks, it also reduces both players' incentives to launch preemptive attacks. To see this formally, we must compute threshold  $\pi_{SU}$ . When weapon endowments are unequal,  $\pi_{SU}$  is the only root of the second degree equation

$$\left( \frac{1}{1-\delta} \pi - F_i \right) \left( \frac{1}{1-\delta} \pi - F_{-i} \right) = (W_i - S_i)(W_{-i} - S_{-i})$$

that is also greater than  $(1 - \delta) \max_{i \in \{1,2\}} F_i$ .

We present our results in two steps. Lemmas 3 and 4 first provide conditions under which the destabilizing impact of inequality is mitigated under strategic risk compared to the complete information case. Proposition 7 then shows that under strategic uncertainty, peace may be possible between unequal contenders while equally strong groups would end up fighting.

**Lemma 3 (mitigated impact of inequality)** *Whenever  $(W_i - S_i)(W_{-i} - S_{-i})$  is decreasing in  $\lambda$ , we have that*

$$\frac{\partial \pi_{SU}}{\partial \lambda} < \frac{\partial \pi_{CI}}{\partial \lambda}.$$

In words, whenever the product  $(W_i - S_i)(W_{-i} - S_{-i})$  is decreasing in  $\lambda$ , strategic risk dampens the adverse impact of inequality on the sustainability of peace.

The term  $W_i - S_i$  corresponds to player  $i$ 's incentives to launch preemptive attacks. The fact that the *product* of these preemptive incentives affect equilibrium selection can be roughly assigned to the idea that the players' fear of suffering a surprise attack compound. Indeed, when the difference  $W_{-i} - S_{-i}$  is large, player  $i$  may worry that player  $-i$  is likely to launch a preemptive attack. This makes player  $i$ 's own incentives to launch preemptive strikes become more salient. As a result the effect of each player's preemptive incentives are complementary.<sup>14</sup> When the product of the preemptive incentives is decreasing in  $\lambda$ , inequality reduces the overall destabilizing effect of preemptive incentives. Next we show that in our benchmark model, large levels of inequality will in fact minimize the preemptive temptation.

**Lemma 4 (appeasing inequality)** *Assume that conflict payoffs  $F$ ,  $S$  and  $W$  are generated by the benchmark model of Definition 2 and that  $D$  is bounded above. Then, the preemptive incentives of both players  $i$  and  $-i$  are minimized when inequality parameter  $\lambda$*

---

<sup>14</sup>For a more detailed discussion of why it is specifically the product of preemptive incentives that matters, the interested reader is referred to Harsanyi and Selten (1988).

grows arbitrarily large:

$$\lim_{\lambda \rightarrow +\infty} W_i - S_i = \inf_{\lambda \geq 1} W_i - S_i \quad \text{and} \quad \lim_{\lambda \rightarrow +\infty} W_{-i} - S_{-i} = \inf_{\lambda \geq 1} W_{-i} - S_{-i}.$$

It is interesting to note that *both* players' incentives to launch preemptive attacks can diminish with  $\lambda$ . The stronger player's incentives diminish because she gets a share of the spoils close to 1 whether she acts second or simultaneously. The weaker player's incentives to launch preemptive attacks also diminish because when facing an overwhelmingly stronger opponent, she obtains the same payoffs whether she is a second mover or attacks simultaneously. Proposition 7 now shows that this positive effect of inequality can be strong enough that in some circumstances, peace is sustainable only when weapon stocks are sufficiently unequal.

**Proposition 7 (stabilizing inequality)** *Assume that  $D$  is bounded above. Whenever*

$$\frac{1}{1-\delta}\pi < \left[ \frac{1}{2} + \frac{\rho_F - \rho_S}{\rho_F + \rho_S} \right] m - D(\rho_S k) - D(k) + D(\rho_F k) \quad (5)$$

$$\text{and } \frac{1}{1-\delta}\pi > m - D(\rho_S k) \quad (6)$$

*then under strategic risk, peace is unsustainable for  $\lambda = 1$  but sustainable for  $\lambda = +\infty$ .*

Proposition 7 provides conditions under which peace is not sustainable if both groups have the same stock of weapons  $k$  but becomes sustainable if one of the players becomes overwhelmingly strong.<sup>15</sup> Condition (5) ensures that peace is not sustainable under strategic risk when  $\lambda = 1$ . This simply corresponds to the negation of condition (3) for our benchmark model. Inversely, condition (6) implies that when a player becomes arbitrarily strong, predatory attacks remain unattractive. When these conditions hold together then peace is sustainable only if players are sufficiently unequal.

---

<sup>15</sup>For a given  $k$  and  $\rho_F$ , there is always a  $\rho_S$  small enough such that these two conditions hold simultaneously for some  $\pi$ .

Note that the term corresponding to the players' preemptive incentives has dropped out in inequality (6). The only term that matters now corresponds to the deviation temptation of the stronger player. This highlights two important points: first, asymmetry can be stabilizing because it rules out preemption as a motive for conflict, second, for asymmetry to be beneficial, it is still necessary for the weaker party to keep sufficient military capacity so that predatory attacks are unattractive for the stronger player. In that sense, Proposition 7 relates to, but nuances, the idea that a monopoly of violence facilitates peace.

## 5 Conflict and intervention

This section explores the impact of peace keeping interventions on the sustainability of peace.<sup>16</sup> First, note that if peace keeping interventions reestablished peace immediately, it is clear they would be beneficial. However, problems arise if peace keeping operations only reestablish peace with delay. Indeed, a complete information model would predict that delayed peace keeping operations are in fact destabilizing. We show that this need not be the case anymore under strategic risk.

To understand whether late intervention can be effective, we unbundle payoffs upon conflict as a discounted sum of flow payoffs, and ask how the timing of third-party peace-enforcing interventions affects peace and conflict. We consider the case of symmetric weapon stocks. Payoffs take the form

$$F = \sum_{t=0}^{+\infty} \delta^t f_t ; \quad S = \sum_{t=0}^{+\infty} \delta^t s_t ; \quad W = \sum_{t=0}^{+\infty} \delta^t w_t$$

where  $\{f_t\}_{t \in \mathbb{N}}$ ,  $\{s_t\}_{t \in \mathbb{N}}$  and  $\{w_t\}_{t \in \mathbb{N}}$  are exogenously given streams of payoffs upon conflict. Peace keeping interventions are characterized by a date  $T$ , at which players anticipate that

---

<sup>16</sup>Note that we never consider the opportunity cost or direct social benefit of such peace keeping operations, but rather focus on how they affect peace and conflict. However, although we do not endeavor to do a full fledged welfare assessment of interventionist policies, we think of our analysis as an important input for such an assessment.

civil war will be interrupted. Some settlement is then imposed and players obtain flow payoffs  $\pi' \leq \pi$  from then on. Hence, if intervention occurs at time  $T \geq 1$ , players' payoffs upon civil war are

$$F^T = \sum_{t=0}^{T-1} \delta^t f_t + \frac{\delta^T}{1-\delta} \pi'; \quad S^T = \sum_{t=0}^{T-1} \delta^t s_t + \frac{\delta^T}{1-\delta} \pi'; \quad W^T = \sum_{t=0}^{T-1} \delta^t w_t + \frac{\delta^T}{1-\delta} \pi'.$$

When intervention occurs at time  $T$ , the minimum value of  $\pi$  for peace to be sustainable under complete information is

$$\pi_{CI}^T = (1-\delta) \sum_{t=0}^{T-1} \delta^t f_t + \delta^T \pi'. \quad (7)$$

For simplicity we make the following assumption.

**Assumption 5 (conflict as punishment)** *We assume that  $f_0 > \pi$  and for all  $t \geq 1$ ,  $f_t < \pi'$ .*

This corresponds to the idea that there are short term benefits to attacking followed by painful conflict payoffs. The following result shows how an expected intervention affects the sustainability of peace under complete information.

**Proposition 8 (intervention under complete information)** *Consider the complete information game in which intervention occurs at time  $T$ . The following hold,*

- (i) *whenever  $T = 0$ , peace is sustainable for any value  $\pi \geq \pi'$ ;*
- (ii) *whenever  $T \geq 1$ , then the cooperation threshold  $\pi_{CI}^T$  is decreasing in  $T$ . Hence if  $T \geq 1$ ,  $\pi_{CI}^T$  is minimized for  $T = +\infty$ .*

Point (i) of Proposition 8 highlights that if intervention were immediate, then peace would be sustainable for any value of  $\pi$ . This happens because a first mover attacker never gets the one shot benefit  $f_0$  but only ever gets settlement payoffs  $\pi' \leq \pi$ . Point (ii) shows

however that anticipating a *delayed intervention* is always destabilizing under complete information. Moreover it shows that if it is only feasible to intervene with some delay, then artificially increasing this response delay improves the sustainability of peace, to the point that committing not to intervene sustains the highest level of peace.

We now examine the impact of intervention under strategic risk. The minimum value of  $\pi$  for which cooperation is sustainable is

$$\pi_{SU}^T = (1 - \delta) \sum_{t=0}^{T-1} \delta^t (f_t + w_t - s_t) + \delta^T \pi'.$$

**Proposition 9 (intervention under strategic risk)** *If intervention occurs at time  $T$  then under strategic risk, the following hold,*

- (i) *whenever  $T = 0$ , peace is sustainable for any value  $\pi > \pi'$ ,*
- (ii) *for any  $T \geq 1$ , the cooperation threshold under strategic risk  $\pi_{SU}^T$  is increasing in  $T$  if and only if  $f_T + w_T - s_T > \pi'$ .*

Point (ii) of Proposition 9 highlights that even when only delayed intervention is feasible, intervention can facilitate the sustainability of peace and that artificially increasing anticipated delays may foster conflict. This occurs because under strategic risk, intervention affects the sustainability of peace via two channels. On the one hand it replaces flow predatory payoffs  $f_t$  by  $\pi'$  which is destabilizing. On the other hand, intervention replaces flow preemptive incentives  $w_t - s_t$  by 0, which is stabilizing. Whenever  $f_t + w_t - s_t > \pi'$  the second effect dominates and the promise of intervention – even delayed – improves the sustainability of peace. The following corollary reinterprets these results in the specific case where flow payoffs  $w_t$  upon simultaneous conflict are constant.

**Corollary 1 (converging and diverging conflicts)** *Assume that for all  $t \geq 0$ ,  $w_t = w_0$ . We have that*

- (i) *if  $f_t - s_t$  is increasing in  $t$  for all  $t \geq 0$ , then  $\pi_{SU}^T$  is increasing in  $T$ ;*

(ii) if  $f_t - s_t$  is decreasing in  $t$  for all  $t \geq 0$  and there exists  $T^*$  such that  $f_{T^*} + w_{T^*} - s_{T^*} \leq \pi'$ , then for all  $T \geq T^*$ ,  $\pi_{SU}^T$  is decreasing in  $T$ .

Point (i) of Corollary 1 states that when flow payoffs between first and second movers diverge with time, the promise of intervention at *some* time  $T$  will always improve the stability of peace, and that even if it is delayed, intervention should occur as early as possible. This corresponds to a setting where the first mover advantage and second mover disadvantage are durable, so that war becomes worse and worse for the victim of the first attack. In contrast, point (ii) of Corollary 1 states that whenever flow payoffs between first and second movers converge – in other words, when the victims can effectively retaliate – then only the promise of sufficiently early intervention can foster peace. If intervention cannot occur before some delay  $T^*$ , intervention unambiguously reduces the stability of peace. In this second case the intuition obtained under complete information survives: intervention only improves the sustainability of peace if it is expected to happen sufficiently early. If intervention can only happen with delay greater than  $T^*$ , then artificial delay (or abstaining from intervening) will improve the chances of peace. This suggests that intervention is most suited when conflicts follow a diverging pattern.

## 6 Conclusion

The purpose of this paper is to contrast the mechanics of conflict with and without strategic risk. It shows that under complete information, the sustainability of peace depends only on the players' predatory incentives. Under strategic risk however, the sustainability of peace depends both on predatory and preemptive incentives. Taking strategic risk seriously highlights the role of fear – rather than just temptation – in the determination of peace and war. This changes intuitions about deterrence and intervention in a number of ways. We focused on three particular insights.

First, while weapons are deterrent under complete information this need not be the case under strategic risk. Indeed, while weapons diminish players' temptation to launch predatory

attacks, they may also increase the temptation to launch preemptive attacks. As a result we show that weapons need not always be deterrent. We show that under natural conditions, sufficiently destructive weapons (i.e. nuclear warheads) will be deterrent, while intermediary weapons (i.e. guns) may be destabilizing. In particular we highlight the danger of disruptive weapons which hurt second movers much more than first movers in times of conflict.

Our second set of results pertains to the impact of unequal military strength on conflict. We show that under strategic risk, inequality may very well facilitate the sustainability of peace. Indeed, while inequality always increases one of the players' predatory temptation, it may also decrease both players' preemptive incentives: if one of the contenders is overwhelmingly stronger than the other, the timing of conflict doesn't change payoffs by much. As a result peace may be sustainable if groups are unequal and unsustainable if groups are equal. The model however doesn't suggest that monopoly of violence sustains the highest level of peace. Indeed, it is necessary in our framework that the weaker party keep sufficient weapon stocks to dissuade the stronger party from unilateral attacks. This result suggests that policies that attempt to level the playing field between conflicting groups may in fact be misguided and that *restrained superiority* may foster the greatest level of peace.

Finally we consider the relationship between intervention and conflict. We show that under complete information, unless intervention occurs immediately, it will make peace harder to sustain. This isn't true anymore under strategic risk as intervention may reduce players' fear of being the victim of a surprise attack. More precisely, we show that when conflict is diverging, in the sense that second movers fare worse and worse compared to first movers, then intervention will always facilitate the sustainability of peace. This result suggests that interventionist policies may improve the sustainability of peace even though they appear to worsen the players' predatory incentives.

The model we use to make these points is particularly simple. On the one hand, we view this as a strength of the paper. It highlights the importance of strategic risk as a fundamental determinant of peace and conflict, that can potentially yield rich comparative statics. Intuitions from our model also apply to many different circumstances of conflict, whether it

occurs between countries, armed groups within a country, or even between individuals. On the other hand, because it is so simple, our model leaves open a number of questions which need to be addressed if we are to gain a comprehensive understanding of the determinants of war and peace. In particular, we think that endogenizing weapon stocks and linking the economic benefits of peace to investment and the likelihood of future conflict are obvious directions for future research.

## A Appendix: Proofs

### A.1 Proofs for Section 2

**Proof of Lemma 1:** Since for all  $i \in \{1, 2\}$ ,  $F_i > W_i > S_i$ , the highest continuation value player  $i$  can expect is  $\max\{F_i, \frac{1}{1-\delta}\pi\}$ . If peace is an equilibrium action for player  $i$ , this implies that  $\pi + \delta \max\{F_i, \frac{1}{1-\delta}\pi\} \geq F_i$ , which yields that necessarily  $\frac{1}{1-\delta}\pi \geq F_i$ . Finally, since  $S_i < W_i$ , peace is an equilibrium action only if both players choose peace. This shows that whenever peace is an equilibrium outcome, then for all  $i \in \{1, 2\}$  we have  $\frac{1}{1-\delta}\pi \geq F_i$ . The reverse implication is straightforward: whenever  $\frac{1}{1-\delta}\pi \geq F_i$ , then being always peaceful is an equilibrium. ■

The proof of Lemma 1 and Proposition 2 is inspired from Chassang (2007) and Chassang (2008). However, because we have only one dominance region, the proofs must be adapted in non-trivial ways. We first introduce some notation and prove intermediary results in Lemmas 5 and 6.

**Definition 3** For any pair of values  $(V_i, V_{-i}) \in \mathbb{R}$  we denote by  $x^{RD}(V_i, V_{-i})$  the risk-dominant threshold of the one shot  $2 \times 2$  game

	$P$	$A$
$P$	$x + \delta V_i$	$S_i$
$A$	$F_i$	$W_i$

which is defined as the greatest solution of the second degree equation:

$$\prod_{i \in \{1, 2\}} (x + \delta V_i - F_i) = \prod_{i \in \{1, 2\}} (W_i - S_i) \quad (8)$$

**Definition 4** (i) A strategy  $s_i$  is said to take a threshold-form if and only if there exists  $x_i \in \mathbb{R}$  such that for all  $h_{i,t}$ ,  $s_i(h_{i,t}) = P \iff x_{i,t} \geq x_i$ . A strategy of threshold  $x_{-i}$  will be denoted  $s_{x_{-i}}$ .

(ii) Given a strategy  $s_{-i}$ , a history  $h_{i,t}$  and continuation value functions  $(V_i, V_{-i})$ , we denote by

$$\begin{aligned} U_{i,\sigma}^P(V_i, h_{i,t}, s_{-i}) &= \mathbb{E} [(\tilde{\pi}_t + \delta V_i) \mathbf{1}_{s_{-i}=P} + S_i \mathbf{1}_{s_{-i}=A} \mid h_{i,t}, s_{-i,t}] \\ U_{i,\sigma}^A(h_{i,t}, s_{-i}) &= \mathbb{E} [F_i \mathbf{1}_{s_{-i}=P} + W_i \mathbf{1}_{s_{-i}=A} \mid h_{i,t}, s_{-i,t}] \end{aligned}$$

the payoffs<sup>17</sup> player  $i$  expects upon playing  $P$  and  $A$ .

(iii) Given a strategy  $s_{-i}$  we denote by  $V_{i,\sigma}(s_{-i})$  the value function that player  $i$  obtains from best-replying to strategy  $s_{-i}$ .

(iv) Given a strategy  $s_{-i}$ , a history  $h_{i,t}$  and a value function  $V_i$ , we define

$$\Delta_{i,\sigma}(h_{i,t}, s_{-i}, V_i) = U_{i,\sigma}^P(V_i, h_{i,t}, s_{-i}) - U_{i,\sigma}^A(h_{i,t}, s_{-i}).$$

(v) Given  $x_i \in \mathbb{R}$  and  $V_i \in \mathbb{R}$ , for all  $\alpha \in [-2, 2]$  we define  $\widehat{\Delta}_{i,\sigma}(x_i, \alpha, V_i) = \Delta_{i,\sigma}(x_i, s_{x_i - \alpha\sigma}, V_i)$ .

**Lemma 5 (intermediary results)** There exists  $\bar{\sigma} > 0$  and  $\kappa > 0$  such that for all  $\sigma \in (0, \bar{\sigma})$ , all the following hold,

(i) Whenever  $s_{-i}$  is threshold-form and  $s'_{-i} \preceq s_{-i}$ , then  $V_{i,\sigma}(s_{-i}) \geq V_{i,\sigma}(s'_{-i})$ .

(ii) Consider  $s_{-i}$  a threshold form strategy and  $s'_{-i}$  any strategy such that  $s'_{-i} \preceq s_{-i}$ .

Whenever  $\Delta_{i,\sigma}(h_{i,t}, s'_{-i}, V_{i,\sigma}(s'_{-i})) \geq 0$  then  $\Delta_{i,\sigma}(h_{i,t}, s_{-i}, V_{i,\sigma}(s_{-i})) \geq \Delta_{i,\sigma}(h_{i,t}, s'_{-i}, V_{i,\sigma}(s'_{-i}))$

(iii) For any  $V_i \in [W_i, \frac{1}{1-\delta}\pi]$ , whenever  $\widehat{\Delta}_{i,\sigma}(x_i, \alpha, V_i) \geq 0$  then  $\frac{\partial \widehat{\Delta}_{i,\sigma}}{\partial x_i} > \kappa$  and  $\frac{\partial \widehat{\Delta}_{i,\sigma}}{\partial \alpha} > 0$ . Furthermore, if in addition there exists  $V_{-i} \in [W_{-i}, \frac{1}{1-\delta}\pi]$  such that  $\Delta_{-i,\sigma}(x_i - \alpha\sigma, -\alpha, V_{-i}) \geq 0$ , then  $\frac{\partial \widehat{\Delta}_{i,\sigma}}{\partial \alpha} > \kappa$ .

**Proof:** We begin with point (i). Let us first show that whenever  $V$  is a constant and  $V'$  a value function such that for all  $h_{i,t}$ ,  $V'(h_{i,t}) \leq V$  then for  $\sigma$  small enough,

$$\max\{U_{i,\sigma}^P(V, h_{i,t}, s_{-i}), U_{i,\sigma}^A(h_{i,t}, s_{-i})\} \geq \max\{U_{i,\sigma}^P(V', h_{i,t}, s'_{-i}), U_{i,\sigma}^A(h_{i,t}, s'_{-i})\}.$$

<sup>17</sup>We drop the  $\sigma$  subscript and the dependency on  $h_{i,t}$  whenever doing so does not cause confusion.

Indeed, since  $F_i > W_i$  it follows that  $U_{i,\sigma}^A(s_{-i}) \geq U_{i,\sigma}^A(s'_{-i})$ . Also, whenever  $U_{i,\sigma}^P(V, s'_{-i}) \geq U_{i,\sigma}^A(s'_{-i})$  then it must be that for some value of  $\tilde{\pi}_t$  with positive likelihood conditionally on  $h_{i,t}$ ,  $\tilde{\pi}_t + \delta V \geq F_i$ . Since  $F_i > S_i$  and  $\tilde{\pi}_t$  has support  $[x_{i,t} - \sigma, x_{i,t} + \sigma]$  conditionally on  $h_{i,t}$ , this implies that there exists  $\bar{\sigma}_1 > 0$  such that for all  $\sigma \in (0, \bar{\sigma}_1)$ ,  $U_{i,\sigma}^P(V, s'_{-i}) \geq U_{i,\sigma}^A(s'_{-i})$  implies that with probability 1 conditionally on  $h_{i,t}$ ,  $\tilde{\pi}_t + \delta V > S_i$ . This yields that whenever  $U_{i,\sigma}^P(V, s'_{-i}) \geq U_{i,\sigma}^A(s'_{-i})$ , then  $U_{i,\sigma}^P(V, s_{-i}) \geq U_{i,\sigma}^P(V, s'_{-i})$ . Since  $U_{i,\sigma}^P(V, s_{-i}) > U_{i,\sigma}^P(V', s_{-i})$ , this yields that indeed for all  $\sigma \in (0, \bar{\sigma}_1)$ ,

$$\max\{U_{i,\sigma}^P(V, s_{-i}), U_{i,\sigma}^A(s_{-i})\} \geq \max\{U_{i,\sigma}^P(V', s'_{-i}), U_{i,\sigma}^A(s'_{-i})\}. \quad (9)$$

Since for any strategy  $s''_{-i}$ , the value  $V_i(s''_{-i})$  is the highest solution of the fixed point equation  $V_i(s''_{-i})(h_{i,t}) = \max\{U_i^P(V_i(s''_{-i}), s''_{-i}), U_i^A(s''_{-i})\}$ , inequality (9) implies that for all  $\sigma \in (0, \bar{\sigma}_1)$ ,  $V_{i,\sigma}(s_{-i}) \geq V_{i,\sigma}(s'_{-i})$ . This proves point (i).

We now turn to point (ii). From point (i), we know that  $V_{i,\sigma}(s_{-i}) \geq V_{i,\sigma}(s'_{-i})$ . Also, since  $S_i - W_i < 0$ , there exists,  $\bar{\sigma}_2 > 0$  such that for all  $\sigma \in (0, \bar{\sigma}_2)$ ,  $\Delta_{i,\sigma}(h_{i,t}, s'_{-i}, V) \geq 0$  implies that  $\tilde{\pi}_t + \delta V - F_i > S_i - W_i$ . This yields that

$$\begin{aligned} \Delta_{i,\sigma}(h_{i,t}, s_{-i}, V_{i,\sigma}(s_{-i})) &= \mathbb{E}[(\tilde{\pi}_t + \delta V_{i,\sigma}(s_{-i}) - F_i)\mathbf{1}_{s_{-i}=P} + (S_i - W_i)\mathbf{1}_{s_{-i}=A} \mid h_{i,t}, s_{-i}] \\ &\geq \mathbb{E}[(\tilde{\pi}_t + \delta V_{i,\sigma}(s_{-i}) - F_i)\mathbf{1}_{s'_{-i}=P} + (S_i - W_i)\mathbf{1}_{s'_{-i}=A} \mid h_{i,t}, s'_{-i}] \\ &\geq \mathbb{E}[(\tilde{\pi}_t + \delta V_{i,\sigma}(s'_{-i}) - F_i)\mathbf{1}_{s'_{-i}=P} + (S_i - W_i)\mathbf{1}_{s'_{-i}=A} \mid h_{i,t}, s'_{-i}] \\ &\geq \Delta_{i,\sigma}(h_{i,t}, s'_{-i}, V_{i,\sigma}(s'_{-i})) \end{aligned}$$

which yields point (ii).

We now turn to point (iii). Denote by  $f_\epsilon$  and  $F_\epsilon$  the distribution and c.d.f. of  $\epsilon_{i,t}$  and define  $G_\epsilon \equiv 1 - F_\epsilon$ . Recall that  $f$  denotes the distribution of  $\tilde{\pi}_t$ . We have that

$$\begin{aligned} \Delta_{i,\sigma}(x_i, \alpha, V_i) &= \mathbb{E}[(\tilde{\pi}_t + \delta V_i - F_i)\mathbf{1}_{x_{-i,t} \geq x_i - \alpha\sigma} + (S_i - W_i)\mathbf{1}_{x_{-i,t} \leq x_i - \alpha\sigma} \mid x_{i,t}] \\ &= \int_{-1}^1 [(x_i - \sigma u + \delta V_i)F_\epsilon(\alpha + u) + (S_i - W_i)G_\epsilon(\alpha + u)] \underbrace{\frac{f_\epsilon(u)f(x_i - \sigma u)}{\int_{-1}^1 f_\epsilon(u')f(x_i - \sigma u')du'}}_{\equiv \Psi_\sigma(x_i, u)} du. \end{aligned}$$

Since  $S_i - W_i < 0$ , there exists  $\bar{\sigma}_3 > 0$  and  $\tau > 0$  such that for all  $\sigma \in (0, \bar{\sigma}_3)$ , whenever  $\Delta_{i,\sigma}(x_i, \alpha, V_i) \geq 0$  then  $\alpha \geq -2 + \tau$ . Otherwise  $F_\epsilon(\alpha + u)$  would be arbitrarily small and we would have  $\Delta_{i,\sigma}(x_i, \alpha, V_i) < 0$ . Standard results on convolution products<sup>18</sup> show that

<sup>18</sup>See for instance Lemma 8 of Chassang (2008)

as  $\sigma$  goes to 0, the posterior  $\Psi_\sigma(x_i, u)$  converges uniformly to  $f_\varepsilon(u)$  and that  $\frac{\partial \Psi_\sigma}{\partial x_i}$  converges uniformly to 0. This yields that there exists  $\bar{\sigma}_4$  and  $\kappa_1 > 0$  such that whenever  $\sigma \in (0, \bar{\sigma}_4)$ , then  $\frac{\partial \Delta_{i,\sigma}}{\partial x_i} > \kappa_1 > 0$ .

Now assume that we also have  $\Delta_{-i,\sigma}(x_i - \alpha\sigma, -\alpha, V_{-i}) \geq 0$ . Since  $S_{-i} - W_{-i} < 0$  there exists  $\bar{\sigma}_5 > 0$  and  $\tau' > 0$  such that for all  $\sigma \in (0, \bar{\sigma}_5)$ ,  $\Delta_{-i,\sigma}(x_i - \alpha\sigma, -\alpha, V_{-i}) \geq 0$  implies that  $-\alpha \geq -2 + \tau'$ . Altogether this implies that  $\alpha \in [-2 + \tau, 2 - \tau']$ . From there, simple algebra yields that there exists  $\bar{\sigma}_6 > 0$  and  $\kappa_2 > 0$  such that for all  $\sigma \in (0, \bar{\sigma}_6)$ ,  $\frac{\partial \Delta_{i,\sigma}}{\partial \alpha} > \kappa_2$ .

To conclude the proof, simply pick  $\bar{\sigma} = \min_{i \in \{1, \dots, 6\}} \bar{\sigma}_i$  and  $\kappa = \min(\kappa_1, \kappa_2)$ . ■

**Proof of Lemma 1:** Point (i) is straightforward and simply results from the assumption that for all  $i \in \{1, 2\}$ ,  $S_{-i} < W_{-i}$ . Points (ii) and (iii) are more delicate and make extensive use of Lemma 5. We prove (ii) and (iii) together.

Let us first show that if  $s_{-i}$  is a threshold-form strategy of threshold  $x_{-i}$ , then the best reply to  $s_{-i}$  is also threshold form. The best reply to  $s_{-i}$  is to play peace if and only if  $\Delta_{i,\sigma}(x_{i,t}, s_{-i}, V_{i,\sigma}(s_{-i})) \geq 0$ . Since the value  $V_{i,\sigma}(s_{-i})$  is constant, point (iii) of Lemma 5 holds and it follows from simple algebra that  $\Delta_{i,\sigma}(x_{i,t}, s_{-i}, V_{i,\sigma}(s_{-i})) \geq 0$  implies that  $\frac{\partial \Delta_{i,\sigma}}{\partial x_i} > 0$ . This single crossing condition implies that the best reply is to play peace if and only if  $x_{i,t} \geq x_i$  where  $x_i$  is the unique solution of  $\Delta_{i,\sigma}(x_i, s_{-i}, V_{i,\sigma}(s_{-i})) = 0$ . Hence the best reply to a threshold form equilibrium is a threshold form equilibrium.

Point (ii) of Lemma 5 also implies a form of monotone best reply. Consider  $s_{-i}$  and  $s'_{-i}$  two strategies and denote  $s_i$  and  $s'_i$  corresponding best replies of player  $i$ . Then whenever  $s_{-i}$  is threshold-form and  $s'_{-i} \preceq s_{-i}$ , then  $s'_i \preceq s_i$  (note that we also know that  $s_i$  is unique and takes a threshold form). We call this property restricted monotone best-reply. It allows to replicate part of the standard construction of Milgrom and Roberts (1990) and Vives (1990). Denote  $BR_{i,\sigma}$  and  $BR_{-i,\sigma}$  the best-reply mappings and  $s_P$  the strategy corresponding to playing peace always. We construct the sequence  $\{[BR_{i,\sigma} \circ BR_{-i,\sigma}]^k(s_P)\}_{k \in \mathbb{N}}$ . Since  $s_P$  is threshold-form (with threshold  $-\infty$ ) and is the highest possible strategy, this sequence is a decreasing sequence of threshold form strategies. Restricted monotone best-reply implies that it also converges to a strategy  $s_{i,\sigma}^H$  that is an upper bound to the set of equilibrium strategies of player  $i$ . Furthermore,  $(s_{i,\sigma}^H, s_{-i,\sigma}^H)$  is itself an equilibrium (where  $s_{-i,\sigma}^H = BR_{-i,\sigma}(s_{i,\sigma}^H)$ ) which takes a threshold form. Point (i) of Lemma 5 implies that the associated values are the highest equilibrium values. This concludes the proof. ■

Let us now turn to the proof of Proposition 2. We begin by characterizing the most peaceful equilibrium for fixed  $f$  as parameter  $\sigma$  goes to 0.

**Lemma 6 (characterizing the most peaceful equilibrium)** For any  $x \in \mathbb{R}$ , define

$$V_i(x) = \frac{1}{1 - \delta \text{prob}(\tilde{\pi} \geq x)} [\mathbb{E}(\tilde{\pi} \mathbf{1}_{\pi \geq x}) + \delta \text{prob}(\tilde{\pi} \leq x) W_i]$$

As  $\sigma$  goes to 0,  $\mathbf{x}_\sigma^{\mathbf{H}}$  converges to a symmetric pair  $(x^H, x^H)$  where  $x^H$  is the smallest value  $x$  such that for all  $i \in \{1, 2\}$ ,  $x + \delta V_i(x) \geq F_i$  and

$$\prod_{i \in \{1, 2\}} (x + \delta V_i(x) - F_i) = \prod_{i \in \{1, 2\}} (W_i - S_i) \quad (10)$$

**Proof of Lemma 6:** We begin by showing the following result: for any upper bound for values  $\bar{V} \in \mathbb{R}$ , there exists  $\bar{\sigma} > 0$  such that for any  $\sigma \in (0, \bar{\sigma})$ , for any  $(V_i, V_{-i}) \in [W_i, \bar{V}] \times [W_{-i}, \bar{V}]$ , the *one-shot* global game with payoffs

	$P$	$A$
$P$	$\tilde{\pi}_i + \delta V_i$	$S_i$
$A$	$F_i$	$W_i$

has a highest equilibrium that takes a threshold-form denoted by  $\mathbf{x}_\sigma^*(V_i, V_{-i}) = (x_{i,\sigma}^*, x_{-i,\sigma}^*)$ . Furthermore as  $\sigma$  goes to 0, the mapping  $\mathbf{x}_\sigma^* : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  converges uniformly over  $[W_i, \bar{V}] \times [W_{-i}, \bar{V}]$  to the mapping  $\mathbf{x}^* : (V_i, V_{-i}) \mapsto (x^{RD}(V_i, V_{-i}), x^{RD}(V_i, V_{-i}))$ .

The existence of a highest threshold form equilibrium results from point (ii) of Lemma 5. As in the dynamic case, one can prove a restricted form of monotone best-reply. Joint with the fact that best-replies to threshold-form strategies are also threshold form, iterative application of the best-reply mapping yields the result.

We now show uniform convergence. The proof uses point (iii) of Lemma 5. The equilibrium threshold  $\mathbf{x}_\sigma^*$  can be characterized as a pair  $(x_i, \alpha)$  where  $\alpha = (x_{i,\sigma}^* - x_{-i,\sigma}^*)/\sigma$  and  $x_i = x_{i,\sigma}^*$ . The pair  $(x_i, \alpha)$  must solve

$$\widehat{\Delta}_{i,\sigma}(x_i, \alpha, V_i) = 0 \quad (11)$$

$$\widehat{\Delta}_{-i,\sigma}(x_i - \alpha\sigma, -\alpha, V_{-i}) = 0. \quad (12)$$

As  $\sigma$  goes to 0,  $\widehat{\Delta}_{i,\sigma}$  converges uniformly to a mapping  $\widehat{\Delta}_i$ . Using point (iii) of Lemma 5 equations (11) and (12) imply that there exists  $\bar{\sigma}$  and  $\kappa > 0$  such that for all  $\sigma \in (0, \bar{\sigma})$  we must have

$$\forall i \in \{1, 2\}, \quad \frac{\partial \widehat{\Delta}_{i,\sigma}}{\partial x_i} > \kappa \quad \text{and} \quad \frac{\partial \widehat{\Delta}_{i,\sigma}}{\partial \alpha} > \kappa.$$

This implies that given  $x_i$  there is at most a unique value  $\alpha_\sigma(x_i)$  such that  $\Delta_{i,\sigma}(x_i, \alpha_\sigma(x_i), V_i) = 0$ . Since  $\frac{\partial \widehat{\Delta}_{i,\sigma}}{\partial \alpha} > \kappa > 0$  we also have that  $\alpha_\sigma(x_i)$  converges uniformly to the unique solution in  $\alpha$  of  $\widehat{\Delta}_i(x_i, \alpha, V_i) = 0$ . Furthermore, it must be that  $\alpha_\sigma(x_i)$  is decreasing in  $x_i$ . Define the mapping  $\zeta_\sigma(x_i) = \widehat{\Delta}_{-i,\sigma}(x_i - \alpha_\sigma(x_i)\sigma, -\alpha_\sigma(x_i), V_{-i})$ . The equilibrium threshold  $x_i$  must satisfy  $\zeta_\sigma(x_i) = 0$ . At any such  $x_i$ , we have that  $\zeta_\sigma$  is strictly increasing with slope greater than  $\kappa$ . Furthermore as  $\sigma$  goes to 0,  $\zeta_\sigma$  converges uniformly to a mapping  $\zeta$ . This yields that as  $\sigma$  goes to 0,  $x_{i,\sigma}^*$  must converge to the unique zero of  $\zeta$ . We know from the global games literature that this unique zero is  $x^{RD}(V_i, V_{-i})$ . This concludes the first part of the proof.

We now prove Lemma 6 itself. The highest equilibrium  $\mathbf{s}_\sigma^H$  of the dynamic game is associated with constant values  $\mathbf{V}_\sigma^H$  and constant thresholds  $\mathbf{x}_\sigma^H$ . This threshold has to correspond to a Nash equilibrium of the one-shot augmented global game

	$P$	$A$
$P$	$\tilde{\pi}_t + \delta V_{i,\sigma}^H$	$S_i$
$A$	$F_i$	$W_i$

where payoffs are given for row player  $i$ . Furthermore since  $\mathbf{s}_\sigma^H$  is the highest equilibrium of the dynamic game, it must be that  $\mathbf{x}_\sigma^H$  also corresponds to the highest equilibrium of the one-shot augmented global game. Hence  $\mathbf{x}_\sigma^H = \mathbf{x}_\sigma^*(\mathbf{V}_\sigma^H)$ . Let us denote by  $V_{i,\sigma}(x_{-i})$  the value player  $i$  obtains from best replying to a strategy  $s_{x_{-i}}$  and  $\mathbf{V}_\sigma(\mathbf{x}) = (V_{i,\sigma}(x_{-i}), V_{-i,\sigma}(x_i))$ . We have that  $\mathbf{V}_\sigma^H = \mathbf{V}_\sigma(\mathbf{x}_\sigma^H)$ . Together this yields that  $\mathbf{V}_\sigma^H$  is the highest solution of the fixed point equation  $\mathbf{V}_\sigma^H = \mathbf{V}_\sigma(\mathbf{x}_\sigma^*(\mathbf{V}_\sigma^H))$ . We know that  $\mathbf{x}_\sigma^*$  converges uniformly to the symmetric pair  $(x^{RD}, x^{RD})$ . Furthermore  $V_{i,\sigma}(x)$  converges uniformly over any compact to  $V_i(x)$ . Hence as  $\sigma$  goes to 0,  $\mathbf{V}_\sigma^H$  must converge to the highest solution  $\mathbf{V}^H$  of the fixed point equation  $\mathbf{V}^H = \mathbf{V}(x^{RD}(\mathbf{V}^H))$ . Equivalently,  $\mathbf{x}_\sigma^H$  must converge to the symmetric pair  $(x^H, x^H)$  where  $x^H$  is the smallest value such that  $x^H = x^{RD}(\mathbf{V}(x^H))$ . This yields that indeed  $x^H$  is the smallest value  $x$  such that for all  $i \in \{1, 2\}, x + \delta V_i(x) \geq F_i$  and  $\prod_{i \in \{1, 2\}} (x + \delta V_i(x) - F_i) = \prod_{i \in \{1, 2\}} (W_i - S_i)$ , which concludes the proof. ■

Using Lemma 6, Proposition 2 follows directly.

**Proof of Proposition 2:** As  $f_n$  converges to the Dirac mass  $d_\pi$ , the mapping  $V_{i,f_n}(x)$  converges to the mapping  $V_{i,d_\pi}(x) = \frac{1}{1-\delta}\pi \mathbf{1}_{x \leq \pi} + W_i \mathbf{1}_{x > \pi}$ . The conditions of Proposition 2 simply correspond to whether  $\pi > x^{RD}(\mathbf{V}(\pi))$  or  $\pi < x^{RD}(\mathbf{V}(\pi))$ . If  $\pi > x^{RD}(\mathbf{V}(\pi))$  then the value of permanent peace generates a cooperation threshold below  $\pi$  and hence permanent peace is self sustainable. If on the other hand  $\pi < x^{RD}(\mathbf{V}(\pi))$  then even the

value of permanent peace generates a cooperation threshold above  $\pi$  so that with very high probability immediate conflict occurs. This concludes the proof. ■

## A.2 Proofs for Section 3

**Proof of Proposition 3:** When  $k_i = k_{-i} = k$ , we have that  $\pi_{CI} = (1 - \delta)F(k)$  and  $\pi_{SU} = (1 - \delta)[F(k) + W(k) - S(k)]$ . Under Assumption 2,  $F$  is decreasing in  $k$ , and hence  $\pi_{CI}$  is decreasing in  $k$ . Clearly,  $\pi_{SU}$  is decreasing in  $k$  if and only if  $F'(k) + W'(k) - S'(k) < 0$ . ■

**Proof of Lemma 2:** Whenever  $D$  is convex over the range  $[\rho_S k, \rho_F k]$ , then  $\rho_S D'(\rho_S k)$  is increasing in  $\rho_S$  and  $\rho_F D'(\rho_F k)$  is increasing in  $\rho_F$ . Hence  $\phi$  is decreasing in  $\rho_F$  and increasing in  $\rho_S$ . ■

**Proof of Proposition 4:** When  $k_i = k_{-i} = k$ , then  $\pi_{SU} = (1 - \delta)(F(k) + W(k) - S(k))$ . We have that

$$\inf_{k \geq 0} \pi_{SU}(k) \leq (1 - \delta) \inf_{k \geq 0} F(k) + (1 - \delta) \inf_{k \geq 0} [W(k) - S(k)].$$

By Assumptions 2, and 3 we get that

$$\inf_{k \geq 0} \pi_{SU}(k) \leq (1 - \delta) \lim_{k \rightarrow \infty} F(k) + (1 - \delta) \lim_{k \rightarrow \infty} [W(k) - S(k)] = \lim_{k \rightarrow \infty} \pi_{SU}(k).$$

This concludes the proof. ■

**Proof of Proposition 5:** We have that

$$\frac{d\pi_{SU}}{dk} = \frac{dF}{dk} + \frac{dW}{dk} - \frac{dS}{dk} = -D'(\rho_S k) - D'(k) + D'(\rho_F k).$$

Using Assumption 4 and the fact that  $\rho_F > 1 > \rho_S$ , we obtain that at  $k = k^*/\rho_F$ ,  $d\pi_{SU}/dk = +\infty$ . Since  $\pi_{SU}$  is continuously differentiable in  $k$ , this concludes the proof. ■

**Proof of Lemma 3 :** Let us compute  $\pi_{SU}$  explicitly in the case where  $k_i$  may be different

of  $k_{-i}$ . The threshold  $\pi_{SU}$  is the only root of the second degree equation

$$\left(\frac{1}{1-\delta}\pi - F_i\right) \left(\frac{1}{1-\delta}\pi - F_{-i}\right) = (W_i - S_i)(W_{-i} - S_{-i})$$

that is also greater than  $\max_i F_i$ . This yields that

$$\pi_{SU} = \frac{F_i + F_{-i} + \sqrt{(F_i - F_{-i})^2 + 4(W_i - S_i)(W_{-i} - S_{-i})}}{2}$$

which can be re written as

$$\begin{aligned} \pi_{SU} &= \pi_{CI} + \frac{\sqrt{(F_i - F_{-i})^2 + 4(W_i - S_i)(W_{-i} - S_{-i})} - \sqrt{(F_i - F_{-i})^2}}{2} \\ &= \pi_{CI} + \eta \end{aligned}$$

where  $\pi_{CI} = (1 - \delta) \max\{F_i, F_{-i}\}$  and

$$\eta = \frac{\sqrt{(F_i - F_{-i})^2 + 4(W_i - S_i)(W_{-i} - S_{-i})} - \sqrt{(F_i - F_{-i})^2}}{2}.$$

Denote  $\mu_1 = (F_i - F_{-i})^2$  and  $\mu_2 = (W_i - S_i)(W_{-i} - S_{-i})$ . We have that

$$\frac{d\eta}{d\lambda} = \frac{1}{2} \left[ \frac{1}{2\sqrt{\mu_1 + 4\mu_2}} \left( \frac{d\mu_1}{d\lambda} + 4\frac{d\mu_2}{d\lambda} \right) - \frac{1}{2\sqrt{\mu_1}} \frac{d\mu_1}{d\lambda} \right].$$

Since  $\mu_1 > 0$ ,  $d\mu_1/d\lambda > 0$  and  $d\mu_2/d\lambda \leq 0$ , we obtain that  $d\eta/d\lambda < 0$ . This concludes the proof. ■

**Proof of Lemma 4:** In the benchmark model, we have that

$$W_i - S_i = \frac{\lambda}{1 + \lambda} m - \frac{\rho_S \lambda}{\rho_F + \rho_S \lambda} m - D(k) + D(\rho_F k) > -D(k) + D(\rho_F k).$$

Hence

$$\lim_{\lambda \rightarrow +\infty} W_i - S_i = -D(k) + D(\rho_F k) = \inf_{\lambda \geq 1} W_i - S_i.$$

We also have that

$$W_{-i} - S_{-i} = \frac{1}{1 + \lambda} m - \frac{\rho_S}{\rho_S + \rho_F \lambda} m - D(\lambda k) + D(\rho_F \lambda k) > 0.$$

Since by assumption  $D$  is increasing in  $k$  and bounded above, this yields that

$$\lim_{\lambda \rightarrow +\infty} W_{-i} - S_{-i} = 0 = \inf_{\lambda \geq 1} W_{-i} - S_{-i}.$$

This concludes the proof. ■

**Proof of Proposition 7:** When  $\lambda = 1$ , peace is sustainable under strategic risk if and only if  $\frac{1}{1-\delta}\pi \geq F(k) + W(k) - S(k)$ . In the benchmark model, this boils down to

$$\frac{1}{1-\delta}\pi \geq \frac{\rho_F}{\rho_F + \rho_S}m - D(\rho_S k) + \frac{1}{2}m - D(k) - \frac{\rho_S}{\rho_F + \rho_S}m + D(\rho_F k).$$

Hence when condition (5) holds, peace is not sustainable under strategic risk.

When weapon stocks are asymmetric ( $\lambda > 1$ ), then peace is sustainable under strategic risk if and only if

$$\prod_{i \in \{1,2\}} \left( \frac{1}{1-\delta}\pi - F_i \right)^+ > \prod_{i \in \{1,2\}} (W_i - S_i). \quad (13)$$

We have just shown that whenever  $D$  is bounded above, as  $\lambda$  goes to  $+\infty$  the difference  $W_{-i} - S_{-i}$  goes to 0. Since for all  $\lambda \geq 1$ ,  $F_i \geq F_{-i}$  and  $\lim_{\lambda \rightarrow +\infty} F_i = m - D(\rho_S k)$ , inequality (13) boils down to

$$\frac{1}{1-\delta}\pi > m - D(\rho_S k).$$

Hence condition (6) guarantees that as  $\lambda$  goes to  $+\infty$ , peace will be sustainable under strategic uncertainty. This concludes the proof. ■

### A.3 Proofs for Section 5

**Proof of Proposition 8:** Point (i) is obvious. As for point (ii), we have that  $\frac{1}{1-\delta}\pi_{CI}^T = \sum_{t=0}^{T-1} \delta^t f_t + \sum_{t=T}^{+\infty} \delta^t \pi'$ . Hence  $\pi_{CI}^{T+1} - \pi_{CI}^T = \delta^T(1-\delta)(f_T - \pi')$ . This concludes the proof. ■

**Proof of Proposition 9:** Point (i) holds since for  $T = 0$ , we have that  $W_i^T - S_i^T = 0$  and  $\frac{1}{1-\delta}\pi - F_i = \frac{1}{1-\delta}(\pi - \pi') > 0$ . This implies that  $(P, P)$  is indeed the risk-dominant equilibrium of the augmented one-shot game.

As for point (ii), we have that  $\frac{1}{1-\delta}\pi_{SU}^T = \sum_{t=0}^{T-1} \delta^t(w_t + f_t - s_t) + \sum_{t=T}^{+\infty} \delta^t \pi'$ . Hence  $\pi_{SU}^{T+1} - \pi_{SU}^T = \delta^T(1-\delta)(f_T + w_T - s_T - \pi')$ , which concludes the proof. ■

## References

- [1] Abreu, Dilip. 1988. "Towards a Theory of Discounted Repeated Games". *Econometrica* 56(2): 383-396.
- [2] Baliga, Sandeep and Tomas Sjöström. 2004. "Arms Races and Negotiations." *Review of Economic Studies* 71(2): 351-369.
- [3] Carlsson, Hans and Eric van Damme. 1993. "Global Games and Equilibrium Selection." *Econometrica* 61(5): 989-1018.
- [4] Chassang, Sylvain. 2007. "Fear of Miscoordination and the Sustainability of Cooperation in Dynamic Global Games with Exit." Working Paper.
- [5] Chassang, Sylvain. 2008. "Uniform Selection in Global Games." *Journal of Economic Theory* 139(1): 222-241.
- [6] Chassang, Sylvain, and Gerard Padró i Miquel. 2007. "Mutual Fears and Civil War." Working Paper.
- [7] Collier, Paul, Lani Elliott, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicolas Sambanis. 2003. *Breaking the Conflict Trap: Civil War and Development Policy*. Oxford and Washington DC: Oxford University Press and The World Bank.
- [8] Doyle, Michael and Nicholas Sambanis. 2006. *Making War & Building Peace: United Nations Peace Operations*. Princeton, NJ: Princeton University Press.
- [9] Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3): 379-414.
- [10] Flint, Julie and Alex de Waal. 2005. *Darfur: A Short History of a Long War*. London and New York: Zed Books.
- [11] Garfinkel, Michelle. 1990. "Arming as a Strategic Investment in a Cooperative Equilibrium" *American Economic Review* 80(4): 50-68.
- [12] Grossman, Herschel I. 1991. "A General Equilibrium Model of Insurrections." *American Economic Review* 81(4): 912-921.

- [13] Harsanyi, John and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*, MIT Press.
- [14] Hirshleifer, Jack. 1995. "Theorizing about Conflict." in Hartley, Keith and Sandler, Todd eds. *Handbook of Defense Economics*. Amsterdam: Elsevier Science.
- [15] Jackson, Matthew O. and Massimo Morelli. 2007. "Political Bias and War." *American Economic Review* 97(4): 1353-1373.
- [16] Jervis, Robert. 1976. *Perception and Misperception in International Politics*. New Jersey: Princeton University Press.
- [17] Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30(1): 167-214.
- [18] Jervis, Robert. 1979. "Deterrence Theory Revisited." *World Politics* 31(1): 289-324.
- [19] Jervis, Robert. 1989. *The Meaning of the Nuclear Revolution*. Ithaca, N.Y: Cornell University Press.
- [20] Jervis, Robert and Jack Snyder. 1999. "Civil War and the Security Dilemma." in Snyder, Jack and Barbara Walter, eds. *Civil War, Insecurity, and Intervention*. New York: Columbia University Press.
- [21] John, A. Andrew, Rowena Pecchenino and Stacey L. Schreft. 1993. "The Macroeconomics of Dr. Strangelove." *American Economic Review* 83(1): 43-62.
- [22] Kydd, Andrew H. 1997. "Game Theory and the Spiral Model." *World Politics* 49(3): 371-400.
- [23] Morris, Stephen and Hyun Song Shin. 2001. "Global Games: Theory and Applications." Cowles Foundation Discussion Paper No. 1275R. <http://ssrn.com/abstract=284813>.
- [24] Milgrom, Paul and John Roberts. 1990. "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities." *Econometrica* 58(6): 1255-1277.
- [25] Oxfam, IANSA and Safeworld. 2007. "Africa's Missing Billions: International Arms Flows and the Cost of Conflict." *Briefing Papers* 107.
- [26] Posen, Barry R. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35(1): 27-47.

- [27] Robert Powell. 1990. *Nuclear Deterrence Theory: The Search for Credibility*. New York: Cambridge University Press.
- [28] Powell, Robert. 2004. "The Inefficient Use of Power: Costly Conflict with Complete Information." *American Political Science Review* 98(3): 231-241.
- [29] Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- [30] Schelling, Thomas. 1966. *Arms and Influence*. New Haven, Conn.: Yale University Press.
- [31] Skaperdas, Stergios. 1992. "Cooperation, Conflict, and Power in the Absence of Property Rights." *American Economic Review* 82(4): 720-739.
- [32] Vives, Xavier. 1990. "Nash equilibrium with strategic complementarities ". *Journal of Mathematical Economics* 19(3): 305-321.