Conflict and Deterrence under Strategic Risk*

Sylvain Chassang
Princeton University
chassang@princeton.edu

Gerard Padró i Miquel
London School of Economics and NBER
g.padro@lse.ac.uk

January 2010

Abstract

We examine the determinants of cooperation and the effectiveness of deterrence 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 information about their environment. These two strategic settings allow us to identify and distinguish the role of predatory and preemptive incentives as determinants of cooperation and conflict. We show that while weapons unambiguously facilitate peace 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 cooperation. 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: cooperation, deterrence, strategic risk, global games, conflict, intervention, exit games.
JEL classification codes: D74, C72, C73

*We are grateful to Sandeep Baliga, Micael Castanheira, Jon Eguia, Joan Esteban, Christian Hellwig, Matt Jackson, Patrick Legros, David Miller, Andrea Prat, Jesse Shapiro, Joel Sobel and Flavio Toxvaerd for comments and conversations. The paper also benefited from feedback by seminar participants at Sussex University, Université Libre de Bruxelles, Rochester, UC Berkeley, UCSD, Cambridge University, and conference participants at the 2008 AEA meetings, the Polarization and Conflict Group Meeting at LSE, the UC Berkeley Conference on Endogenous Institutions and Conflict, as well as the ESSET meeting at Gerzensee. Andrew Robinson provided excellent research assistance. 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 – i.e. to cooperate – or to launch a surprise attack on each other. A peaceful equilibrium can only be sustained if the short-run gains from a surprise attack are counterbalanced by the long-run costs of triggering conflict. In this context, if both groups accumulate weapons, the cost of conflict increases, thereby improving incentives for peaceful behavior. This is the logic of deterrence, which reflects the idea frequently highlighted in the literature on repeated games that harsher punishments should improve incentives for cooperation.[1] The symmetric accumulation of weapons, insofar as it generates higher costs of war, should facilitate peace.

This paper examines the limits of this argument by contrasting the mechanics of cooperation and deterrence under complete information and under strategic risk, i.e. when players do not share a common understanding of their environment. While the complete information model suggests unambiguous predictions about the effect of weapons on peace, and about the impact of inequality on cooperation, these predictions need to be considerably nuanced once strategic risk is taken into account. We develop these points in detail and emphasize the importance of both predatory and preemptive incentives in determining the sustainability of cooperation under strategic risk.

We model conflict as a very stylized dynamic exit game, keeping grim-trigger strategies in a repeated game as a benchmark. 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 on to the next period. However, if one of the players attacks, conflict begins and players obtain exogenous continuation values.[2] Our model of strategic

[1]See for instance Abreu (1988) on penal codes. Garfinkel (1990) makes a similar point in the context of conflict and armament.

[2]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

risk follows the global games literature.[3] 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. At a state around which behavior switches, there will be a high probability that one player will choose peace while the other one attacks. This causes the players to second guess each other's move, and significantly affects the sustainability of peace. These effects remain even as the players' information becomes arbitrarily precise and we approach the complete information case. Throughout the paper we compare and contrast the conditions under which cooperation is sustainable in environments with and without strategic uncertainty.

To understand the difference that strategic risk makes, it is important to distinguish between the two motives for conflict that exist in this game. First, one may be tempted to attack an otherwise peaceful opponent – this is the *predatory motive* for conflict. Second, one may attack to avoid suffering a surprise strike from an opponent who is expected to be aggressive – this is the *preemptive motive* for conflict. Under complete information, it is easy for players to coordinate and only predatory motives matter. Under strategic uncertainty however, the sustainability of peace depends significantly on *both* predatory and preemptive incentives. Because weapon stocks can affect preemptive and predatory incentives differently, many comparative statics that were unambiguous under complete information become much more nuanced under strategic risk.

Our first set of results considers symmetric increases in weapon stocks. Under complete information, increased weapons stocks facilitate peace by diminishing payoffs upon conflict. Under strategic risk however, the symmetric accumulation of weapons may very well be destabilizing. Indeed, while weapons diminish predatory incentives, they may increase preemptive incentives if being the victim of a surprise attack is particularly weakening. It follows

exit framework encompasses the insights we obtain from a repeated prisoners' dilemma. See Chassang and Takahashi (2009) for a full-fledged analysis of repeated games under related incomplete information perturbations.

[3]See for instance Carlsson and van Damme (1993) and Morris and Shin (1998) for seminal work on global games, and Morris and Shin (2003) for a review.

that under general conditions the impact of weapons on peace will be non-monotonic. In particular, very large stocks of weapons (e.g. nuclear stocks sufficiently large to guarantee mutually assured destruction) will foster peace, whereas intermediate stocks of weapons (e.g. a few nuclear warheads that could be destroyed by a surprise strike) may be destabilizing.

Our second set of results explores how inequality in military strength affects stability. It is easy to show that unequal military power is always destabilizing under complete information. This is because inequality increases the predatory temptation of the stronger player. However, inequality reduces the preemptive motive for conflict for two reasons. First, the stronger player knows she has little to fear from the weaker one and hence she has smaller preemptive needs. Second, when the strong player is overwhelmingly dominant, the weaker player 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 player keeps enough weapons to limit the stronger player's predatory incentives. This suggests that restrained superiority may sustain the greatest level of peace.

Finally, we examine the impact of peace-enforcing interventions on peace and conflict. We first highlight that under complete information, unless intervention is immediate and war is prevented altogether, intervention will always have a destabilizing impact. Indeed, as in the familiar case of grim trigger strategies, it is precisely the prospect of a long and painful conflict that deters players from attacking in the first place. This conclusion, however, is not robust to strategic risk. By alleviating the potential costs of being the victim of a surprise attack, intervention reduces preemptive incentives. In that setting we show that the promise of intervention may promote peace even if it can only happen with delay.

This paper focuses entirely on the impact of strategic risk on the mechanics of deterrence and peace. As a result, the paper abstracts from a number of other realistic dimensions of conflict already emphasized in the literature. These include several frictions that induce bargaining failures, such as imperfect information (see Fearon (1995) or Powell (1999)), leader bias (see Jackson and Morelli (2007)), and commitment problems (as in Powell (2004) or Yared (2009)). Also, we do not consider the question of endogenous investment in weapons and the guns vs butter trade-off (see for instance Grossman (1991), Skaperdas (1992), Esteban and Ray (2008), as well as Jackson and Morelli (2009) who examine a model based on this trade-off that exhibits deterrence). Rather, our purpose here is to revisit a more primitive question: how does the accumulation of weapons affect the stability of peace?

While our contribution here is mostly applied, this paper also belongs to the recent theoretical literature on dynamic global games.[4] It is closely related to the work of Steiner (2008), Chassang (2009), Giannitsarou and Toxvared (2009), or Ordoñez (2009), all of which use a simple dynamic programming approach to simplify the analysis of large global games. In these papers, as well as in ours, payoff shocks are independent across periods and the focus is on how incomplete information affects the provision of incentives, rather than on how players may learn the underlying state of the world. A complementary literature focuses on such learning by considering dynamic global games in which the state is constant or follows a random walk. See for instance Chamley (1999), Angeletos, Hellwig and Pavan (2007), Dasgupta (2007) or Dasgupta, Steiner and Stewart (2008).

Because the exit game we consider can be thought of as a reduced form for trigger strategies in a repeated game, the basic insights of the paper can be applied in other environments usually modeled using repeated games. Whenever predatory and preemptive incentives move in different directions, taking strategic risk seriously will significantly affect comparative statics. One possible application is the model of price wars during booms of Rotemberg and Saloner (1986) which shows that collusion is hardest to sustain during times

[4]It is also useful to relate this paper to some of our other applied work on conflict. In a small extension of the current paper (Chassang and Padro i Miquel (2009a)), we use the framework developed here to discuss the relative merit of defensive weapons and defensive alliances as means to sustain peace. In an other recent paper (Chassang and Padro i Miquel (2009b)) we use a complete information model to discuss the impact of wealth on conflict in a context where wealth is expropriable. We highlight that it is temporary changes in wealth, rather than the level of wealth, that determine conflict. We note that in contrast to the current paper, considerations of strategic risk do not change the intuitions obtained in the complete information setting.

of temporary high demand since this is when predatory incentives are maximized. To the extent that preemptive incentives might be highest when demand is low (failing to react might put a firm out of business), introducing strategic risk may alter comparative statics. Similarly, the relational contracting literature (see for instance, Shapiro and Stiglitz (1985), Bull (1987), Baker, Gibbons and Murphy (1994, 2002), or Levin (2003)) often makes the point that reducing the players' outside option facilitates cooperation. This need not hold anymore in a model with strategic risk if reducing the players' outside option increases their incentives to preempt.

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 Appendix A.

2 Framework

2.1 A Simple Class of Cooperation Games

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 δ. Each period t, players simultaneously decide whether to be peaceful (P) or attack (A). If both players are peaceful at time t, they obtain a flow payoff π and the game moves on to period $t + 1$. When either of the players attacks, the game enters a conflict mode. Players receive an exogenously specified stream of payoffs and strategic interaction per-se ends. When player i attacks while $-i$ is peaceful, she is a first mover and gets a stream of payoffs $(f_{i,n})_{n\geq 0}$, where n denotes the number of periods elapsed since conflict began.[5] If the opposite happens, player i is a second mover and gets a stream of payoffs $(s_{i,n})_{n\geq 0}$. If both players attack at the same time, simultaneous

[5] i.e. if conflict started at time t, the flow payoff obtained by a first mover i at time $t + n$ is $f_{i,n}$.

war begins and player i gets a stream of payoffs $(w_{i,n})_{n\geq 0}$. We define F_i, S_i and W_i the present discounted values of starting conflict as a first, second or simultaneous mover. More specifically, we define,

$$F_i = \sum_{n=0}^{+\infty} \delta^n f_{i,n} \ ; \ \ S_i = \sum_{n=0}^{+\infty} \delta^n s_{i,n} \ ; \ \ W_i = \sum_{n=0}^{+\infty} \delta^n w_{i,n}.$$ [6]

Throughout the paper F_i, S_i and W_i will depend on the respective stocks of weapons k_i and k_{-i} of each player. More specifically, there are functions F, S and W such that,

$$F_i = F(k_i, k_{-i}) \ \ , \quad S_i = S(k_i, k_{-i}) \ \ , \quad W_i = W(k_i, k_{-i}).$$

Whenever $k_i = k_{-i} = k$, we use the notation $F_i = F(k)$, $S_i = S(k)$ and $W_i = W(k)$. We maintain the following assumption.

Assumption 1 *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 fairly 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. Throughout the paper, we discuss weapon stocks k_i and k_{-i} affect the sustainability of peace under different informational environments.

In any period t, given continuation values $(V_i)_{i\in\{0,1\}}$ upon joint cooperation, players can

[6] Note that trigger strategies in a repeated game are naturally mapped into this framework. Consider for instance, in the Prisoners' Dilemma, with stage game payoffs given by

	P	A
P	π	$-c$
A	b	0

where $\pi < b$ and $b - c < 2\pi$ so that peace is efficient. Trigger strategies correspond to payoffs upon conflict $f_{i,0} = b$, $s_{i,0} = -c$, $w_{i,0} = 0$ and $f_{i,n} = s_{i,n} = w_{i,n} = 0$ for $n > 0$.

be thought of as facing the one-shot game,

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

where payoffs are given for row player i.[7] This representation of payoffs allows us to identify two distinct motives for conflict. The payoff difference $F_i - \pi - \delta V_i$ corresponds to player i's *predatory incentives*, that is, how much player i would gain from attacking a consistently peaceful opponent. When players expect permanent peace upon continuation, predatory incentives take the form $F_i - \frac{1}{1-\delta}\pi$. 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 an opponent that is expected to attack. We make the following assumption.

Assumption 2 (early mover advantage) *For all $i \in \{1, 2\}$, $F_i > W_i > S_i$.*

Assumption 2 simply states that if conflict occurs, there is an advantage to attacking early. This assumption is natural in many instances of conflict, including military conflict, conflict between firms, or even conflict between individuals, as the first mover benefits from additional time to prepare her moves.

Throughout the paper, we contrast a situation in which the flow benefits of peace π are common knowledge, and a situation in which players make noisy but precise private assessments of the value of π. In the first case, common knowledge of payoffs allows players to coordinate their actions effectively and only predatory incentives matter for the sustainability of peace. Under incomplete information however, coordination becomes difficult as players attempt to second guess one another's value for peace. In that case the sustainability of peace depends significantly on both predatory and preemptive incentives.

Note that while we emphasize the players' uncertainty over the common returns to peace

[7] We look at a situation where the benefits of cooperation π are symmetric for the purpose of simplicity. Extending the model to a setting with asymmetric benefits presents no conceptual difficulty and simply adds to the notational burden.

π, our results would be identical if we consider uncertainty over the returns F from a surprise attack.[8] Indeed, it is uncertainty over predatory incentives $F - \pi - \delta V_i$ as a whole that drives our results. Note in addition that unfavorable economic shocks are in fact a major driver of conflict (see Miguel et al (2004) or Ciccone (2008)).

2.2 The Complete Information Benchmark

In the benchmark complete information setting, payoff π is fixed and common knowledge among players. We denote by Γ_{CI} the corresponding dynamic game.

Proposition 1 (cooperation under complete information) *Peace is (permanently) sustainable in an equilibrium of* Γ_{CI} *if and only if*

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

This means that under complete information, the sustainability of peace depends only on the magnitude of predatory incentives. Preemptive incentives play no role as neither S_i nor W_i enter condition (1). Note that this condition is analogous to the condition for cooperation in a Prisoners' Dilemma under grim trigger strategies. We denote by π_{CI} the smallest value of π such that inequality (1) holds. Let us turn to the case of strategic risk.

2.3 Strategic Risk

We model 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 what actions will be taken. This leads players to try to second guess each other's next move in order to avoid suffering a surprise attack. This second guessing is closely related to the idea of "reciprocal

[8] See Chassang (2009) for a general framework in which perturbations can affect all entries of the payoff matrix.

fear of surprise attacks" developed by Schelling (1960). We are ultimately interested in determining when such thought processes lead to an unraveling of peace.[9]

We consider an environment in which the returns to peace are not common knowledge. Specifically, we follow the framework of Chassang (2009) and consider the 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 g and support $(-\infty, +\infty)$. The payoff of cooperation $\tilde{\pi}_t$ 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 $[-1, 1]$, and $\sigma > 0$. For simplicity we assume that $\tilde{\pi}_t$ is observable in period $t + 1$ via the flow payoffs. Let us denote this game by $\Gamma_{\sigma,g}$.

To perform a robustness check on the complete information environment we are interested in the sustainability of peace in $\Gamma_{\sigma,g}$ as first, σ goes to 0, and second, g approaches a point mass at π.[10] This corresponds to an environment where players have approximately complete information about the state of the world, but remain uncertain about whether they are more or less optimistic than the other player. Analysis is facilitated by the fact that given a distribution g, as σ becomes small, game $\Gamma_{\sigma,g}$ admits a most peaceful equilibrium $\mathbf{s}^H_{\sigma,g}$ which sustains the highest equilibrium values $\mathbf{V}^H_{\sigma,g}$. Equilibrium $\mathbf{s}^H_{\sigma,g}$ also takes a simple threshold form, i.e., there exists $(x^H_{i,\sigma,g})_{i\in\{1,2\}} \in \mathbb{R}^2$ such that player i plays peace whenever she gets a signal $x_{i,t} \geq x^H_{i,\sigma,g}$ and attacks otherwise.[11]

[9] For a related model of reciprocal fears see Baliga and Sjoström (2004).

[10] Note that the order of limits we take is important. By taking σ to 0 first, we insure that the players always care about their private information, so that there is indeed second guessing and strategic risk. When we take the other order of limits, the players have such strong priors that they regard their private signals as completely noisy and we are essentially back in the complete information setting.

[11] See the appendix for more formal statements and proofs. It is important to note that we do not restrict

Before characterizing this most peaceful equilibrium in the limit case where players have very precise information, it is useful to delineate why a small amount of incomplete information can radically affect equilibrium behavior. For this purpose let us focus on the case where payoffs and signalling structures are symmetric. In that setting, the most peaceful equilibrium is symmetric with both players using the same threshold $x^H_{\sigma,g}$. With σ small, a player that gets a signal well below or well above $x^H_{\sigma,g}$ has little uncertainty about her opponent's behavior. The likelihood of surprise attacks is small. However, when a player gets as a signal the threshold $x^H_{\sigma,g}$, then there is roughly probability a half that her opponent got a higher signal and probability a half that her opponent got a lower signal. This means that an equilibrium threshold must be such that at that state of the world, a player is willing to be peaceful even though there is probability roughly a half that her opponent will launch a surprise attack. Note that in aggregate, the overall probability of a surprise attack maybe vanishing. What matters is that conditional on being at an equilibrium threshold, there is a high likelihood of an attack. This is why a small amount of incomplete information can significantly affect the way players interact even though, in aggregate, surprise attacks are quite rare.

We now characterize explicitly when peace can be sustained under strategic risk. For this purpose we introduce some notation. Given any pair $\mathbf{V} = (V_i, V_{-i})$ of continuation values, we consider the following 2×2 game $G(\mathbf{V})$

	P	A
P	$\pi + \delta V_i$	S_i
A	F_i	W_i,

where payoffs are given for row player i. Following Harsanyi and Selten (1988), we say that

attention to threshold-form strategies. Rather, we prove that game $\Gamma_{\sigma,g}$ admits a most peaceful equilibrium, and that this equilibrium is necessarily in threshold-form strategies.

$(Peace, Peace)$ is risk-dominant in game $G(\mathbf{V})$ if and only if

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

Inversely, we say that $(Attack, Attack)$ is risk-dominant if the opposite strict inequality holds. We also denote by

$$\overline{V} \equiv \frac{1}{1-\delta}\pi$$

the value of permanent peace. We can now state the main result of this section, which we use throughout the paper. Recall that $\mathbf{V}^H_{\sigma,g}$ denotes the highest equilibrium pair of values in game $\Gamma_{\sigma,g}$. It is supported by the most cooperative equilibrium.

Proposition 2 (cooperation under strategic risk) *For any sequence $\{g_n\}_{n\in\mathbb{N}}$ such that for all $n \in \mathbb{N}$, g_n has support $(-\infty, +\infty)$ and $\{g_n\}_{n\in\mathbb{N}}$ converges in mean to the unit mass at π, the following hold:*

(i) Whenever $(Peace, Peace)$ is risk-dominant in game $G(\overline{V}, \overline{V})$, then permanent peace is sustainable under strategic risk, in the sense that

$$\lim_{n\to\infty} \lim_{\sigma\to 0} \mathbf{V}^H_{\sigma,g_n} = \left(\overline{V}, \overline{V}\right).$$

(ii) Inversely, whenever $(Attack, Attack)$ is risk-dominant in game $G(\overline{V}, \overline{V})$, then peace is unsustainable under strategic risk, in the sense that

$$\lim_{n\to\infty} \lim_{\sigma\to 0} \mathbf{V}^H_{\sigma,g_n} = (W_i, W_{-i}).$$

Proposition 2 provides a convenient criterion to check whether peace is sustainable under strategic risk. Point (i) shows that when $(Peace, Peace)$ is risk-dominant in $G(\overline{V}, \overline{V})$, then the highest sustainable equilibrium value in $\Gamma_{\sigma,g}$ converges to the value of permanent peace $\overline{V}$, which implies that the most cooperative equilibrium of $\Gamma_{\sigma,g}$ sustains approximately permanent peace. Inversely, point (*ii*) shows that when $(Attack, Attack)$ is risk dominant, then the values associated with the most cooperative equilibrium of $\Gamma_{\sigma,g}$ converge to the value of immediate conflict. This implies that permanent conflict is the only equilibrium sustainable under strategic risk. Altogether, points (*i*) and (*ii*) imply that peace is robust to strategic risk if and only if

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

where $(z)^+ \equiv \max\{0, z\}$. Let us denote by π_{SR} the smallest value of π such that (2) holds.[12]

Condition (2) shows that just as under complete information, it is necessary that both players' predatory incentives $(F_i - \frac{1}{1-\delta}\pi)$ be negative to sustain peace.[13] In addition, condition (2) emphasizes the role of preemptive incentives $(W_i - S_i)$. The larger preemptive incentives are, the harder it is to sustain peace. When payoffs are symmetric, peace is sustainable under strategic risk if and only if $F - \frac{1}{1-\delta}\pi + W - S < 0$, i.e. peace is sustainable if and only if the sum of predatory and preemptive incentives is negative. As Sections 3, 4 and 5 show, there will often be a conflict between minimizing predatory incentives and minimizing preemptive incentives. As a consequence, taking strategic risk seriously can refine in important ways our understanding of cooperation and conflict.

From a modeling perspective, we consider the limit where the distribution g of returns from peace $\tilde{\pi}_t$ becomes concentrated around a given value π both for the purpose of tractability and because it allows us to focus exclusively on the role of preemptive incentives in determining the players' ability to cooperate. A drawback of taking this limit is that in our model, conflict either begins in the first period (if $(Peace, Peace)$ is not risk-dominant in $G(\overline{V}, \overline{V})$), or the likelihood of conflict in finite time is zero (if $(Peace, Peace)$ is risk-dominant in $G(\overline{V}, \overline{V})$). It is not difficult to resolve this problem since our analysis extends

[12]Note that the equilibrium in which players always attack is always robust to strategic risk. In fact, "attacking always" is an equilibrium of $\Gamma_{\sigma,g}$ for all σ and all g. As Chassang (2009) notes, in games with an infinite horizon, the global games perturbation cannot be used as a trick to select a unique equilibrium. Rather, the global games perturbation serves as a model of strategic risk in equilibrium that introduces preemption as a motive for conflict.

[13]Indeed, since $W_i - S_i > 0$, Condition (2) holds only if $\frac{1}{1-\delta}\pi - F_i > 0$ for $i \in \{1, 2\}$.

easily to circumstances where information is precise but the distribution g is not degenerate (see Chassang (2009)). In that case, the most peaceful equilibrium will still be in threshold strategies, but there will be some probability of conflict in each period depending on whether the realized return to peace $\tilde{\pi}_t$ is above the threshold or not. Most importantly, the equilibrium threshold will still be determined by risk-dominance concerns and the broad qualitative points we make in the paper would be unchanged. However, because payoffs upon conflict would now enter continuation values upon peace and change the potential surplus available in the game, the analysis would become richer and obscure the role played by preemptive incentives.

3 Deterrence with Symmetric Weapon Stocks

3.1 General Results

This section investigates how a symmetric increase in weapon stocks affects the sustainability of peace by studying the comparative statics of thresholds π_{CI} and π_{SR}. These thresholds correspond respectively to the minimum flow returns to peace π necessary for peace to be sustainable under complete information and under strategic risk. This implies that the lower π_{CI} and π_{SR} are, 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 π required to sustain peace.

The following proposition describes how the deterrent effect of weapons may differ across strategic settings. Recall that payoffs upon conflict F_i, S_i and W_i depend on the players' respective weapon stocks, k_i and k_{-i}. In addition, when weapon stocks are symmetric, i.e. $k_i = k_{-i} = k$, then all payoffs upon conflict are decreasing in k.

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

(i) π_{CI} is always strictly decreasing in k.

(ii) π_{SR} *is strictly decreasing in k if and only if*

$$\frac{dF}{dk} + \frac{dW}{dk} - \frac{dS}{dk} < 0. \tag{3}$$

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 facilitates the sustainability of peace by reducing predatory incentives. This holds independently of how weapons stocks affect W or S.

This prediction does not necessarily hold anymore once strategic risk is taken into account. Indeed, as point (ii) of Proposition 3 shows, the effectiveness of weapons as deterrent depends on their effect on preemptive incentives. If second movers suffer especially when weapon stocks increase, i.e. if $\frac{\mathrm{d}S}{\mathrm{d}k}$ is large and negative, the accumulation of weapons will increase preemptive incentives. As a consequence, whenever the value S of being a second mover falls more sharply than the value W of simultaneous war and the value F of initiating conflict, an increase in weapons will be destabilizing.[14] To make the discussion more specific and flesh out condition (3) we introduce the following benchmark model.

3.2 A Benchmark Model of Payoffs upon Conflict

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.

[14]Note that our interest here lies in the particular channel by which comparative statics in the benchmark model are overturned. There are many ways to change the complete information model so that larger symmetric weapon stocks are destabilizing. Consider for instance a complete information model where conflict occurs on the equilibrium path. Increasing weapon stocks may reduce continuation values upon peace more than it reduces the value of initiating conflict. In that model, increasing weapon stocks increases predatory incentives and preemptive incentives play no role. In contrast, a more refined prediction specific to our model with strategic risk is that changes in payoffs which reduce predatory incentives but increase preemptive incentives may in fact make cooperation harder.

Definition 1 (benchmark payoffs) *Payoffs upon conflict F, S and W are as follows*

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

(ii) $F(k_i, k_{-i}) = W(\rho_F k_i, \rho_S k_{-i})$ *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 $W(k_i, k_{-i})$ is a classic contest function.[15] It corresponds to the idea that players are competing for a prize m, and that the likelihood of obtaining m depends on the relative stocks of arms. The second term $D : \mathbb{R}^+ \to \mathbb{R}^+$ is a continuously differentiable increasing function that represents the amount of destruction incurred by player i upon conflict, independent of whether she wins prize m or not. We capture first mover advantage by allowing weapon stocks to be inflated or deflated by factors ρ_F and ρ_S depending on the timing of attacks. When weapon stocks are (k_i, k_{-i}) and player i unilaterally initiates conflict, it is as if players were engaged in a simultaneous conflict where weapon stocks are $(\rho_F k_i, \rho_S k_{-i})$. The difference $\rho_F - 1$ is positive and measures the increased effectiveness of a first mover's arsenal. We refer to $\rho_F - 1$ as the first mover advantage. The difference $1 - \rho_S$ is also positive and measures the decreased effectiveness of the second mover's arsenal. We refer to $1 - \rho_S$ as the second mover's disadvantage.[16] Note that payoffs F, S and W corresponding to this benchmark model satisfy Assumptions 1 and 2.

As of now we do not specify D any further, but we think of it as bounded (in the event of complete destruction, the amount of weapons used in the process does not change payoffs). The damage function D may also display convex parts. This may be because the way weapons are used changes as k increases (a team of ten soldiers with guns may cause more damage than ten times the damage of an individual soldier), or because the nature of weapons changes as k increases (for instance, rifles may be replaced by machine guns). Altogether, the typical damage function we envision is bounded with S-shaped portions.

[15] See for instance Hirshleifer (1995).

[16] Note that upon conflict, the payoffs of a first mover depend on the magnitude of both first mover advantage and second mover disadvantage.

3.3 Deterrence in the Benchmark Model

To better understand the circumstances in which weapons will be destabilizing, we now examine the meaning of condition (3) when payoffs upon conflict are those of our benchmark model. The threshold π_{SR} takes the form

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

Weapons are deterrent under strategic uncertainty if and only if

$$\frac{\mathrm{d}\pi_{SR}}{\mathrm{d}k} = -(1-\delta)[\rho_S D'(\rho_S k) + D'(k) - \rho_F D'(\rho_F k)] < 0;$$

accumulating weapons is counter-productive otherwise. The derivative $\frac{\mathrm{d}\pi_{SR}}{\mathrm{d}k}$ characterizes the destabilizing impact of marginal weapons. If it is positive and large, additional weapons will make it much harder to sustain peace. If it is negative and large, then additional weapons facilitate the sustainability of peace.

We are now interested in how the first strike advantage $\rho_F - 1$ and the second strike disadvantage $1 - \rho_S$ may affect the sign and magnitude of $\frac{\mathrm{d}\pi_{SR}}{\mathrm{d}k}$, i.e. the destabilizing impact of weapons. Fact 1 shows that under reasonable conditions a large first mover advantage and a large second mover disadvantage increase the destabilizing impact of weapons.

Fact 1 *If D is weakly convex over the range $[\rho_S k, \rho_F k]$, then $\frac{d\pi_{SR}}{dk}$ is increasing in ρ_F and decreasing in ρ_S.*

Fact 1 states that whenever the damage function D is weakly convex over the range $[\rho_S k, \rho_F k]$, then a large first strike advantage and a large second strike disadvantage will make weapons more destabilizing. Consider the case of a linear damage function D: if first mover advantage and second mover disadvantage are large, then when weapon stocks increase the amount of destruction suffered by second movers rises faster than the amount of destruction suffered by a first mover. The greater the first mover advantage, the greater this discrepancy and the more likely it is that weapons are destabilizing. Whenever D is not linear, this reasoning needs to be qualified. Indeed if the damage function is very concave, an increase in first mover advantage ρ_F may reduce the marginal impact of weapons on second mover damages $D(\rho_F k)$ to the point where $\rho_F D'(\rho_F k) < D'(k)$. This would reduce the destabilizing impact of weapons. Conversely, if the damage function D is convex over the range $[\rho_S k, \rho_F k]$, the destabilizing effects of first strike advantage and second strike disadvantage are magnified.

Interestingly, because the deterrent effect of weapons depends on the local shape of the destruction function D, the marginal effect of weapons will depend on existing weapon stocks. As a consequence, our model can generate rich comparative statics. In the following subsection, we highlight that under reasonable assumptions our model predicts that large stocks of weapons (i.e. enough to guarantee mutually assured destruction) are deterrent, while intermediate stocks of weapons (i.e. enough to cause damage, but small enough to be wiped by a surprise strike) may be destabilizing.[17]

3.4 Mutually Assured Destruction and Incapacitating Strikes

This section explores the possibility that different levels of weapons may have different deterrent effects. We introduce the following assumption.

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

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

[17] Another question concerns the deterrent impact of defensive versus offensive weapons. See Chassang and Padró i Miquel (2009a) on the subject.

This assumption corresponds to the idea that when weapon stocks are large, destruction is unavoidable and the benefits from preemption are minimized. This assumption is weaker than what is typically understood by mutually assured destruction. In its more narrow sense, mutually assured destruction would correspond to the assumption that $\lim_{k\to+\infty} F(k) = \lim_{k\to+\infty} W(k) = \lim_{k\to+\infty} S(k) = \inf_{k\geq 0} S(k)$. In that case, when weapon stocks are large, destruction is so complete that payoffs upon conflict are independent of who initiated the first attack. Clearly, we have that $\lim_{k\to+\infty} W(k) - S(k) = 0 = \inf_{k\geq 0} W(k) - S(k)$, so that Assumption 3 holds. Whenever Assumption 3 holds both predatory and preemptive incentives are minimized when weapon stocks become large. This yields the following result.

Fact 2 (MAD and stability) *If Assumption 3 holds, peace is most sustainable under strategic risk (or complete information) when the stock of weapons becomes arbitrarily large. More formally*

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

Note that the benchmark model of Section 3.2 satisfies Assumption 3 whenever the destruction function D is bounded above. In such a situation, when weapon stocks are symmetric, sufficient destructive power will guarantee the highest possible level of peace independently of whether there is strategic risk. Note that this does not imply an ambiguous statement about welfare. Indeed if we are in a situation where returns are low compared to threshold π_{SR}, then increasing weapon stocks does not change the fact that conflict will occur anyway and only reduces the players' welfare.

Fact 2 also does not imply that weapons monotonically increase stability in a world with strategic risk. In fact we now present a stark example highlighting how convexities in the destruction function D may cause intermediate stocks of weapons to be destabilizing.

Assumption 4 (disruptive weapons) *There exists a weapon level k^* such that*

$$D'(\rho_F k^*) = +\infty \quad \textit{and for all } \; k \neq \rho_F k^*, \quad D'(k) < +\infty. \tag{4}$$

Note that Assumption 4 is consistent with Assumption 3. This would be the case if the damage function D is bounded and S-shaped with a sharp inflexion point at $\rho_F k^*$. The stock k^* corresponds to a level of weapons at which the marginal damages $\rho_F D'(\rho_F k^*)$ caused by a first mover are much larger than the marginal damage $\rho_S D'(\rho_S k^*)$ caused by a second mover. Intuitively this corresponds to a level of weaponry where incapacitating strikes are possible. For instance, consider a situation where each party owns a few destructive weapons (airplanes, nuclear warheads...), which could be potentially wiped out by a surprise strike. Whenever the stock of weapons is close to k^*, increasing the stock of weapons will reduce the sustainability of peace.

Fact 3 (disruptive weapons precipitate conflict) *Whenever Assumption 4 holds, there exists an open interval $I \subset \mathbb{R}$ containing k^* such that π_{SR} is strictly increasing in k over I.*

While Assumption 4 facilitates the statement of Fact 3, the assumption that $D'(\rho_F k^*)$ be infinite for some stock of weapons k^* is by no means necessary. For instance, if the damage function D is S-shaped with a sufficiently steep inflexion point a similar result would hold.

Altogether, these results suggest that the relationship between peace and weapon stocks may be non-monotonic. Indeed under Assumptions 3 and 4 very large stocks of weapons ensuring MAD will facilitate peace, while intermediate stocks of weapon may precipitate war if incapacitating attacks are possible.

4 Stabilizing Asymmetry

In the previous section we analyzed the case of contenders with equal weapon stocks. We now turn to the question of how asymmetry in military strength affects the sustainability of peace. Asymmetry is parameterized by a constant $\lambda \in [1, +\infty)$ so that $k_i = \lambda k$ and $k_{-i} = k$. As λ becomes large, one player becomes arbitrarily stronger than the other.[18] Let us denote

[18]Note that starting from a symmetric pair of weapon stocks (k, k), the transition to any pair (k_i, k_{-i}) with $k_{-i} \leq k_i$ can be decomposed as a symmetric move to (k_{-i}, k_{-i}) (which we studied in Section 3) followed by an asymmetric increase of the stronger player's weapon stock (of the kind we study now) to the pair

by $\Delta^i_{pred} \equiv F_i - \frac{1}{1-\delta}\pi$ player i's predatory incentives and by $\Delta^i_{preempt} \equiv W_i - S_i$ player i's preemptive incentives.

A marginal increase in the relative stock of weapons λ affects predatory and preemptive incentives as follows

$$\frac{\mathrm{d}\Delta^i_{pred}}{\mathrm{d}\lambda} = k\frac{\partial F_i}{\partial k_i} > 0 \quad (5)$$

$$\frac{\mathrm{d}\Delta^{-i}_{pred}}{\mathrm{d}\lambda} = k\frac{\partial F_{-i}}{\partial k_i} < 0 \quad (6)$$

$$\frac{\mathrm{d}\Delta^i_{preempt}}{\mathrm{d}\lambda} = k\left[\frac{\partial W_i}{\partial k_i} - \frac{\partial S_i}{\partial k_i}\right] \quad (7)$$

$$\frac{\mathrm{d}\Delta^{-i}_{preempt}}{\mathrm{d}\lambda} = k\left[\frac{\partial W_{-i}}{\partial k_i} - \frac{\partial S_{-i}}{\partial k_i}\right]. \quad (8)$$

Equations (5) and (6) show that as asymmetry in military strength increases, the predatory temptation of the stronger player increases, while the predatory temptation of the weaker player decreases. We obtain the following result.

Proposition 4 (asymmetry is bad under complete information) *Keeping k constant, greater asymmetry makes peace harder to sustain under complete information. Formally, π_{CI} is strictly increasing in λ.*

Indeed, under complete information, condition (1) implies that the sustainability of peace is entirely determined by the predatory incentives of the stronger player. As equation (5) shows, these are unambiguously increasing with respect to λ.

Because we keep the weapon stock of the weaker country constant, greater values of λ are associated with both greater asymmetry and greater total weapon stocks $(\lambda + 1)k$. It is useful to highlight the role played by asymmetry rather than total weapon stocks. Here, as asymmetry and total weapon stocks increase, the sustainability of peace diminishes (under complete information). This contrasts with Section 3 where players have identical weapon stocks and symmetric increases in total weapon stocks $2k$ unambiguously increase the

(k_i, k_{-i}).

sustainability of peace (under complete information). The reason why we keep the weapon stocks of the weaker country constant will be clarified shortly.

As equations (7) and (8) highlight, the effect of inequality on preemptive incentives can be ambiguous for both players, since $\frac{\partial W_i}{\partial k_i}$ and $\frac{\partial S_i}{\partial k_i}$ have the same sign. In particular the following proposition shows that under reasonable conditions, increasing inequality will reduce the preemptive incentives of *both* players.

Proposition 5 (appeasing inequality) *Assume that conflict payoffs F_i, S_i and W_i are generated by the benchmark model of Definition 1.*

(i) If D is bounded above then

$$\lim_{\lambda\to+\infty}\Delta^i_{preempt}=\inf_{\lambda\geq 1}\Delta^i_{preempt}\quad\text{and}\quad\lim_{\lambda\to+\infty}\Delta^{-i}_{preempt}=\inf_{\lambda\geq 1}\Delta^{-i}_{preempt}.$$

(ii) If $D(k')$ is concave for $k'>\underline{\lambda}k$ then $\Delta^i_{preempt}$ and $\Delta^{-i}_{preempt}$ are decreasing in λ over $[\underline{\lambda},+\infty)$.

Point (i) states that when damage D is bounded above, then arbitrarily large inequality will minimize the preemptive incentives of both players. Point (ii) provides a local version of this result and shows that when D is concave, preemptive incentives are decreasing in inequality.

The reasons why the players' incentives to launch preemptive attacks can diminish with inequality λ are intuitive. The stronger player's incentives to preempt 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 fighting an overwhelmingly stronger opponent, she faces complete destruction and obtains similar payoffs whether she is a second mover or attacks simultaneously. Proposition 6 now shows that this effect is strong enough that in some circumstances peace is sustainable only when weapon stocks are sufficiently unequal.[19]

[19] Appendix B provides a complementary example in which peace is sustainable if weapon stocks are large and asymmetric, but not sustainable if weapon stocks are large and symmetric.

Proposition 6 (stabilizing asymmetry) *Assume that conflict payoffs F_i, S_i and W_i are generated by the benchmark model of Definition 1 and that D is bounded above. Whenever*

$$\frac{1}{1-\delta}\pi \quad < \quad \left[\frac{1}{2}+\frac{\rho_F-\rho_S}{\rho_F+\rho_S}\right] m - D(\rho_S k) - D(k) + D(\rho_F k) \tag{9}$$

$$\text{and } \frac{1}{1-\delta}\pi \quad > \quad m - D(\rho_S k) \tag{10}$$

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

Note that inequalities (9) and (10) can hold simultaneously since $D(\rho_F k) - D(k) > 0$ and $(\rho_F - \rho_S)/(\rho_F + \rho_S)$ approaches 1 when ρ_F is large compared to ρ_S.

Proposition 6 provides conditions under which peace is not sustainable if both players have the same stock of weapons k but becomes sustainable if one of the players becomes overwhelmingly strong.[20] Condition (9) ensures that peace is not sustainable under strategic risk when $\lambda = 1$. This simply corresponds to the negation of condition (2) for our benchmark model. Condition (10) implies that when a player becomes arbitrarily strong, predatory attacks remain unattractive. When these conditions hold together, peace is sustainable only if players are sufficiently unequal.

Note that the term $D(\rho_F k) - D(k)$, corresponding to the strong player's preemptive incentives, does not appear in inequality (10). Indeed, because the preemptive incentives $\Delta^{-i}_{preempt}$ of the weaker players go to 0, peace is approximately dominant for the weaker player and strategic risk no longer affects the players' behavior. The only term that matters now corresponds to the predatory 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 that predatory attacks are unattractive for the stronger player. This is the reason why we focus on comparative statics that keep the weapon stocks of the weaker party constant. Altogether, Proposition 6 suggests that

[20]For a given k and ρ_F, there is always a ρ_S small enough such that these two conditions hold simultaneously for a range of π.

restrained superiority may guarantee the highest level of peace. This relates to, but qualifies, the idea that a monopoly of violence facilitates peace. A natural question, which we discuss a little further in Appendix B, is whether one of the players may willingly give up some of its weapons to improve the sustainability of peace.

5 Conflict and Intervention

This section explores the impact of peace-keeping interventions on the sustainability of peace.[21] First, note that if peace-keeping interventions reestablished peace immediately, it is clear they would be beneficial. However, problems arise if peace-keeping operations can only reestablish peace with some delay. Indeed, a complete information model would predict that delayed peace-keeping operations are in fact destabilizing.[22] We show that this need not be the case anymore under strategic risk.

To understand whether late intervention can be effective, it is important to unbundle payoffs upon conflict as a discounted sum of flow payoffs, and consider how the timing of third-party peace-enforcing interventions affects peace and conflict. For simplicity, we consider the case of symmetric weapon stocks, so that

$$F = \sum_{n=0}^{+\infty} \delta^n f_n \ ; \ \ S = \sum_{n=0}^{+\infty} \delta^n s_n \ ; \ \ W = \sum_{n=0}^{+\infty} \delta^n w_n$$

where $\{f_n\}_{n\in\mathbb{N}}$, $\{s_n\}_{n\in\mathbb{N}}$ and $\{w_n\}_{n\in\mathbb{N}}$ are exogenously given streams of payoffs upon conflict, and n denote the number of periods elapsed since the initiation of conflict. We extend Assumption 2 so that flow payoffs satisfy $f_n \geq w_n \geq s_n$ for all $n \in \mathbb{N}$. Peace keeping interventions are characterized by a number of periods N, which is the delay after which players anticipate that conflict will be interrupted. Some settlement is then imposed and

[21] 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. 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.

[22] In the sense that they increase the minimum returns to peace π necessary to sustain peace.

players obtain flow payoffs $\pi' < \pi$ from then on. Hence, if intervention occurs with delay $N \geq 1$, players' payoffs upon civil war are

$$F^N = \sum_{n=0}^{N-1} \delta^n f_n + \frac{\delta^N}{1-\delta}\pi' \; ; \quad S^N = \sum_{n=0}^{N-1} \delta^n s_n + \frac{\delta^N}{1-\delta}\pi' \; ; \quad W^N = \sum_{n=0}^{N-1} \delta^t w_n + \frac{\delta^N}{1-\delta}\pi'.$$

When intervention occurs N periods after the initiation of conflict, the minimum returns to peace π such that peace is sustainable under complete information is

$$\pi_{CI}^N = (1-\delta) \sum_{n=0}^{N-1} \delta^n f_n + \delta^N \pi'. \tag{11}$$

Note that $\lim_{N \to +\infty} \pi_{CI}^N = \pi_{CI}$. We make the following assumption.

Assumption 5 (conflict as punishment) *We assume that $f_0 > \pi$ and for all $n \geq 1$, $f_n < \pi'$.*

This corresponds to a type of war in which there are short-term benefits to attacking (looting, social prestige...) followed by painful conflict payoffs (guerrilla and retaliation). This is also the pattern of flow payoffs upon conflict generated by trigger strategies in the repeated Prisoner's Dilemma with stage game payoffs

	P	A
P	π	$-c$
A	b	0

where $\pi < b$ and $b - c < 2\pi$ so that peace is efficient. Under trigger strategies, flow payoffs upon conflict for a first mover are $f_0 = b > \pi$ and for $n > 0$, $f_n = 0 < \pi'$.

The following result shows how expected intervention affects the sustainability of peace under complete information.

Proposition 7 (intervention under complete information) *Consider the complete in-*

formation game in which intervention occurs at time T. The following hold:

(i) whenever $N = 0$, $\pi_{CI}^{N} = \pi' < \pi$, i.e. peace is always sustainable;

(ii) whenever $N \geq 1$, then π_{CI}^{N} is decreasing in N and $\pi_{CI}^{N} > \pi_{CI}$.

Point (i) of Proposition 7 highlights that if intervention were immediate, then peace would be sustainable for any value of π. This happens because a first mover attacker never gets the one-shot benefit f_0 and only ever gets settlement payoffs $\pi' < \pi$. Point (ii) shows that in contrast, 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 postponing intervention improves the sustainability of peace, to the point that committing not to intervene induces the highest level of peace. The intuition is clear: peace in this model is sustained by the perspective of a long, drawn-out and painful war. An intervention stops such wars and hence increases predatory incentives.

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

$$\pi_{SR}^{N} = (1-\delta) \sum_{n=0}^{N-1} \delta^{n} (f_n + w_n - s_n) + \delta^{N} \pi'.$$

Proposition 8 (intervention under strategic risk) *If intervention occurs at time N, the following hold:*

(i) whenever $N = 0$, $\pi_{SR}^{N} = \pi' < \pi$, i.e. peace is always sustainable,

(ii) for any $N \geq 1$, the cooperation threshold under strategic risk π_{SR}^{N} is increasing *in N if and only if $f_N + w_N - s_N > \pi'$.*

Point (ii) of Proposition 8 highlights that even when only *delayed* intervention is feasible, expecting intervention can facilitate peace. This occurs because under strategic risk, intervention affects the sustainability of peace via two channels. On the one hand, it replaces future flow predatory payoffs f_n by π'. This is destabilizing as it increases predatory incentives ($f_n < \pi'$). On the other hand, intervention replaces flow preemptive incentives $w_n - s_n$ by 0. This is stabilizing because it improves the situation of the victim of a surprise attack, thereby reducing preemptive incentives. Whenever $f_n + w_n - s_n > \pi'$, 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_n upon simultaneous conflict are constant.

Fact 4 (converging and diverging conflicts) *Assume that for all $t \geq 0$, $w_n = w_0$. We have that*

(i) if $f_n - s_n$ is increasing in n for all $n \geq 0$, then π^N_{SR} is increasing in N and $\pi^N_{SR} < \pi_{SR}$;

(ii) if $f_n - s_n$ is decreasing in n for all $n \geq 0$ and there exists N^* such that $f_{N^*} + w_{N^*} - s_{N^*} \leq \pi'$, then for all $N \geq N^*$, π^N_{SR} is decreasing in N.

Point (i) of Corollary 4 states that when flow payoffs between first and second movers diverge with time, even the promise of delayed intervention at *some* time N will improve the stability of peace. Furthermore when delay is unavoidable, intervention should still 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 4 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 N^*, intervention unambiguously reduces the stability of peace. In this second case the intuition obtained under complete information survives. If intervention can only happen with delay greater than N^*, then increasing such delay (or abstaining from intervening altogether) will improve the chances of peace. This suggests that intervention is most suited when conflicts follow a diverging pattern.

6 Conclusion

This paper contrasts 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 greed – as a determinant of cooperation and conflict. 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 the players' temptation to launch predatory attacks, they may also increase the urgency to launch preemptive attacks. As a result we show that weapons need not always be deterrent. We show that under natural conditions, sufficiently large stocks of weapons will be deterrent, while intermediary stocks of weapons may be destabilizing. In particular we highlight the danger of holding stocks of weapons that allow for incapacitating strikes, i.e. levels of weapons such that second movers are hurt 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. As a result peace may be sustainable if groups are unequal, and unsustainable if groups are equal. The model, however, does not imply that a 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 is not true anymore under strategic risk, as intervention may reduce players' fears 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 very streamlined. 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 cooperation 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 individuals. Our results may also apply to non-violent conflict settings such as price wars between firms. 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, allowing for investment in both productive and military capital would be a useful direction for future research.

A Appendix: Proofs

A.1 Proofs for Section 2

Proof of Proposition 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 Proposition 2 follows closely Chassang (2008) and Chassang (2009). 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 A.1 and A.2.

Definition A.1 *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 *2×2 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:

$$\prod_{i \in \{1,2\}} (x + \delta V_i - F_i) = \prod_{i \in \{1,2\}} (W_i - S_i) \tag{12}$$

Definition A.2

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

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

(ii) *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}.

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

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

the payoffs[23] *player* i *expects upon playing* P *and* A.

(iv) *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}.

[23] We drop the σ subscript and the dependency on $h_{i,t}$ whenever doing so does not cause confusion.

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

$$\Delta_{i,\sigma}(h_{i,t}, s_{-i}, V_i) = U^P_{i,\sigma}(V_i, h_{i,t}, s_{-i}) - U^A_{i\sigma}(h_{i,t}, s_{-i}).$$

(vi) Given $x_i \in \mathbb{R}$ and $V_i \in \mathbb{R}$, for all $\alpha \in [-2, 2]$ we define $\widehat{\Delta}_{i,\sigma}(x_i, \alpha, V_i) = \Delta_{i,\sigma}(x_i, s_{x_i - \alpha\sigma}, V_i)$. [24]

Lemma A.1 (intermediary results) *There exists $\overline{\sigma} > 0$ and $\kappa > 0$ such that for all $\sigma \in (0, \overline{\sigma})$, 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 $\frac{\partial \hat{\Delta}_{i,\sigma}}{\partial x_i} > \kappa$ and $\frac{\partial \hat{\Delta}_{i,\sigma}}{\partial \alpha} > 0$. 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 $\frac{\partial \hat{\Delta}_{i,\sigma}}{\partial \alpha} > \kappa$.

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

$$\max\{U^P_{i,\sigma}(V, h_{i,t}, s_{-i}), U^A_{i,\sigma}(h_{i,t}, s_{-i})\} \geq \max\{U^P_{i,\sigma}(V', h_{i,t}, s'_{-i}), U^A_{i,\sigma}(h_{i,t}, s'_{-i})\}.$$

Indeed, since $F_i > W_i$ it follows that $U^A_{i,\sigma}(s_{-i}) \geq U^A_{i,\sigma}(s'_{-i})$. Also for any history $h_{i,t}$ such that $U^P_{i,\sigma}(V, h_{i,t}, s'_{-i}) \geq U^A_{i,\sigma}(h_{i,t}, s'_{-i})$, there must exist some value of $\tilde{\pi}_t$, occurring with with positive likelihood conditional on $h_{i,t}$, such that $\tilde{\pi}_t + \delta V \geq F_i$. 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 $\overline{\sigma}_1 > 0$ such that for all $\sigma \in (0, \overline{\sigma}_1)$, if $U^P_{i,\sigma}(V, s'_{-i}) \geq U^A_{i,\sigma}(s'_{-i})$ then, $\tilde{\pi}_t + \delta V > S_i$ with probability 1 conditional on $h_{i,t}$. This yields that whenever $U^P_{i,\sigma}(V, s'_{-i}) \geq U^A_{i,\sigma}(s'_{-i})$, then $U^P_{i,\sigma}(V, s_{-i}) \geq U^P_{i,\sigma}(V, s'_{-i})$. Since $U^P_{i,\sigma}(V, s_{-i}) \geq U^P_{i,\sigma}(V', s_{-i})$, this yields that indeed for all $\sigma \in (0, \overline{\sigma}_1)$,

$$\max\{U^P_{i,\sigma}(V, s_{-i}), U^A_{i,\sigma}(s_{-i})\} \geq \max\{U^P_{i,\sigma}(V', s'_{-i}), U^A_{i,\sigma}(s'_{-i})\}. \tag{13}$$

[24] Note that $\Delta_{i,\sigma}(x_i, s_{x_i - \alpha\sigma}, V_i)$ is a slight abuse of notation since the first argument of $\Delta_{i,\sigma}$ should be a history. Since threshold-form strategies only depend on the current signal, we only keep track of the relevant part of history h_i: the signal x_i.

Since for any strategy s''_{-i}, the value $V_i(s''_{-i})$ is the highest solution of the fixed point equation $V_i(s''_{-i})(h_{i,t}) = \max\{U_i^P(V_i(s''_{-i}), h_{i,t}, s''_{-i}), U_i^A(h_{i,t}, s''_{-i}))$, inequality (13) implies that for all $\sigma \in (0, \overline{\sigma}_1)$, $V_{i,\sigma}(s_{-i}) \geq V_{i,\sigma}(s'_{-i})$. 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, $\overline{\sigma}_2 > 0$ such that for all $\sigma \in (0, \overline{\sigma}_2)$, $\Delta_{i,\sigma}(h_{i,t}, s'_{-i}, V) \geq 0$ implies that $\tilde{\pi}_t + \delta V - F_i \geq 0 > S_i - W_i$. 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_ϵ and F_ϵ the distribution and c.d.f. of $\epsilon_{i,t}$ and define $G_\epsilon \equiv 1 - F_\epsilon$. Recall that g denotes the distribution of $\tilde{\pi}_t$. We have that

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

Since $S_i - W_i < 0$, there exists $\overline{\sigma}_3 > 0$ and $\tau > 0$ such that for all $\sigma \in (0, \overline{\sigma}_3)$, whenever $\hat{\Delta}_{i,\sigma}(x_i, \alpha, V_i) \geq 0$ then $\alpha \geq -2 + \tau$. Otherwise $F_\epsilon(\alpha + u)$ would be arbitrarily small and we would have $\hat{\Delta}_{i,\sigma}(x_i, \alpha, V_i) < 0$. Standard results on convolution products[25] show that as σ 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 $\overline{\sigma}_4$ and $\kappa_1 > 0$ such that whenever $\sigma \in (0, \overline{\sigma}_4)$, then $\frac{\partial \hat{\Delta}_{i,\sigma}}{\partial x_i} > k_1 > 0$.

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

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

[25] See for instance Lemma 8 of Chassang (2008)

We can now prove the claim made in Section 2.3 that for σ small, game $\Gamma_{\sigma,g}$ admits a most peaceful equilibrium taking a threshold form. The proof makes extensive use of Lemma A.1.

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 A.1 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 $\Delta_{i,\sigma}(x_i, s_{-i}, V_i(s_{-i})) = 0$. Hence the best reply to a threshold form strategy is a threshold form strategy.

Point (ii) of Lemma A.1 also implies a form of monotone best reply. Consider two strategies s_{-i} and s'_{-i}, and denote by s_i and s'_i the 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 us to replicate part of the standard construction of Milgrom and Roberts (1990) and Vives (1990). Denote by $BR_{i,\sigma}$ 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 $s^H_{i,\sigma}$ that is an upper bound to the set of equilibrium strategies of player i. Furthermore, $(s^H_{i,\sigma}, s^H_{-i,\sigma})$ is itself an equilibrium (where $s^H_{-i,\sigma} = BR_{-i,\sigma}(s^H_{i,\sigma})$) which takes a threshold form. Point (i) of Lemma A.1 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 g as parameter σ goes to 0.

Lemma A.2 (characterizing the most peaceful equilibrium) *For any $x \in \mathbb{R}$, define*

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

As σ goes to 0, $\mathbf{x}^{\mathbf{H}}_{\sigma}$ 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

$$\prod_{i\in\{1,2\}} (x + \delta V_i(x) - F_i) = \prod_{i\in\{1,2\}} (W_i - S_i). \tag{14}$$

Proof of Lemma A.2: We begin by showing the following result: for any upper bound for values $\overline{V} \in \mathbb{R}$, there exists $\overline{\sigma} > 0$ such that for any $\sigma \in (0, \overline{\sigma})$ and for any $(V_i, V_{-i}) \in [W_i, \overline{V}] \times [W_{-i}, \overline{V}]$, the *one-shot* global game with payoffs

	P	A
P	$\tilde{\pi}_t + \delta V_i$	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 σ goes to 0, the mapping $\mathbf{x}^*_\sigma : \mathbb{R}^2 \to \mathbb{R}^2$ converges uniformly over $[W_i, \overline{V}] \times [W_{-i}, \overline{V}]$ 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 A.1. 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 A.1. The equilibrium threshold $\mathbf{x}^*_\sigma$ can be characterized as a pair $(x^*_{i,\sigma}, \alpha)$ where $\alpha = (x^*_{i,\sigma} - x^*_{-i,\sigma})/\sigma$. The pair (x_i, α) must solve

$$\widehat{\Delta}_{i,\sigma}(x_i, \alpha, V_i) = 0 \qquad (15)$$

$$\widehat{\Delta}_{-i,\sigma}(x_i - \alpha\sigma, -\alpha, V_{-i}) = 0. \qquad (16)$$

As σ goes to 0, $\widehat{\Delta}_{i,\sigma}$ converges uniformly to a mapping $\widehat{\Delta}_i$. Using point (iii) of Lemma A.1 equations (15) and (16) imply that there exists $\overline{\sigma}$ and $\kappa > 0$ such that for all $\sigma \in (0, \overline{\sigma})$ we must have

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

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 α 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 $x^*_{i,\sigma}$ must satisfy $\zeta_\sigma(x^*_{i,\sigma}) = 0$. At any such $x^*_{i,\sigma}$, we have that ζ_σ is strictly increasing with slope greater than κ. Furthermore, as σ goes to 0, ζ_σ converges uniformly to a mapping ζ. This yields that as σ goes to 0, $x^*_{i,\sigma}$ must converge to the unique zero of ζ. 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 A.2 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 let $\mathbf{V}_\sigma(\mathbf{x}) = (V_{i,\sigma}(x_{-i}), V_{-i,\sigma}(x_i))$. 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 set to $V_i(x)$. Hence as σ goes to 0, V_σ^H must converge to the highest solution $\mathbf{V}_H$ of the fixed point equation $\mathbf{V}^H = \mathbf{V}(\mathbf{x}^{RD}(\mathbf{V}^H))$. 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 A.2, Proposition 2 follows directly.

Proof of Proposition 2: As g_n converges to the Dirac mass d_π, the mapping $V_{i,g_n}(x)$ 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 π 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 π 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

$\pi_{SR} = (1-\delta)[F(k) + W(k) - S(k)]$. Under Assumption 1, F is decreasing in k, and hence π_{CI} is decreasing in k. Clearly, π_{SR} is decreasing in k if and only if $F'(k)+W'(k)-S'(k) < 0$. ■

Proof of Fact 1: Whenever D is convex over the range $[\rho_S k, \rho_F k]$, then $\rho_S D'(\rho_S k)$ is increasing in ρ_S and $\rho_F D'(\rho_F k)$ is increasing in ρ_F. Hence ϕ is decreasing in ρ_F and increasing in ρ_S. ■

Proof of Fact 2: The complete information case is immediate and we focus on the case of strategic risk.

When $k_i = k_{-i} = k$, then $\pi_{SR} = (1-\delta)(F(k) + W(k) - S(k))$. We have that

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

By Assumptions 1, and 3 we get that

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

This concludes the proof. ■

Proof of Fact 3: We have that

$$\frac{\mathrm{d}\pi_{SR}}{\mathrm{d}k} = \frac{\mathrm{d}F}{\mathrm{d}k} + \frac{\mathrm{d}W}{\mathrm{d}k} - \frac{\mathrm{d}S}{\mathrm{d}k} = -\rho_S D'(\rho_S k) - D'(k) + \rho_F D'(\rho_F k).$$

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

Proof of Proposition 5: We first prove point (i). In the benchmark model, we have that

$$W_i - S_i = \frac{\lambda}{1+\lambda} m - \frac{\rho_S \lambda}{\rho_F + \rho_S \lambda} m - D(k) + D(\rho_F k) > -D(k) + D(\rho_F k). \tag{17}$$

Hence

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

We also have that

$$W_{-i} - S_{-i} = \frac{1}{1+\lambda}m - \frac{\rho_S}{\rho_S + \rho_F\lambda}m - D(\lambda k) + D(\rho_F \lambda k) > 0. \tag{18}$$

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

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

This proves point (i). Point (ii) follows from taking derivatives in equations (17) and (18) and using the fact that $D'(\lambda \rho_F k) < D'(\lambda k)$. ■

Proof of Proposition 6: 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 (9) 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

$$\prod_{i \in \{1,2\}} \left(\frac{1}{1-\delta}\pi - F_i \right)^+ > \prod_{i \in \{1,2\}} \left(W_i - S_i \right). \tag{19}$$

We have just shown that whenever D is bounded above, as λ goes to $+\infty$ the difference $W_{-i} - S_{-i}$ goes to 0. Since for all $\lambda \geq 1$, $F_i \geq F_{-i}$ and $\lim_{\lambda \to +\infty} F_i = m - D(\rho_S k)$, inequality (19) boils down to

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

Hence condition (10) guarantees that as λ goes to $+\infty$, peace will be sustainable under strategic uncertainty. This concludes the proof. ■

A.3 Proofs for Section 5

Proof of Proposition 7: 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 8: 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 $\frac{1}{1-\delta}\pi_{SR}^T = \sum_{t=0}^{T-1} \delta^t(w_t + f_t - s_t) + \sum_{t=T}^{+\infty} \delta^T \pi'$. Hence $\pi_{SR}^{T+1} - \pi_{SR}^T = \delta^T(1-\delta)(f_T + w_T - s_T - \pi')$, which concludes the proof. ■

B Sustaining Peace with Asymmetric Weapon Stocks

In this appendix, we provide an example such that under strategic risk, large and asymmetric weapon stocks will sustain peace, whereas large and symmetric weapon stocks won't.

We use the benchmark payoffs described in Section 3.2, and make the following assumptions:

(B.i) the early mover advantage is sufficiently large that

$$(\rho_F - \rho_S)/(\rho_F + \rho_S) + 1/2 > 1;$$

(B.ii) the damage function D is bounded above, with $\overline{D} \equiv \sup_{k \geq 0} D(k)$;

(B.iii) the value m from victory is such that

$$\frac{1}{1-\delta}\pi \quad < \quad \left(\frac{1}{2} + \frac{\rho_F - \rho_S}{\rho_F + \rho_S}\right) m - \overline{D} \tag{20}$$

$$\frac{1}{1-\delta}\pi \quad > \quad m - \overline{D}. \tag{21}$$

Note that point (B.i) above implies that there exist a range of values m satisfying point (B.iii).

Straight forward algebra shows that under Condition (20), if weapon stocks k_i and k_{-i} grow arbitrarily large with $k_i = k_{-i}$, then asymptotically, peace is not sustainable under strategic risk.[26] In contrast, Condition (21) implies that if weapon stocks k_i and k_{-i} grow arbitrarily large with the ratio k_{-i}/k_i going to 0, then asymptotically, peace is sustainable under strategic risk.

[26]The result holds if the ratio k_i/k_{-i} goes to 1.

This occurs because when players are of equal strength, preemptive incentives are asymptotically equal to $[1/2 - \rho_S/(\rho_F + \rho_S)]m$ which is strictly positive given Assumption (B.i) above. If instead both players have high weapon stocks, but one player is arbitrarily stronger than the other, preemptive incentives for the stronger player go to 0 since she is obtains victory payoff m independently of whether she attacks first or second. As a consequence the right hand term of Condition (2) is equal to zero and the sustainability of peace is guaranteed if and only if the stronger player has negative predatory incentives, which follows from Condition (21).

This suggests that in some settings one player may be willing to disarm unilaterally in order to facilitate the sustainability of peace. However, our model however is too streamlined to treat properly the question of demilitarization. In particular, in a model where conflict does happen with positive probability on the equilibrium path, weapons have value on the equilibrium path and each player will prefer the other player to be the one relinquishing weapons. A more reasonable prediction is that a well armed player may be quite willing to tolerate the unilateral acquisition of weapons by a peaceful opponent.

References

[1] Abreu, Dilip. 1988. "Towards a Theory of Discounted Repeated Games". *Econometrica* 56(2): 383-396.

[2] Angeletos, George-Marios, Christian Hellwig and Alessandro Pavan. 2007. "Dynamic Global Games of Regime Change: Learning, Multiplicity, and the Timing of Attacks." *Econometrica.* 75(3): 711-756.

[3] Baker, George, Robert Gibbons and Kevin Murphy. 1994. "Subjective performance measures in optimal incentive contracts." *Quarterly Journal of Economics.* 109(4): 1125-56.

[4] Baker, George, Robert Gibbons and Kevin Murphy. 2002. "Relational Contracts and the Theory of the Firm." *Quarterly Journal of Economics.* 117(1):39-83.

[5] Baliga, Sandeep and Tomas Sjöström. 2004. "Arms Races and Negotiations." *Review of Economic Studies* 71(2): 351-369.

[6] Bull, Clive. 1987. "The Existence of Self-Enforcing Implicit Contracts." *Quarterly Journal of Economics.* 102(1): 147-59.

[7] Carlsson, Hans and Eric van Damme. 1993. "Global Games and Equilibrium Selection." *Econometrica* 61(5): 989-1018.

[8] Chamley, Christophe. 1999. "Coordinating Regime Switches." *Quarterly Journal of Economics.* 114(3): 869-905.

[9] Chassang, Sylvain. 2008. "Uniform Selection in Global Games." *Journal of Economic Theory* 139(1): 222-241.

[10] Chassang, Sylvain. 2009. "Fear of Miscoordination and the Sustainability of Cooperation in Dynamic Global Games with Exit." Forthcoming in *Econometrica*.

[11] Chassang, Sylvain, and Gerard Padró i Miquel. 2009a. "Defensive Weapons and Defensive Alliances." *American Economic Review, Papers and Proceedings.* 99(2): 282-286.

[12] Chassang, Sylvain, and Gerard Padró i Miquel. 2009b. "Economic Shocks and Civil War."Forthcoming in the *Quarterly Journal of Political Science.*

[13] Chassang, Sylvain and Satoru Takahashi. 2009. "Robustness to Incomplete Information in Repeated Games." Princeton University working paper.

[14] Ciccone, Antonio. 2008. "Transitory Economic Shocks and Civil Conflict." Universitat Pompeu Fabra working paper.

[15] Dasgupta, Amil. 2007. "Coordination and Delay in Global Games." *Journal of Economic Theory.* 134(1):195-225.

[16] Dasgupta, Amil, Jakub Steiner and Colin Stewart. 2008. "Efficient Dynamic Coordination with Individual Learning." LSE Working paper.

[17] Esteban, Joan and Debraj Ray. 2008. "On the Saliency of Ethnic Conflict." *American Economic Review* 98:2185-2202.

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

[19] Garfinkel, Michelle. 1990. "Arming as a Strategic Investment in a Cooperative Equilibrium " *American Economic Review* 80(4): 50-68.

[20] Giannitsarou, Chrissy and Flavio Toxvared. 2009. "Recursive Global Games." Cambridge University working paper.

[21] Grossman, Herschel I. 1991. "A General Equilibrium Model of Insurrections." *American Economic Review* 81(4): 912-921.

[22] Harsanyi, John and Reinhart Selten. 1988. *A General Theory of Equilibrium Selection in Games*, MIT Press.

[23] Hirshleifer, Jack. 1995. "Theorizing about Conflict." in Hartley, Keith and Sandler, Todd eds. *Handbook of Defense Economics*. Amsterdam: Elsevier Science.

[24] Jackson, Matthew O. and Massimo Morelli. 2007."Political Bias and War." *American Economic Review* 97(4): 1353-1373.

[25] Jackson, Matthew O. and Massimo Morelli. 2009."Strategic Militarization, Deterrence and Wars." Stanford University working paper.

[26] Miguel, Edward, Shanker Satyanath and Ernest Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach" Journal of Political Economy 112(4):725-753.

[27] Milgrom, Paul and John Roberts. 1990. "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities." *Econometrica* 58(6): 1255-1277.

[28] Morris, Stephen and Hyun Song Shin. 2003. "Global Games: Theory and Applications." Cowles Foundation Discussion Paper No. 1275R. http://ssrn.com/abstract=284813.

[29] Ordoñez, Guillermo. 2009. "Fragility of Reputation and Clustering of Risk-Taking." Yale University working paper.

[30] Powell, Robert. 1999. *In the Shadow of Power: States and Strategies in International Politics*. Princeton: Princeton University Press.

[31] Powell, Robert. 2004. "The Inefficient Use of Power: Costly Conflict with Complete Information." *American Political Science Review* 98(3): 231-241.

[32] Rotemberg, Julio and Garth Saloner. 1986. "A Supergame-Theoretic Model of Price Wars during Booms". *American Economic Review.* 76(3):390-407.

[33] Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

[34] Shapiro, Carl and Joseph Stiglitz. 1984. "Equilibrium Unemployment as a Worker Descipline Device." *American Economic Review.* 74(3):433-444.

[35] Skaperdas, Stergios. 1992. "Cooperation, Conflict, and Power in the Absence of Property Rights." *American Economic Review* 82(4): 720-739.

[36] Steiner, Jakub. 2008. "Coordination of Mobile Labor". *Journal of Economic Theory.* 139(1): 25-46.

[37] Vives, Xavier. 1990. "Nash Equilibrium with Strategic Complementarities ". *Journal of Mathematical Economics* 19(3): 305-321.

[38] Yared, Pierre. 2009. "A Dynamic Theory of War and Peace". Columbia University mimeo.