A Theory of Brinkmanship, Conflicts, and Commitments*

Michael Schwarz[†] and Konstantin Sonin[‡]

December 20, 2004

Abstract

Many conflicts and negotiations can be viewed as a dynamic game, where parties have no commitment power. In our model, a potential aggressor demands concessions from the weaker party threatening a war. The absence of commitment makes a continuous stream of transfers a more effective appeasement strategy than a lump sum transfer. Based on such strategy, it is possible to construct a self-enforcing peace agreement between risk-neutral parties even if transfers shift the balance of power. When parties are risk averse, a self-enforcing peace agreement may not be feasible. The bargaining power of the potential aggressor increases dramatically if she is able to make probabilistic threats, e.g. by taking an observable action that leads to a war with positive probability. The 'brinkmanship strategy' allows a blackmailer to extract a positive stream of payments from the victim even if carrying out the threat is harmful to both parties. Although we use the terminology of international relations, our results are applicable to environments ranging from diplomacy to negotiations within or among firms, and are aimed to bring together 'parallel' investigations in the nature of commitment in economics and political science.

---

*The authors are grateful to Daron Acemoglu, Andrei Shleifer, Luigi Zingales, and conference and seminar participants at the European Meeting of the Econometric Society, Kellogg Business School, Pennsylvania State University, and UCSC for valuable comments.

[†]University of California at Berkeley and NBER; mschwarz@berkeley.edu. Part of this paper was written while this author was a National Fellow at the Hoover Institution at Stanford. He would like to express his gratitude to the Hoover Institution for enabling him to spend a productive year at Stanford.

[‡]Institute for Advanced Study, Princeton, NJ, New Economic School/CEFIR, Moscow, and CEPR; ksonin@ias.edu.

1 Introduction

When the rifle was invented, bows and arrows became obsolete. After the invention of nuclear weapons many observers expected that the conventional weapons would go the way of bows and arrows. For instance, one of the leading columnists of his time, Drew Pearson, declared in the Fall of 1945 that: "the Navy and huge armies are now obsolete".[1] Later it became apparent that nuclear weapons do not give expected political leverage. In 1957, reflecting on the time when the US had a nuclear monopoly, Henry Kissinger commented with a measure of surprise that "We never succeeded in translating our military [nuclear] superiority into a political advantage." Today, the verdict of history is clear: conventional arsenal remains important in the nuclear era.

Unfortunately, history is an infallible judge that never writes opinions to explain its verdicts. This paper offers a framework that deepens our understanding of the role of threats in the environments ranging from international negotiations to negotiations among or within firms. We show that an "indivisible" threat such as such as nuclear capabilities gives a party far less bargaining power than a threat that is divisible, such as a large arsenal of conventional weapons. We also show that a self-enforcing peace agreement between negotiating parties should rely on a sequence of transfers or pre-negotiated concessions (as opposed to one large concession that buys peace). Fearon (1995) argues that indivisibility of the object of bargaining might be a rationalist explanation for conflicts,[2] while Powell (2004) argues that the commitment problem lies in the heart of the 'indivisibility argument': if commitment is possible, a lottery over the indivisible object would be Pareto-superior to any conflict. We focus on *how* the ability to make observable probabilistic or divisible threats affects parties' bargaining power.

The study of blackmail received little attention from economists. At the first glance, this may seem justified because relevance of blackmail may appear to be limited to small time crooks and nations with nuclear capabilities. However, in a broader sense many economic relationships involve an element of blackmail. For instance, a manager induces a worker to put forth more effort by a threat of firing, a tenant wants the landlord to do timely maintenance threatening to terminate the lease etc. Legal scholars devoted considerable

[1]For analysis of evolution of conventional wisdom about the role of nuclear weapons in national defense, see Besse and Lasswell (1950).

[2]Early studies of war as a bargaining process are Schelling (1966), Wagner (1982), and Pillar (1983). Recent models of negotiations in/and wars include Acemoglu and Robinson (2001, 2002, 2005), Fearon (1995), Filson and Werner (2002), Smith and Stam (2003), Powell (2002, 2004), Slantchev (2003), Wittman (2003), Baliga and Sjostrom (2004).

attention to blackmail.[3] The celebrated Coase theorem (Coase, 1962) was in part an outcome of a decades-long intellectual inquiry of the author in legal questions related to blackmail (Coase, 1988).[4] A defining feature of a successful blackmail is that the threat must be credible (e.g., Shavell, 1992). Shavell and Spier (2002) analyze blackmail in the absence of binding commitment. In particular, they find that if carrying out a threat is costly, then extracting a payment from a victim is impossible even in an infinite horizon setting. The main innovation of our paper relative to Shavell and Spier (2002) is that we introduce brinkmanship into the model. We define brinkmanship as an ability of the aggressor to undertake an observable action that will lead with positive probability to a war or some other mutually undesirable outcome; "shooting over the head" is a metaphor for the kind of brinkmanship we have in mind. We show that being able to engage in this kind of visible brinkmanship may allow the potential aggressor to extract the whole surplus from the victim.

Though we employ the international-relations terminology in this paper, applications of this idea stretch far beyond the nuclear power politics. For example, in a relationship between a manager and an employee we observe the same pattern: if the only options available to a manager are to fire or to retain a worker, a worker only has to put forth enough effort to make the manager indifferent between firing and retaining him. All the bargaining power resides with the worker. If the manager can impose arbitrarily small penalties on the worker he can induce a higher effort level (though it might be that in equilibrium penalties are never imposed).

For an economist, the phenomenon of war is in striking contradiction with the Coase conjecture.[5] At the first glance, it may seem that the weaker party can avoid a conflict by transferring resources to the stronger side.[6,7] However, in the absence of external en-

[3] The fascination of legal scholars with blackmail dates to at least the first half of the 20th century, as Campbell (1939) puts it: "The law of blackmail has something in common with the blackmailer: it allows its student no peace of mind."

[4] Other works on economics of blackmail include Ginsburg and Shechtman (1993) and Posner (1993).

[5] Theories explaining failure of Pareto-efficiency abound (see, e.g., very different perspectives in Coase, 1962, Myerson and Satterthwaite, 1983, Acemoglu, 2003). However, neither transaction costs, nor informational asymmetry can fully explain the phenomenon of war (e.g., Fearon, 1995).

[6] Leng's (1993) typology of influence strategies refers to such a strategy as *appeasing*. In Huth (1988) typology, any policy of concessions is *conciliatory*. Hensel and Diehl (1994) analyze various non-militarized responses on an immediate military threat. Wittman (2003) considers appeasement in a static setting and argues that it should be possible to redraw the map so that peace becomes a self enforcing outcome.

[7] There is a strand of literature that argues that most wars are wanted from the ex-ante perspectives (e.g., Blainley, 1988). Still, this case is much harder to make from the Pareto-efficiency perspective: in rare circumstances a war is wanted on both sides.

forcement, no peaceful agreement could ensure that a strong party would not ask for more concessions in the future. Furthermore, the more transfers the strong party receives, the stronger it becomes. One problem is that the commitment power of parties involved in any destructive conflict is very limited.[8] Indeed, a binding agreement that calls for a lump sum wealth transfer from one nation to the other in exchange for a promise of peace, can make both parties better off. However, this type of agreement is not self-enforcing, except for an unlikely scenario, where the pay-off to the potential aggressor from starting a war decreases substantially as a result of receiving a transfer.[9] Rajan and Zingales (2000) formalize this point and show that the lack of commitment power leads to inefficient outcomes ranging from wars to wasteful political conflicts in organizations (see also, Fearon, 1995, Acemoglu and Robinson, 2005). Acemoglu (2003) emphasizes that ”parties holding political power cannot make commitments to bind their future actions because there is no outside agency with the coercive capacity to enforce such arrangements.” Slantchev (2001) analyzes empirically the interactions of European powers in post napoleonic era. He concludes that "the pattern of cooperative behavior is seen to result from the commitment to uphold the settlement [concluded at the Vienna Congress of 1815], which hinged on the credibility of enforcement threats...".

We show that if agents are risk-averse, then a self enforcing peace agreement may not be viable, since the weaker party foresees that any appeasement may lead to a shift in the balance of power that makes its final destruction inevitable. This is consistent with Fearon (1995) who argues that preventive wars should be understood as a result of commitment problems.[10] In contrast, if parties have constant marginal utility of consumption, which is mathematically equivalent to risk-neutrality, then there exists a continuous stream of transfers that supports a sub-game perfect peaceful equilibrium (Proposition 2). Acemoglu and Robinson (2004, 2005), Shavell and Spier (2002) , Powell (2004) also argue that unavailability of external contract enforcement can be mitigated by spreading transfers over time.[11]

[8] Barzel (2002) considers the absence of commitment to be a major cause for violent conflicts: ”The problem of ... the inability to commit is also the root of theft and some wars”. Classic work that focused on the problem of commitment is North and Weingast (1989) and Shepsle (1991). Most recently, the problem has become the central element in dynamic models of development suggested by Acemoglu and Robinson (e.g., Acemoglu, 2003). Powell (2004) discusses commitment problems in various contexts close to ours.

[9] This is not entirely impossible. According to a legend when Attilla the Hun came to Rome, the citizens of Rome gave him all the gold of Rome in exchange for peace, thus making conquest of the city less attractive for him.

[10] Powell (2004) provides a survey of the literature related to the commitment problem.

[11] In fact, arrangements of this nature were practiced throughout history. A number of medieval cities bought peace by making annual contributions to their stronger neighbors (e.g., Shavell and Spier (2002)

.

In Section 4 we show that the bargaining power of the potential aggressor increases dramatically if she is able to make her threat divisible. For instance, she may achieve so by taking observable action that lead to a war with positive probability. What might come as a surprising result is that ability of the aggressor to make probabilistic threats may dramatically increase her bargaining power. Here probabilistic threat is an observable action that leads to war with small exogenously given probability. (If the weak party believes that the strong party does not start a war, it does not transfer any resources to appease the potential aggressor.) If the observable action that might leads to a war is taken, but the war does not occur, the other party sees the aggressor's commitment to action. In this case, there may exist a subgame perfect equilibrium where the aggressor extracts the whole surplus from the weak party. We can view brinkmanship as an instance of a probabilistic threat. Divisible threats may appear in blackmail practices, where the blackmailer have a possibility to divide the harm he could impose on the victim infinitely.[12]

There are two related strands in the theoretical literature on origins of conflict: the first one is concerned with determinants of wars between countries (see Powell, 1998 and 2002 and references therein). The second is related to the famous Coase theorem (Coase, 1962). Acemoglu (2003) illustrates the failure of bargaining between social groups toward an efficient policy (see also Becker, 1983, and Dixit et al, 2000). Fearon (1995) surveys massive literature on origins of conflicts. Wilkenfeld (1991) shows that the two parties relative military capacity influence the likelihood of weaker party concessions. Garnham (1976) and Weede (1976) demonstrated empirically that two states are less likely to have a war when one of them have a substantial military advantage. Blainley (1988) and Holti (1991) investigate the origins of wars. In particular, they suggest that a war may be caused by expectations of a shift in the balance of power. Undoubtedly, changes in the balance of power are important component in emergence of international conflicts. We show that even if the balance of powers does not have an exogenously given time trend, a self-enforcing peace agreement may not be feasible.

The rest of the paper is organized as follows. Section 2 offers a description of a bargaining game that captures some essential features of wars. In Section 3, we analyze conditions under which a peace treaty is viable. In Section 4 we show that the bargaining power of the potential aggressor increases dramatically if she is able to engage in brinkmanship by taking

invoke Vikings history to illustrate the importance of succession of payments).

[12] If the blackmailer can impose harm only once (or do this a finite number of times), there is no sub-game perfect equilibrium where the victim's transfers are non-zero.

an observable action that leads to a war with positive probability. Section 5 concludes.

2 The Setup

Consider a world consisting of two countries, A and B. Country A is a potential aggressor, whose expected payoff from a war is greater than that of B. Country's B expected payoff from a war is negative. Consequently, B may be willing to pay tribute (bribes) to A in order to preserve peace. At each instant of time $t > 0$ player A (the potential aggressor) chooses whether or not to start a war, $a_t \in \{P, W\}$, while the player B can adjust the rate of transfer to A. [13] The game ends if a war starts. During the pre-war period, the consumption of both parties is determined by the history of transfers. If the war occurs, the post war consumption is determined by the war outcome, while the probability distribution of war outcomes may be in turn influenced by pre-war wealth transfers.[14] Both countries maximize expected discounted utility of consumption.

In a reduced form, a strategy of the weaker country (player B) can be represented by a non-negative differentiable function $b : \mathbf{R}_+ \to \mathbf{R}_+$, $t \longmapsto b_t$. In the event that the game proceeds until time t without a war, b_t represents the rate at which B transfers resources to A at time t. To define an aggressor's strategy, let us summarize the history of the game up to time T by an element from the set H^T of non-negative differentiable function with support on $[0, T[$.[15] Each strategy of player A prescribes her to start or not to start a war for each history of transfers, i.e. a strategy of player A is a function $F : H^T \to \{W, P\}$. The payoff to player i is the discounted utility of the players consumption profile c_{it}, the payoff equals to $V^i = \int_0^\infty \beta^t U_i(c_{it})dt$ where $U_i'(.) > 0$ and $U_i''(.) \leq 0$, $i = A, B$ and β is the discount rate.[16] Instantaneous consumption c_{it} equals to some base rate plus a transfer received. Thus, without loss of generality, we normalize total consumption to zero, so that consumption profile prior to war is given by $c_{At} = b_t$ and $c_{Bt} = -b_t$ for $t \leq T$, where T denotes the time

[13] The same situation might be modeled assuming full symmetry between countries.

[14] Gains/losses from a war tend to be spread out over time. For example, the party that prevails in the conflict may receive a stream of reparation payments or a perpetual stream of benefits from concurred territories. Consequently, we assume that the post war consumption is a constant stream of benefits/losses determined by the outcome of the war.

[15] The assumption that the function representing transfers is differentiable is essentially equivalent to the assumption that the agent chooses the level of transfers at time zero, and chooses the rate of change at each subsequent moment.

[16] For the sake of simplicity of notations, we assume here that both players are described by the same utility function $U(.)$.

when agent A selects action W. If a war never occurs, we can say that $T = \infty$. If a war occurs, the total consumptions is permanently reduced so that $c_{At} + c_{Bt} = -L$ for $t > T$, where L is the loss of resources due to war. The post-war consumption of players are $c_{At} = x$ and $c_{Bt} = -x - L$ for all $t > T$, where x is random variable that represents the outcome of the war. In order to allow for a possibility that transfers shift the balance of power in favor of the recipient we assume that x is distributed according to probability density function $g(x|c)$ where c is the consumption of party A at the time immediately preceding the war, the distribution $g(x|c)$ is atomless, continuous, and $\frac{dE[x|c]}{dc} > 0$.

We leave a more formal description of the game to the appendix. Throughout the paper, our analysis is confined to subgame-perfect Nash equilibria. A *peaceful equilibrium* is any equilibrium in which a war occurs with zero probability.

3 Peace and Transfers

In this section, we use the model developed above to obtain insights into the origins of war and international peace treaties. Proposition 1 establishes necessary and sufficient conditions for maintaining peace without one-sided concessions or wealth transfers. Proposition 3 shows that a self-enforcing peace agreement is viable whenever transfers do not change the balance of powers. In such an equilibrium, all gains from peace are captured by the weaker party, leaving the potential aggressor indifferent between starting and not starting a war. Even if transfers increase the benefit from a war to the potential aggressor, thus increasing his demands for more concessions, a self-enforcing peace agreement between risk neutral parties is feasible (see Proposition 2). However, if parties are risk-averse, a self-enforcing peace agreement may not be feasible.

The first two Propositions of this section replicate and generalize the result of Shavell and Spier (2002) for our setting. Unlike Shavell and Spier (2002), we consider a continuous time model and allow for risk aversion. The results of this sections are important for understanding the main result presented in the next section. The following straightforward proposition makes this claim that peace without transfers is possible only if no party expects to benefit from a war formal.

Proposition 1 *A peaceful equilibrium without transfers exists if and only if the agent A's expected instantaneous utility from starting a conflict does not exceed her instantaneous utility under the status-quo, that is* $\int_{-\infty}^{+\infty} U_A(x)g(x|0)dx \leq U_A(0)$. *If the inequality is strict, any sub-game perfect equilibrium is peaceful.*

Now let us consider a more interesting case: a possibility of a peaceful equilibrium with positive transfers. A possibility of peace, especially when transfers change the balance of powers is not obvious, because in this case concessions will lead to demands for new and bigger concessions (see Rajan and Zingales, 2000, for an example of a model where inefficient outcome obtains for exactly this reason). Hirshleifer (2001) argues that this might be the case to start a war as early as possible. The following proposition shows that risk-neutral parties can always negotiate a self-enforcing peace agreement.

Proposition 2 *Suppose that players have a constant marginal utility of consumption, $U(c) = c$. Then there exists an equilibrium transfer stream such that a war does not occur on the (sub-game perfect) equilibrium path.*

The idea of the proof is to construct a transfer profile that makes party A indifferent between starting and not starting a war at any point in time. Such strategy gives all rents to B, and so we refer to such a profile as a B-profile. By construction the A's best response to a B-profile is to refrain from starting a war. It is also an equilibrium for player B to follow B-transfer profile because B will be strictly worse off if she deviates (because a war will result). Both parties recognize that transfers received by a potential aggressor in the present will allow her to gain more transfers (or wage a more profitable war) in the future. Indeed, consumption stream resulting from starting a war at period t is the amount $\overline{x}_t \equiv E[x|b_t]$. Player A may be willing to postpone the war at time t even if the transfer that he receives at time t and immediately afterwards is substantially less than the post war consumption level would have been. If transfers shift the balance of power, an increase in transfers to A will trigger even greater demands by the potential aggressor tomorrow. In equilibrium future transfers will increase over time to some limit denoted b_∞ where $b_\infty = \lim_{t\to\infty} b_t$. For a given probability distribution of war outcomes $g(x|c)$, the asymptotic value of transfer is given by the non-negative root of the following equation $E[x|c = b_\infty] = b_\infty$.[17]

History provides us with a number of examples illustrating that a stream of transfers could support peace better than lump-sum transfer. For decades since the last days of Chinguiz Khan, the Mongol Empire was the largest state in the history of Earth by size (and the largest of the medieval empires by population). Most of the empire parts were controlled by relatively small forces, which collected modest, but regular tributes from occupied lands. Mongols used a threat of war to extract tributes from the territories under their control. In the political domain, a concession by one party often changes the balance of powers

[17] For instance, if $E[x|c] = 1 + \frac{c}{c+1}$ then solving $1 + \frac{b_\infty}{b_\infty+1} = b_\infty$ we obtain $b_\infty = c = \frac{1}{2} + \frac{1}{2}\sqrt{5}$. Also note that $E[x|c = b_\infty] = b_\infty$ only if b_∞ is finite.

that in turn leads to an escalating sequence of concessions. (The history of the civil-rights movement provide a vivid example of this.) Making a small change in a political institution may shift the balance of power making an institution unstable and leading to a sequence of reforms. Lagunoff (2004) investigates the set of stable political institutions in a static setting. Our model suggests that political institutions may also be in a dynamic equilibrium where stakeholders initiating reform (as well as these opposed to it) understand and correctly anticipate that the change will trigger future reforms.

Now let us turn to the case where transfers do not influence the balance of power.

Proposition 3 *Suppose that A and B are risk-averse, and that past transfers do not affect the probability distribution of war outcomes, i.e. $g(x|c) = g(x)$ for any c. Then there exists a peaceful sub-game perfect equilibrium.*

From Proposition 1, we know that there exists a peaceful equilibrium whenever $E[U_A(X)] \leq U_A(0)$. Let us consider the case when $E[U_A(X)] > U_A(0)$, and hence $E[X] \geq 0$ by the risk-aversion assumption. Consider the following strategy profile: the strategy of player B calls for B-profile transfer schedule, i.e. the schedule that makes A indifferent between starting a war and waiting a little. The strategy of A is to start a war, if B ever deviates from B-profile. These strategy profiles constitute a peaceful equilibrium. By construction, it is the best response for player A to start a war if B ever deviates, and to abstain from a war as long as B follows this strategy. Now let us show that B's strategy is a best response to A's strategy. Indeed, the transfers should be constant as they do not affect the balance of power: $b_\tau = b_\infty$ such that $U_A(b_\infty) = E[U_A(X|b_\infty)]$ for all τ. The choice of player B is between deviating and receiving the consumption stream of $-x - L$ thereafter, or receiving $-b_\infty$. Since B is risk-averse, B's best response is indeed to follow B-profile transfer schedule.

The economic intuition behind the above propositions does not require the assumption that consumption profile is continuous . However, this assumption does have some interesting implications. For instance, in the context of Proposition 3, the following strategy profile constitutes a Nash equilibrium: party B transfers at a rate $b_t = E[X]$ (recall that we assume here that the distribution of X, the outcome of war, does not depend on transfers), and if it ever deviates, it rapidly decreases transfer rate to zero. And party A starts a war if and only if party B deviates. It is easy to see that this strategies form an equilibrium, where no war occurs on the equilibrium path. However, this equilibrium is not sub-game perfect. Indeed, if B deviates and transfers become slightly less then $E[X]$, it is not in the interest of party A to carry out the threat of immediately starting a war because party A is strictly better off not starting a war as far as transfer level stays over b_∞ defined by $U_A(b_\infty) = E[U_A(X)]$.

Indeed, even if party A believes that B is going to reduce transfers to zero very soon, there is no reason not to wait until transfer rate would drop to b_∞. Consequently, continuity implies that out of a large set of Nash equilibria, only the least favorable for A survives subgame perfection. The following simple proposition generalizes this argument.

Proposition 4 *In any sub-game perfect equilibrium, the expected pay-off of A is equal to the expected pay-off of A in the case of war at $t = 0$. The pre-war consumption profile coincides with the B-profile.*

Propositions 2 and 3 show that a carefully negotiated treaty may be a self-enforcing peace agreement. However, for some parameter values, a self-enforcing peace agreement is not feasible. Suppose that both agents' utility functions are strictly concave. In this case when consumption is large, the marginal utility is very small, and when consumption is negative, the marginal utility of consumption is large. If transfer shift the balance of power the weaker party will have to transfer more and more resources in order to appease the potential aggressor. Due to concavity of the utility function the weaker party may find the prospect of making big transfers later more painful than fighting the war earlier. The example demonstrates that a self enforcing peace agreement between risk averse parties may not be viable if transfers shift the balance of power.

Example 1 *Let c_t denote the consumption level of A at time t. Suppose that the outcome of a conflict is deterministic: if a war occurs at time τ, the strong side, A, receives post conflict consumption flow at the rate of $c_\tau + 2$, while post conflict consumption of the weak side, B,is $-c_\tau - 2 - L$ per period, $L > 0$. Let instantaneous utility of both parties be $U(c) = -e^{-c}$ and the total utility of either player is $V_i = \int_0^\infty 0.9^t U(c_i(t))dt$ where $i = A, B$. Then there does not exist a sub-game perfect peaceful equilibrium.*

This logic helps us understand many historic events. Every year since 1933, then-European powers, including England, France, and Poland faced a dilemma: either to fight Hitler now, or appease him and wait (Taylor, 1961). First concession was postponement of the World War I debt payments, next came tolerance of mobilization in Germany and launching of German air forces, and then re-occupation of Ruhrgebiet. The Munich treaty of 1938 year, where the (future) allies agreed with occupation of Sudetenland, a part of Czechoslovakia, is probably the most prominent concession (for a game-theoretic analysis, see Hirshleifer, 2001). On October 5, 1938 Winston Churchill, the leader of the opposition in the British parliament, said that "it seems to me that all those countries of Middle Europe, all those Danubian countries, will, one after another, be drawn into this vast system of power politics - not only power military politics but power economic politics – radiating from Berlin, and I believe this can be achieved quite smoothly and swiftly and will not necessarily entail the firing of a single shot...." In 1940, it became too clear that Germany is already too strong, and England actually entered the World War II out of the fear that Germany would be even stronger in the future.[18] On May 13, Churchill, a prime-minister for five days, said: "What is our aim? I can answer in one word: Victory - victory - at all costs, victory, ... however long and hard the road may be; for without victory, there is no survival."

The same reasoning may explain why some rulers, being presumably risk-averse individuals, often miss opportunities to make concessions to the opposition in the days preceding revolutions. Why would not the Russian czar agree to share power with the elected parliament (where the majority was very far from any leftist movement) in January 1917? Why Louis XVI gambled on counter-revolutionary plots during the first years of French revolution, instead of trying to work out a real political compromise? The answer is that the reluctance of these rulers can be explained by the fear that any concession would have empowered the opposition to ask for even more concessions. Then, they would have a choice between making an escalating sequence of greater and greater concessions or an open struggle under less favorable conditions. Nikolai II of Russia and Louis XVI were the ultimate losers, but there are examples of rulers who ultimately prevailed.[19]

The above example shows that not only the violation of power parity may prompt a war, but the mere expectation of a future power shift may prompt a war. Indeed, a party whose relative power is expected to diminish if it were to pursue an appeasement strategy may prefer to fight a war sooner rather than later (provided that the discount factor is sufficiently high).[20] Fearon (1995) surveys a number of authors that support this conclusion, citing, among many others, Taylor (1954): "Every war between the Great Powers [in the period of 1848-1918] started as a preventive war, not a war of conquest", and Carr (1964): "The most serious wars are fought in order to make one's own country military stronger or, more often, to prevent another country from becoming military stronger". Within this logic, Example 1 formalizes the Hirshleifer (2001) assertion that if the opponent (A, the potential aggressor in our setup) is hostile and non-appeasable, the best possible strategy is to keep the aggressor poor, so that if he opts for a war, he has as little resources as possible.

[18]Formally, Britain announced that she is in war with Germany on September 3, 1939, but the active military stage begun no earlier than Summer 1940.

[19]Louis XIV of France stratefy during the Fronda years might be such an example. John the Landless policy is a textbook example of an appeasement strategy towards the political opposition.

[20]Formally, there are no beliefs in this model. However, it is useful to think of conditional strategies as those of based on expectations.

4 Main Result: Brinkmanship and Blackmail

Unless the potential aggressor expects to benefit from a conflict, extracting any concessions from the weaker party is not subgame-perfect. In light of this, any blackmail is doomed to failure if the blackmailer suffers some costs associated with inflicting the damage on the victim. In this section, we demonstrate that this result hinges on the assumption that the potential aggressor has a small action set: previously, we assumed that the only possible action are peace or ending the game with an all-out war. Here, we show that the bargaining power of the potential aggressor is dramatically increased if action sets includes actions that are in between the two extremes mentioned above.

For instance, we argue that possession of nuclear weapons may not be useful for extracting concessions from weaker countries. A country with air force capable of delivering comparable destruction using conventional weapons may be able to extract concessions from other countries, because, unlike a nuclear bomb, conventional weapons are a "divisible threat". The importance of "divisible threats" is not limited to international negotiations. We argue that the "divisibility of a threat" matters as much in conflicts and negotiations within organizations as it does in international relations. It captures essential features of blackmail in a wide range of situations. For instance, a manager who has the power to fire a worker at any point in time is far less powerful than a manager who has a range of punishments less severe than firing. Of course, in an everyday usage the world "blackmail" is not an appropriate description of a manager threatening a worker with firing unless the worker puts forth good effort. However, from the modeling perspective, it is not different from blackmail.

In the previous sections, we considered an infinite game between a potential aggressor and a victim where the aggressor can exercise the threat and end the game and the victim can make a single transfer or a sequence of transfers to the aggressor in order to appease him. If carrying out the threat is costly for the potential aggressor, the unique subgame perfect equilibrium entails no transfers Ithis statement is made formal in Proposition 1 and is also emphasized by Shavell and Spier, 2002). We show that a threatener is able to extract a stream of positive payments if the threat is probabilistic or infinitely 'divisible.'

Before proceeding to the model, we illustrate the "probabilistic threats" concept with the story of a French diplomat Eon de Beaumont who blackmailed the French king Louis XV.[21] During his service as a diplomat in London, Eon de Beaumont found himself involved in a feud with the French ambassador. When d'Eon was asked by his superior to resign from embassy and return to France, he decided to resort to blackmail. Chevalier d'Eon

[21] For a complete account see Broglie (1879).

had in his possession a copy of the diplomatic correspondence that was so secret that even the French ambassador to England did not know about its existence[22]. Chevalier d'Eon threatened to reveal the diplomatic correspondence to the Englishmen; among other things, the correspondence contained information about Louis XV secret preparation for invasion of the British Islands. d'Eon demanded from the French king money and a permission to remain in England. However, the king was not the only party who stood to loose from revelation of the secret correspondence. Publishing the correspondence could have made d'Eon the most despised traitor on both sides of the channel. At the first glance d'Eon did not have a credible threat; understandably, the crown found d'Eon demands ridiculous and was reluctant to offer concessions.

To persuade the king of the seriousness of his intentions, d'Eon published a pamphlet that included small bits of the text from the secrete correspondence and vague hints about its existence. That prompted France to grant d'Eon's request: he was allowed to stay in England and received the allowance from the embassy. The bargaining power of d'Eon was enhanced by his ability to inflict arbitrary small amounts of harm by leaking small bits of secrete correspondence. The actions of d'Eon were a sort of brinkmanship. Each public hint about the correspondence could have tipped of the English that d'Eon stored the secret documents outside the embassy; in which case, the English might have tried to steal the documents.

To model brinkmanship, we modify the model of the previous section to allow for probabilistic threats. In the previous section, the only threat available to the aggressor was to start an all-out war, that would impose a cost x on him and a cost $-x + L$ on the weaker party, where $L > 0$. We model divisible threats by allowing the aggressor, A, to use probabilistic threats. Specifically, the aggressor's may take an observable action that leads to war with some positive probability p, $p < 1$.[23] We will show that in this case the bargaining power of the aggressor is much greater than in the case where probabilistic threat is not available. Indeed, if the only threat the blackmailer has is to launch a war, he could not do better than extracting the expected gain from the war, which is zero if $x < 0$. In stark contrast to Proposition 1, with probabilistic threats, the strong side can extract the whole surplus from

[22] An institution known as the "king's secret" was a unique feature of the French diplomacy during the reign of Louis XV. The members of the king's secret, of which Chevalier d'Eon was one, received special directives from the king's secret. The rest of the embassy staff including the ambassador were not aware of the existence of this secret correspondence. A detailed account of the working of the king's secret can be found in Broglie (1879).

[23] Alternatively, one might assumes that the blackmailer has a possibility to distribute harm between different periods instead of launching a war at once.

the victim.[24]

There are two players, the blackmailer, A, and the victim, B. For the sake of simplicity, let agents be risk-neutral, and the outcome of war to be deterministic: if a war occurs, the A's instantaneous consumption thereafter is x and the B's is $y = -x - L$. We consider the case where war is an unambigiously bad outcome for both sides that is $x < 0$ and $y < 0$.[25] As above, the time is continuous. In each period $t \in [0, +\infty)$, the victim makes a non-negative transfer b_t to the blackmailer. Prior to the conflict the consumption of parties A and B is b_t and $-b_t$ respectively. At each period party A (the blackmailer) chooses action $a_t \in \{W, P\}$. If $a_t = W$, then the nature moves: with probability p a war occurs and the game ends, and with probability $1-p$ the game continues.[26] An important assumption is that the history of actions is fully observable. The following proposition shows that availability of a probabilistic threat allows the potential aggressor to extract a share of surplus close to one.

Proposition 5 *Suppose that $(1-p)y \geq px$. For any $0 < b < y$ there exists a sub-game perfect equilibrium where along the equilibrium path party B transfers resources to A at rate b and party A never starts a conflict.*

Let us sketch the intuition behind this result. There exists an equilibrium where party A does not attack B as far as it receives transfers at rate b. If party B ever deviates party A takes action W to punish party B. Party A continues to repeat the punishment at frequent intervals as long as party B continues to deviate. If a deviation of party B remains unpunished then party B stops transferring resources to party A. There are no beliefs in the model because the solution concept here is a subgame perfect Nash equilibrium; however, to understand the intuition behind the equilibrium it is useful to talk in terms of beliefs. If any deviation of party B remains unpunished, party B comes to a conclusion that future deviations will not trigger punishments from party A. Thus, party A will punish deviations of B in order to make sure that party B will continue to transfer resources in the future.

Intuitively, the crucial feature of the above blackmail game is that the harm is infinitely-divisible. The condition $-px + (1-p)y \geq 0$ means that the aggressor, A, prefers gambling on war (i.e. choosing $a_t = W$) that yields a stream of losses at rate x with probability p and a stream of transfers at rate b with probability $(1-p)$ over receiving zero transfers from now

[24] The same logic applies to wars: a 'bully' state that has only two options to either start a large war, or not to start a war at all has much less bargaining power than a state that can start a tiny war or a large war or anything in between.

[25] In particular, this assumption means that there is no impact of transfers on balance of power.

[26] We omit the formal description of the space of histories because it is analogous to the previous section.

on. Thus, it is the best response for party A to punish deviations by B and it is the best response for B to transfer resource to A at rate b.

It is worth noting that the blackmail game has multiple equilibria. A war never occurs on the equilibrium path in the equilibrium sketched above. However, there are many economically plausible equilibria where wars occur with positive probability because probabilistic threats or "brinkmanship" occurs on the equilibrium path. For instance, the equilibrium strategy of player A may call for taking action W (engaging in brinkmanship) every five years. If party A deviates from this strategy it looses credibility as a potential aggressor and never receives transfers in the future. There is an equilibrium where along the equilibrium path the weak party occasionally reduces transfer level and the strong party immediately takes action W, with subsequent resumption of transfers at the normal level. Note that action W is not a punishment for a deviation, the victim was supposed to reduce the transfers and the aggressor is supposed to "punish" this equilibrium behavior. The intuition for this result is as following. According to Proposition 4 there exists a subgame perfect equilibrium with transfers $\frac{9}{10}y$ and there is an equilibrium with no transfers. It is easy to construct an equilibrium where transfers are $\frac{1}{2}y$ per period and once a year transfers by B are reduced to zero for one month. At the end of that month A plays action W and B resumes transfers at rate $\frac{1}{2}y$. If aggressor fails to take action W both players switch to playing the equilibrium without transfers in the remaining subgame. If party B fails to reduce transfer to zero at the end of the year both parties play the equilibrium where transfers are $\frac{9}{10}y$ in the remaining subgame.

Extending the range of applications of this model can shed light on economics of terror. For decades, the Irish Republican Army terrorizes the British government and people. A distinct feature of the IRA terror tactics is that acts of violence usually cause relatively small death toll, but, in retrospect, each attack might have caused much more human loss. The British making slow concessions and IRA engaging in occasional terrorist acts is consistent with an equilibrium path of play. Occasional bomb threats may be necessary to make the terrorists' threat credible and thus induce more concessions from the British. Of course, there exists a more efficient equilibrium where concessions are obtained without violence.

5 Conclusion

The bargaining game described in this paper captures some essential features of relationships where neither party has commitment power. A self-enforcing agreement is feasible if parties are either risk neutral or transfers do not change the balance of power. A successful peace agreement is likely to consists of a sequence of concessions. We show that if both parties are risk-averse and if transfers change balance of powers, an inefficient conflict may be inevitable.

The model of brinkmanship is our main contribution. We model brinkmanship as an ability to take an observable action that with large probability have no consequences and with small probability ends the game with payoffs that are very undesirable for both sides. Examples of brinkmanship can be found situations ranging from international relations to Hollywood movies were gangster shoot under the feet to force a victim to cooperate. We show that if the action space of the potential aggressor allows him only two actions such as 'cooperate' or 'go to the ultimate threat point,' his bargaining power is far less than in the case when the potential aggressor can make probabilistic threats.

References

Acemoglu, D. (2003) Why Not a Political Coase Theorem? Social Conflict, Commitment and Politics, *Journal of Comparative Economics,* 31, 620–652.

Acemoglu, D. and Robinson, J. (2004) Economic Backwardness in Political Perspective, NBER Working Paper No. 8831.

Acemoglu, D. and Robinson, J. (2005) *Economic Origins of Dictatorship and Democracy*, forthcoming.

Baliga, S. and Sjostrom, T. (2004) Arms Races and Negotiations, *Review of Economic Studies,* 71:2, 351-369.

Barzel, Y. (2002) *A Theory of the State*, Cambridge Univ. Press.

Becker, G. (1983) A Theory of Competition among Pressure Groups for Political Influence, *Quarterly Journal of Economics*, 98, 371-400.

Besse, J. and Lasswell, H. (1950) Our Columnists on the A-Bomb, *World Politics*, 3:1, 72-87.

Blainey, G. (1988) *The Causes of War,* 2d ed. New York: Free Press.

Broglie, Albert, duc de, (1879) The king's secret: being the secret correspondence of Louis XV with his diplomatic agents, from 1752 to 1774. By the Duc de Broglie. London, New York [etc.] Cassell, Petter & Galpin.

Carr, E. (1964) *The Twenty Years Crisis, 1919-1939,* New York: Harper and Row.

Coase, R. (1960) The Problem of Social Cost, *Journal of Law and Economics,* 1.

Coase, R. (1988) Blackmail, *Virginia Law Review,* 74: 4, 655-676.

Dixit, A., Grossman, G., and Gul, F. (2000) The Dynamics of Political Compromise, *Journal of Political Economy,* 108: 531-568.

Ginsburg, D. and Shechtman, P. (1993) Blackmail: An Economic Analysis of the Law, *U. Penn Law Review,* 141:1849-1876.

Fearon (1995) Rationalist Explanations for War, *International Organization,* 49 (3), 379-414.

Filson, D., and Werner, S. (2002) A Bargaining Model of War and Peace: Anticipating the Onset, Duration, and Outcome of War. American Journal of Political Science, 46, 819–38.

Garnham, D. (1988) Power Parity and Lethal International Violence, 1969-1973, *Journal of Conflict Resolution, 20: 379-394.*

Hensel, P. and Diehl, P. (1994) It Takes Two to Tango: Nonmilitarized Response in Interstate Disputes, *Journal of Conflict Resolution,* Vol. 38, No 3, 479-506.

Hirshleifer, J. (2001) Appeasement: Can it Work? *American Economic Review Papers and Proceedings*, v. 91, n. 2, 342-346.

Holsti, K. (1991) *Peace and War: Armed Conflicts and International Order,* London: Routledge.

Huth, P. (1988) *Extended Deterrence and the Prevention of War*, New Haven, CT: Yale University Press.

Lagunoff, R. (2004) Dynamic Stability and Reform of Political Institutions, Georgetown University Manuscript

Leng, R. (1993) *Interstate Crisis Behavior, 1816-1980: Realism versus Reciprocity,* Cambridge: Cambridge University Press.

North, D. and Weingast, B. (1989) Constitutions and Commitment: The Evolution of Institutional Governing Public Choice in Seventeenth-Century England, *Journal of Economic History.*

Pillar, P. (1983) *Negotiating Peace: War Termination as a Bargaining Process.* Princeton: Princeton University Press.

Posner, R. (1993) Blackmail, Privacy, and Freedom of Contract, *U. Penn Law Review,* 141:1817-1843.

Powell, R. (2002) Bargaining Theory and International Conflict, *Annual Review of Political Science*, 5: 1-30.

Powell, R. (2004) War as a Commitment Problem, mimeo.

Powell, R. (1998) *In the Shadow of Power: States and Strategies in International Politics,* Princeton Univ. Press, Princeton, NJ.

Rajan, R. and Zingales, L (2000) The Tyranny of Inefficient: An Enquiry into the Adverse Consequences of Power Struggles, *Journal of Public Economics*, Vol. 76, no.3, 521-558.

Schelling, T. (1966) *Arms and Influence.* New Haven: Yale University Press.

Shavell, S. (1993) An Economic Analysis of Threats and Their Illegality: Blackmail, Extortion, and Robbery, *U. Penn Law Review,* 141:1877-1903.

Shavell, S. and Spier, K. (2002) Threats Without Binding Commitment, *Topics in Economic Analysis and Policy,* Vol. 2. Issue 1.

Shepsle, K. (1991). Discretion, Institutions and the Problem of Government Commitment in Bourdieu, Pierre and Coleman, James (eds) *Social Theory for a Changing Society,* Boulder: Westview Press.

Slantchev, B. (2001) Territory-Induced Credible Commitments in the European Concert System, 1815-54, mimeo.

Slantchev, B. (2003) The Power to Hurt: Costly Conflict with Completely Informed States, *American Political Science Review,* Vol. 47, No. 1. (March, 2003), pp. 123-135.

Slantchev, B. (2004) How Initiators End Their Wars: The Duration of Warfare and the Terms of Peace, *American Journal of Political Science,* 48:4. 813-829.

Smith, A., and Stam, A. (2003) Bargaining and the Nature ofWar, mimeo.

Taylor, A. (1954) *The Struggle for Mastery in Europe, 1848-1918,* London: Oxford University Press.

Taylor, A. (1961) *The Origins of the World War Two,* New York: Atheneum.

Wagner, H. (1982) Deterrence and Bargaining, *Journal of Conflict Resolution,* Vol. 2, No 6, 329-358.

Weede, E. (1976) Overwhelming Preponderance as a Pacifying Condition among Contiguous Asian Dyads, 1950-1969, *Journal of Conflict Resolution, 20: 395-411.*

Wilkenfeld, J. (1991) Trigger-response Transitions in Foreign Policy size, 1929-1988, *Journal of Conflict Resolution, 35: 143-69.*

Wittman, D. (2003) Credible Commitments and War: a Re-Evaluation, working paper, UCSC Economics Department

Appendix

5.1 The setup

The set of players is $N = \{A, B\}$. The set of histories is denoted H and $\emptyset \in H$. A vector function h, $h_t = (a_t, k_t)$ belongs to the set of non-terminal histories if and only if the support of h is some interval $[0, T[$, and for any $t \in [0, T[$ we have $a_t = P$ and $b_t \geq 0$ where $b_t = k_0 + \int_0^t k_t dt$. There is a natural interpretation of k_t: we assumed that the function representing transfer profile is differentiable and thus continuous. Consequently, although player B can change the transfer level at any instant his choice must be consistent with continuity assumption, essentially, moment by moment player B chooses the derivative of transfer profile, which we denote k_t. (We slightly abuse notation by denoting agent's choice of the level of transfer at time zero by k_0.)

Let Z denote the set of all terminal histories. The set Z consists of two types of terminal histories: infinite (these where a conflict never occurs) and finite (these that end in a war). A vector function $h_t = (a_t, k_t)$ belongs to the set of infinite histories if and only if h has support $[0, +\infty[$ and for all t, $a_t = P$ and $b_t \geq 0$. A vector function $h_t = (a_t, k_t)$ belongs to a set of finite terminal histories if and only if it has support $[0, \tau]$, $b_t \geq 0$ and $a_t = P$ for all $t < \tau$, and $w(\tau) = W$.

The choice correspondence $P(h) = \{A, B\}$ for all $h \in H \backslash Z$. In other words following any non terminal history both players take action simultaneously.

The preferences of players are given by the discounted sum of consumption utilities. The consumption profiles and utilities corresponding to terminal nodes remain as defined in Section 2.

Proofs

Proof of Proposition 1.

If $\int_{-\infty}^{+\infty} U(x)g(x|0)dx \leq U(0)$, then it is an equilibrium for B to send no transfers and for A not to start a war. On the other hand if $\int_{-\infty}^{+\infty} U(x)g(x|0)dx > U(0)$ player's A best response is war if B chooses a strategy where transfers are zero.

Starting a war is a strictly dominated action if and only if $\int_{-\infty}^{+\infty} U(x)g(x|0)dx < U(0)$. Consequently, if the inequality is strict player A will never choose to start a war and player's B unique best response is to send no transfers: $b_t = 0$.■

Definition. A strategy σ of the agent B is a *B-strategy*, if, given that B follows σ, A's pay-off is the same regardless of the strategy A plays.

Proof of Proposition 2.

First, we prove that there exists a B-strategy. Let b_t be a strategy of the player B. Starting a war at the moment τ (from the stand-point of today, which is $t = 0$ for simplicity), brings A the utility of

$$V(b_t, a_\tau = W) = \int_0^\tau U(b_t)e^{-\beta t}dt + \frac{1}{\beta}E[U(X|b_\tau)]e^{-\beta\tau}.$$

The condition that A's expected pay-off from starting a war is the same at any τ means that

$$\frac{d}{d\tau}\left[\int_0^\tau U(b_t)e^{-\beta t}dt + \frac{1}{\beta}E[U(X|b_\tau)]e^{-\beta\tau}\right] = 0$$

for all τ. Re-writing the above equation (and dropping $e^{-\beta t}$), one gets

$$U(b_\tau) - EU(X|b_\tau) + \frac{1}{\beta}E\left[\frac{d}{db}U(X|b_\tau)\right]\frac{d}{d\tau}b_\tau = 0. \tag{1}$$

At $\tau = 0$, we have b_0 defined by

$$b_0 = \frac{1}{\beta}E[U(X|b_0)]. \tag{2}$$

There exists a unique solution to the above differential equation (1) satisfying the initial condition (2).■

Proof of Proposition 3. We need to consider the case of $E[U(X)] > U(0)$. Consider the following strategy profile: suppose the strategy of player B calls for B-profile transfer schedule (see the proof of Proposition 2). To formally define a strategy of player B, assume that if at some decision node the history of the game is inconsistent with the B-profile then player B stops transferring resources to A forever. The strategy of A is to start a war, if B ever deviates from B-profile. Let us show that this strategy profile constitute a peaceful sub-game perfect equilibrium. By construction of B-profile, it is the best response for player A to start a war if B ever deviates and to abstain from a war as long as B follows this strategy. Now let us show that B's strategy is a best response to A's strategy. The differential equation governing the B-profile is

$$U(b_\tau) - E[U(X|b_\tau)] = 0,$$

since $\frac{d}{dm}E[U(X|b_\tau)] = 0$ by assumption. Therefore, $b_\tau = b_\infty$ for all τ, and

$$U(b_\infty) = E[U(X|b_\infty)].$$

Since A is risk-averse, $b_\infty \leq E[X|b_\infty]$. The choice of player B is between deviating and receiving a consumption stream of $-x - L$ thereafter, or receiving $-b_\infty$. In expectation, the former action brings the utility of $\frac{1}{\beta}E[U(-X-L|b_\infty)]$, while the latter brings $\frac{1}{\beta}U(-b_\infty)$. Since $b_\infty \leq E[X|b_\infty]$,

$$E[U(-X-L|b_\infty)] < U(-b_\infty).$$

Thus, B's best response is indeed to follow the B-profile transfer schedule.■

Proof of Proposition 4. By inspection.■

Proof of Example 1. By inspection. Suppose $b(t)$ is a transfer profile that leaves A indifferent between starting and not starting a war, we call this B-profile, among all appeasement strategies this one gives the highest possible payoff to the weaker side. To show that peaceful equilibrium does not exist we first characterize B-profile and then show that even B-profile gives the weaker side lower payoff than immediate war. To make the argument more intuitive we start with comparing a payoff to party A from starting a war at time t versus waiting till time $t+\delta$. We take a limit of $\delta \to 0$ and thus neglect terms of order $o(\delta)$. If party A starts a war at time t its payoff is $\frac{U_A(b(t)+2)}{1-\beta}$ if a war is postponed till time $t+\delta$ the payoff of party A becomes $\delta U_A(b(t)) + \frac{U_A(b(t+\delta)+2)}{1-\beta}$, equating the payoff from starting a war at time t and $t+\delta$ we obtain the following differential equation describes B-profile: $U_A(b(t)+2) - U_A(b(t)) = \frac{U'_A(b(t)+2)b'(t)}{1-\beta}$, Substituting $U_A = -e^{-c}$ and solving the differential equation we find that transfers are linearly increasing in time $b(t) = (exp(2)-1)(1-\beta)t$. It is easy to check that this transfer profile gives utility of negative infinity to player B. Indeed, $V_B = -\int_0^\infty e^{(exp(2)-1)(1-\beta)t}\beta^t\, dt = -\infty$ whenever $(exp(2)-1)(1-\beta)+\ln\beta > 0$. Thus, opting for a war gives player B a higher utility.■

Proof of Proposition 5. To prove the existence, we construct a sub-game perfect equilibrium as follows. First, we describe player A's equilibrium strategy. If the victim, B transfers all resources so that her own instantaneous consumption is equal to $-b$ to the blackmailer, A, the latter makes no harm to B. If B transfers less than b (i.e. B's instantaneous consumption is larger than $-x$), A chooses $a_t = W$. With probability p a war erupts, and so the A's pay-off is $-y$. With probability $1-p$, the A's pay-off is b. Choosing $a_t = P$ brings 0 in the future because the strategy of player B is to stop transfers if a deviation remains unpunished). So, provided that $(1-p)x \geq py$, it is profitable for A to choose $a_t = W$. For A, not triggering a war while B transfers at rate b is clearly the best response. It is straightforward to fully specify strategies of both players off the equilibrium path and to show that the equilibrium is subgame perfect because after an action W is taken players either find themselves at a terminal node or in a subgame that is essentially identical to the original game. ■