Mechanism Design Goes to War: Peaceful Outcomes with Interdependent and Correlated Types*

Mark Fey†
Kristopher W. Ramsay‡

February 16, 2008

Abstract

In this paper, we consider the possibility of identifying peaceful mechanisms such as bargaining protocols, international institutions, or norms that can enable countries to settle disputes in the absence of binding contracts. In particular, we are interested in the existence of mechanisms with zero probability of war. Here, we focus on situations where the countries' payoffs to war are interdependent or correlated and where efficient settlements are not required but subsidies are unavailable. Most importantly, countries can choose to go to war at any time and can use information learned from the negotiation process in making this choice. We characterize the conditions under which no peaceful mechanisms exist and discuss how weakening our war consistency condition can change this result.

Keywords: conflict, war, private information, mechanism design, war consistency

JEL Classification: D74; D82; C7

*We thank Massimo Morelli, Adam Meirowitz, Hein Goemans and two anonymous reviewers for helpful advice and comments and John Duggan and Dan Bernhardt for encouraging this direction of inquiry.

†Department of Political Science, University of Rochester. email: mark.fey@rochester.edu

‡Department of Politics, Princeton University, email: kramsay@princeton.edu

1 Introduction

A central concern in the study of international conflict is to understand the obstacles that prevent countries from reaching mutually beneficial settlements in times of crisis. Why do countries engage in lengthy wars when the two sides would be better off if they could settle their dispute without war? As an answer, the conflict literature has long held that uncertainty is a central cause of war among states (Waltz 1979, Wittman 1979, Blainey 1988, Fearon 1995). One prominent finding of this literature is that informational differences play a key role in determining bargaining and war behavior (Fearon 1995, Schultz 1998, Powell 1999, Wagner 2000, Powell 2004, Smith & Stam 2004, Ponsati 2006). As in the literature on bargaining in economics, the fact that countries have incentives to misrepresent their private information can prevent the two sides from being able to agree on peaceful settlement terms.

Given this theoretical underpinning, it is natural to ask if there are bargaining or mediation procedures that could mitigate the problems of private information in international conflict. In particular, we may want to know if there exist bargaining protocols, which might be incorporated into international institutional procedures, mediation norms, or international law, that could solve the bargaining problem in the shadow of war. More generally, we might be interested in constructing peaceful mechanisms that resolve crises between countries without resort to the destructive process of war.

Scholars have long been interested in the capabilities and constraints of conflict resolution in the international arena. The study of how international institutions and bargaining norms affect the likelihood of conflict is one of the most active research programs in the international relations subfield of political science. While researchers use a variety of methods to explore the effects of international organizations on dispute settlement, one important area of research is built on a game theoretic foundation of strategic choice. Central ques-

tions in this literature are how international institutions and bargaining norms may matter, why institutional designs vary, and how states might construct institutions to pursue their common goals and reduce wasteful conflict (Mitchell 1994, Martin & Simmons 1998, Abbott & Snidal 1998, Abbott & Snidal 2000, Koremenos, Lipson & Snidal 2001, Langlois & Langlois 2004).[1]

Building on this approach, the nature of these questions suggest that mechanism design tools could be quite usefully applied to the study of international conflict. In particular, applying these techniques to the study of bargaining and war could contribute to a better understanding of the constraints that limit the effectiveness of international institutions to prevent war. However, there are several unique features of international conflict that we must incorporate into our analysis and that prevent a direct application of the standard results in the literature. First, it is a central facet of international conflict that bargaining occurs between "sovereign states with no system of law enforceable among them, with each state judging its grievances and ambitions according to the dictates of its own reason or desire" (Waltz 1959). That is, a country can, at any time, choose to go to war and there is no way a country can commit to not do so. As no enforcement is possible, the type of binding contracts that are implicitly or explicitly assumed in the standard mechanism design literature do not exist. A second important element of international conflict is that countries can learn about their opponents through the process of negotiation (Schelling 1960, Pillar 1983, Slantchev 2003). Specifically, we assume that when a country receives a settlement offer, it can update its prior about the private information of the opposing country by inferring what must be true of the other country in order to generate the received offer. In the language of mechanism design, these two facts dictate our choice of a participation constraint, which we call war

[1]Similar literatures have developed in the study of mediation (Wall & Lynn 1993, Kleiboer 1996, Kydd 2003), international law (Slaughter, Tulemello & Wood 1998), and dispute settlement (Smith 2000).

consistency.[2]

A third important aspect of international conflict is that war is always a destructive option and it is therefore common knowledge that there is some Pareto-superior settlement available. This is unlike the standard model of bilateral trade due to Myerson & Satterthwaite (1983) in which trade is sometimes inefficient.[3] Similarly, we are primarily interested in peaceful mechanisms, rather than mechanisms that extract the maximum surplus of agents as in the auction literature.

The final distinction we identify relates to the nature of uncertainty in our models. It is standard in the mechanism design literature to suppose that agents are uncertain about their individual values of the object or public decision under consideration. In the context of international conflict, however, scholars have also examined uncertainty over the costs of fighting (Fearon 1995) and uncertainty over relative strength (Organski & Kugler 1980), for example. How crisis bargaining outcomes depend on these different types of uncertainty is therefore an important question to address.

The application of mechanism design to international conflict has been addressed in Fey & Ramsay (2007*b*). There, we investigate the robustness of existing claims regarding the link between incomplete information and war and show how the existence of peaceful mechanisms depends on the nature of uncertainty and whether types are independent, interdependent, or correlated. One assumption maintained throughout Fey & Ramsay (2007*b*) is that any negotiated settlement between countries must be efficient. That is, the total pie is always divided between the two sides with no subsidy and with no penalty. Clearly, if

[2]Similar ideas have been discussed under the names ex post individual rationality (Matthews & Postlewaite 1989), individual rationality relative to the status quo (Cramton & Palfrey 1995), *a posteriori* individual rationality (Forges 1999), and ex post veto constraints (Compte & Jehiel 2006).

[3]More specifically, in the standard model of bilateral trade, efficient trade is possible if and only if there is common knowledge of gains from trade.

sufficient subsidies were available, both sides could be bought off and peaceful mechanisms would exist.[4] A more interesting question is whether permitting the mechanism designer, or some third party, to appropriate portions of the pie could, in some circumstances, lead to peaceful outcomes that otherwise would not be possible. The intuition for this conjecture is the following. As unilateral war is always possible, a peaceful mechanism must award high settlements to strong types. Naturally, the basic problem with this is that it creates an incentive for a weak type to mimic a strong type. It may then be possible that this problem could be mitigated by shrinking the size of the pie for strong types. This intuition is reinforced by work in mechanism design that shows when types are correlated (Crémer & McLean 1988, McAfee & Reny 1992, Dasgupta & Maskin 2000) or interdependent (Jehiel & Moldovanu 2001, Fieseler, Kittsteiner & Moldovanu 2003, Mezzetti 2004, Mezzetti 2007), efficient mechanisms exist that also extract all of the available surplus.

In this paper, we show that this conjecture is, in fact, false. That is, it does not become easier to produce peaceful outcomes if we relax the efficiency assumption to allow for allocations among the players that sum to less than the total of the pie. In each case that we consider, we show that if there is no peaceful mechanism that is war consistent and ex post efficient, then there is no peaceful mechanism that is war consistent and ex post unsubsidized. In addition, we investigate how a modified form of war consistency might lead to peaceful mechanisms.

We present results for the case of interdependent types and correlated types.[5] The case of interdependent types has also been investigated by Bester & Warneryd (2006) in the standard mechanism design setting. Our work moves beyond this setting to show what happens with

[4]For interesting papers that model third party stakeholders directly and their strategic implications in bargaining see Manzini & Ponsati (2005) and Manzini & Ponsati (2006).

[5]The case of independent values is trivial, as we discuss below.

war consistent and ex post unsubsidized mechanisms, which are more natural in the context of international conflict. For the correlated types case, in recent independent work Compte & Jehiel (2006) establish, as we do, that inefficiencies must always arise. However, in addition to considering interdependent types and relating our work to international conflict, we also show how changing the war consistency condition in the correlated types case changes the possibility of peaceful outcomes.

The rest of this paper proceeds as follows. The next section formally describes the crisis bargaining model underlying the mechanism design question and lays out the conditions we will impose on our mechanism. Section 3 gives results for the cases in which the war payoffs of countries are interdependent or correlated. The results are followed by a short conclusion that discussion of some substantive and theoretical issues in modeling international conflict in this way.

2 The model

We begin with the standard framework in the literature on international conflict that views an international crisis as a bargaining problem in the shadow of war. Suppose there are two states that are involved in a dispute which may lead to war. We conceptualize the conflict as occurring over a divisible item of unit size, such as an area of territory or an allocation of resources. The expected payoff to war depends on the probability that a country will win, the utility of victory and defeat, and the inefficiencies present in fighting. We normalize the utility of countries to be 1 for victory in war and 0 for defeat, and we suppose there is a cost $c_i > 0$ for country i fighting a war. Thus, if p_i is the probability that country i wins the war, the expected payoff for country i of going to war is simply $w_i = p_i - c_i$.

At the outset, each country has private information about their ability to contest a war. That is, each country has private information on attributes of their military regarding their chance of prevailing in a war and/or the costs of conducting a military campaign. For example, a country is likely to possess better knowledge about its relative value for the issue of dispute (captured by the relative cost of fighting c_i) or the strength and capabilities of its military force or both. Formally, country i's private information, $t_i \in T_i$, is referred to as its type. Type pairs $t = (t_1, t_2)$ are distributed according to the joint density $f(t)$ on $T = T_1 \times T_2$.

The countries can attempt to avoid war by resolving their dispute through some bargaining process. In general, this bargaining process can be any extensive form Bayesian game. Whatever this bargaining process is, though, the final outcome is either a peaceful settlement or the outbreak of war. Of course, the *revelation principle* (Myerson 1979) states that any Bayesian equilibrium of any mechanism may be achieved as a truth-telling equilibrium of a direct mechanism that maps players' private information directly into outcomes. Without loss of generality, then, we focus on direct mechanisms. In our case, a direct mechanism is characterized by three functions: the probability of war, $\pi(t)$, and, in the case of a settlement, the value of the settlement to country 1, $v_1(t)$, and the value of the settlement to country 2, $v_2(t)$.

A fundamental feature of the international system is that no enforcement body exists to permit binding contracts. Thus, a country always has the choice of unilaterally going to war. In particular, it is only reasonable to assume that a country can always reject the proposed settlement $v_i(t)$ if it thinks it will be better off by going to war. A second fact that we must account for is that the process of bargaining can reveal, to a greater or lesser extent, the private information of the two sides. Of course, a country should use this additional

information in choosing whether to reject a proposed settlement in favor of going to war. In order to accommodate this, let $\mu_i(v_i, t_i)$ be country i's updated belief about the type of country j after observing the settlement offer v_i. As in Cramton & Palfrey (1995), we assume that this belief is formed via Bayes' Rule, whenever possible. These two aspects of international conflict motivate our version of the standard participation constraint, which we call *war consistency*.

Definition 1 *A mechanism satisfies* ***war consistency*** *if* $v_i(t) \geq E[w_i(t) \mid \mu_i(v_i, t_i)]$ *for all* $t \in T$ *such that* $\pi(t) \neq 1$.

War consistency implies that any negotiated settlement must give each country a payoff at least as large as the payoff that they expect to get from settling the dispute by going to war, given what they have inferred about their opponent as a consequence of the negotiations. That is, any negotiated settlement must be consistent with the fact that either side can start a unilateral war at any time.

Another important feature of international conflict is that war is always inefficient. That is, the war outcome is always Pareto-inferior to some peaceful settlement. As such, we are interested in the prospects of finding settlement procedures that eliminate the occurrence of war. Therefore, we are interested in the class of *peaceful* mechanisms.

Definition 2 *A mechanism is* ***peaceful*** *if* $\pi(t) = 0$ *for all* $t \in T$.

A peaceful mechanism is one in which there is never an impasse that prevents an agreement. Naturally, if we are interested in the conditions under which an international institution or norm can be constructed that leads to the peaceful settlement of disputes, then we are interested in the existence of peaceful mechanisms.

We are also interested in the efficiency of negotiations. In Fey & Ramsay (2007*b*), we considered the class of mechanisms that satisfy ex post efficiency, i.e., $v_1(t) + v_2(t) = 1$ for all $t \in T$. Here we are interested in settlements that may or may not efficiently divide the stakes of the conflict between the disputing parties. This broader class of mechanisms is interesting for several reasons. First, negotiations may be costly, either in terms of delay or the transaction costs associated with reaching an agreement. Second, outside parties may have incentives to see conflicts end peacefully or to take advantage of conflict situations for their own benefit. Finally, allowing solutions that penalize countries in a conflict by imposing transfers to a mediator could potentially make it easier to find peaceful solutions to disputes. Therefore, we are interested in what we call ex post unsubsidized mechanisms.

Definition 3 *A mechanism is ex post unsubsidized if* $v_1(t) + v_2(t) \leq 1$ *for all* $t \in T$.

An ex post unsubsidized mechanism, then, is one that generates total settlements that do not exceed the stakes of the dispute. Thus, we have an environment where a third party can impose punishments on the countries, but not subsidize an agreement, as it is often an easier political task to punish with sanctions than to provide subsidies to maintain the status quo.

3 Results

We are interested in the existence of mechanisms that are war consistent, peaceful, and ex post unsubsidized. As we show in Fey & Ramsay (2007*b*), if the types of players are independent, then it is trivial to construct such a mechanism. In particular, we show that if the costs of war, c_1 and c_2, are independently drawn from some distribution and all other aspects of the game are common knowledge, then the "flat" mechanism with outcomes $v_1 = p$

and $v_2 = 1 - p$ is war consistent, peaceful, and ex post unsubsidized.[6] However, the problem becomes more complicated when types are interdependent or correlated.

3.1 Interdependent Types

We first consider the case of interdependent types. In the context of international conflict, the probability of winning is a natural example of interdependence. That is, if we think of each country as having private information about its relative strength, then it follows that the likelihood of winning a war depends not only a country's own type but also the type of their opponent. In this way, uncertainty about the likelihood of success in warfare is different than uncertainty about the costs of conflict.

We implement this in our framework by supposing that the costs of engaging in a war, c_1 and c_2 are common knowledge, but that the countries have private information regarding p, the probability of winning. Formally, country i's type, $t_i \in [\underline{t}_i, \bar{t}_i] = T_i$, is distributed according to a distribution function F_i and the probability that country 1 wins the war, $p(t_1, t_2)$, is a function of both types. For convenience we make the assumption that this probability is monotonic in the private information of country 1. That is, higher types have a greater chance of winning, all other things being equal. Formally, this assumption is

$$t_1 > t_1' \quad \Longrightarrow \quad p(t_1, t_2) \geq p(t_1', t_2),$$

for all $t_2 \in T_2$. Likewise, we assume that p is monotonically decreasing in t_2. Also, to ensure there is uncertainty, we assume that p is not everywhere constant.

In this setting, it will be useful to restate our war consistency condition. In doing so, we

[6] A flat mechanism is one that does not depend on the types of the agents.

use the fact that the costs of war are common knowledge and the structure of the interdependence of the probability of success in war. Specifically, we say that a mechanism is war consistent if, for all pairs t_1, t_2,

$$v_1(t_1, t_2) \geq \mathrm{E}[p(t_1, t_2) \mid V_1(t_1, v_1)] - c_1 \tag{1}$$

where $V_1(t_1, v) = \{t_2 \mid v_1(t_1, t_2) = v\}$, and

$$v_2(t_1, t_2) \geq \mathrm{E}[1 - p(t_1, t_2) \mid V_2(v_1, t_2)] - c_2. \tag{2}$$

where $V_2(v, t_2) = \{t_1 \mid v_1(t_1, t_2) = v\}$.[7]

For convenience, we use the following notation to simply some of the expressions to follow.

$$P_1(t_1) = \int_{T_2} p(t_1, y) dF_2(y)$$
$$P_2(t_2) = \int_{T_1} p(x, t_2) dF_1(x)$$

In words, $P_1(t_1)$ is the expected probability of winning a war for type t_1 of country 1 and $P_2(t_2)$ is the expected probability of *losing* a war for type t_2 of country 2. A key value in our results is $\bar{c} = P_1(\bar{t}_1) - P_2(\bar{t}_2)$. This value is the difference between the expected probability of winning of the strongest type of country 1 and the strongest type of country 2.

We first give a result from Fey & Ramsay (2007*b*) which shows if the costs of war are

[7] In general, this conditional expectation must be defined abstractly. But this abstract definition simplifies in many cases. For example, if $V_1(t_1, v_1)$ is an interval, then

$$\mathrm{E}[p(t_1, y) \mid V_1(t_1, v_1)] = \frac{\int_{V_1(t_1, v_1)} p(t_1, y) dF_2(y)}{\int_{V_1(t_1, v_1)} dF_2(y)}.$$

sufficiently small, then there is no mechanism that is war consistent, ex post efficient, and peaceful.[8]

Theorem 1 (Fey & Ramsay (2007*b*)) *If $c_1+c_2 < \bar{c}$, then there is no peaceful mechanism that is war consistent and ex post efficient.*

This inability to find peaceful mechanisms persists if we relax the condition of ex post efficient settlement to only require unsubsidized agreements. This is established in the following theorem.

Theorem 2 *If $c_1 + c_2 < \bar{c}$, then there is no peaceful mechanism that is war consistent and ex post unsubsidized.*

The proof of this result is contained in the Appendix. Put another way, Theorem 2 states that if the costs c_1 and c_2 are such that there is no peaceful mechanism that is war consistent and ex post efficient, then there is no peaceful mechanism that is war consistent and ex post unsubsidized.

Thus, unlike the case of independent values, war may be unavoidable with interdependent values. The reason for this stems from the natural assumption that the probability of winning function $p(t_1, t_2)$ is monotonic (and non-constant) in t_1. Because of this, the private information of certain types may lead them to be overly optimistic regarding their likelihood of success in war. For example, if the type realizations for countries 1 and 2 are $t_1 = \bar{t}_1$ and $t_2 = \bar{t}_2$, then *both* countries believe almost surely that they are the stronger of the two countries. In particular, if the cost of war is not too high, then averaging over the possible types of its opponent leads at least one of these very optimistic types to demand

[8]It is easily seen that Theorem 1 is implied by Theorem 2.

more from the settlement than is available—even though there is common knowledge of gains from peace. Central to this result is the fact that any single country can opt out of the mechanism and start a war.[9] This optimism is sufficient to undermine the mechanism when the social cost of war is sufficiently small. Moving from ex post efficiency to ex post unsubsidized mechanisms does not help; it only makes the constraint for peace harder to satisfy.

On the other hand, if $c_1 + c_2 \geq \bar{c}$, then there do exist peaceful mechanisms that are war consistent and unsubsidized. For completeness, we state the following result which is implied by a result of Fey & Ramsay (2007*b*).

Theorem 3 *If $c_1 + c_2 \geq \bar{c}$, then there exists a peaceful mechanism that is war consistent and ex post unsubsidized.*

Proving this result just requires verifying that the flat, peaceful mechanism with outcomes $v_1 = P_1(\bar{t}_1) - c_1$ and $v_2 = 1 - v_1$ is war consistent.

A final corollary of our result in this section is that even if the costs of war are small, war can be avoided if a third party provides a subsidy equal to $P_1(\bar{t}_1) - P_2(\bar{t}_2) - (c_1 + c_2)$. Unfortunately, in some cases this subsidy can be as large as the entire value of the disputed territory or policy.

3.2 Correlated Types

Another natural way to model international conflict is to assume that the value of war to the two countries is correlated. In particular, since victory for one side implies defeat for the other, it is logical to suppose that the war payoffs of the two countries are negatively

[9] In contrast, we show in Fey & Ramsay (2007*a*) that in a model of mutual optimism where war occurs only if both sides agree to fight this result does not hold.

correlated. More generally, if one country's expected payoff to war is high because its probability of success is high, then the opponent's expected payoff from war must be low. As the relative cost of war also varies, there is a whole range of negatively correlated war payoffs consistent with our underlying assumption that war is inefficient.

To formalize this approach, we suppose that each side's type is simply its expected payoff of war, w_i.[10] Thus, the values of w_1 and w_2 are private information and negatively correlated. Specifically, we assume that $w = (w_1, w_2)$ is drawn from a uniform distribution on $W = \{(w_1, w_2) \in [0,1]^2 \mid w_1 + w_2 < 1\}$. This assumption insures that war is inefficient, as $w_1 + w_2 < 1$ always holds, and further insures that war payoffs are negatively correlated.

In Fey & Ramsay (2007*b*), we show that there is no mechanism that is war consistent, ex post efficient, and peaceful. As with the case of interdependent types, here we show that if we relax the efficiency requirement from ex post efficiency to ex post unsubsidized, then the conclusion does not change.

Theorem 4 *If f is uniformly distributed on W, then there is no peaceful mechanism that is war consistent and ex post unsubsidized.*

The proof of this result is given in the Appendix. This result shows that permitting settlements that do not divide the entire pie is not enough to insure peaceful outcomes.

On the other hand, if we relax the war consistency condition from being an ex post condition to an interim condition, then there exists a variety of peaceful mechanisms, some of which allow the designer to extract all the surplus benefit of avoiding war, while others can be biased, giving all of the benefit to one side. Of course, since peaceful mechanisms are possible under an interim condition, they also are possible under weaker versions of war

[10]We use the notation w_i for country i's type instead of t_i in order to emphasize the identification of a country's type as its expected payoff of war.

consistency.

The interim formulation of war consistency is defined as follows.

Definition 4 *A mechanism satisfies* ***interim war consistency*** *if, for* $i = 1, 2$*,*

$$\frac{1}{1-w_i}\int_0^{1-w_i} \pi(w_i, w_j)w_i + (1-\pi(w_i, w_j))v_i(w_i, w_j)\, dw_j \geq w_i$$

for all $w_i \in [0, 1]$ *and* $j \neq i$*.*

Thus, interim war consistency for peaceful mechanisms requires that each country, knowing only its own war payoff, receives an expected settlement that is at least as large as its war payoff. In particular, this interim condition means that a country is willing to subject themselves to some inferior settlements (and possibly war) as long as the average outcome of the mechanism is at least as good as choosing to use force. Substantively, we can interpret this condition to mean that a country can only choose war in the place of settlement negotiations; it cannot choose war after an agreement has been reached. Thus, this condition could be satisfied, for example, in the case of a powerful third party that has the capability of enforcing an agreement.

Our first result is that with interim war consistency, it is possible for a mechanism designer to extract all of the benefit from avoiding war, just as it is possible to extract the full surplus in auction settings (Crémer & McLean 1988, McAfee & Reny 1992, Dasgupta & Maskin 2000).

Theorem 5 *If* f *is uniformly distributed on* W*, then there exists a peaceful mechanism that is interim war consistent and ex post unsubsidized such that* $v_1(w) + v_2(w) = w_1 + w_2$ *for all* $w \in W$*.*

Proof: We give a direct proof by constructing a mechanism with the desired properties. Specifically, consider a peaceful mechanism defined by

$$v_i(w) = \begin{cases} w_i & \text{if } w_1 + w_2 < 1 \\ w_i - 1 & \text{if } w_1 + w_2 \geq 1 \end{cases}$$

for $i = 1, 2$. By construction, this mechanism satisfies interim war consistency and is ex post unsubsidized. To show that it is incentive compatible, we first note that it is clearly never profitable for a country to report a type $\hat{w}_i < w_i$. On the other hand, the expected utility of falsely reporting a type $\hat{w}_1 > w_1$ is

$$\begin{aligned} U_i(\hat{w}_i \mid w_i) &= \int_0^{1-\hat{w}_i} \hat{w}_i \frac{dy}{1-w_i} + \int_{1-\hat{w}_i}^{1-w_i} (\hat{w}_i - 1) \frac{dy}{1-w_i} \\ &= \frac{1-\hat{w}_i}{1-w_i} \hat{w}_i + \frac{\hat{w}_i - w_i}{1-w_i} (\hat{w}_i - 1) \\ &= \frac{1-\hat{w}_i}{1-w_i} w_i. \end{aligned}$$

As this is strictly less than w_i, the payoff from a truthful report, we conclude that the mechanism is incentive compatible. This establishes the result. ∎

This theorem shows that if a third party can credibly "punish" detectable deviations, by imposing cost on the countries (possibly by military or diplomatic means), then this third party can not only insure peaceful outcomes, but can also lay claim to the entire benefit of avoiding war.[11] It is notable that not only is the mechanism given in the proof of this theorem the "minimally destructive" punishment that generates peaceful mechanisms, but

[11]Theorem 5 relies on a punishment condition similar to *punishability* in Baron & Meirowitz (2006) Proposition 4, where they give sufficient conditions for the existence of fully separating perfect Bayesian equilibria of signaling games.

also that such a "minimally destructive" punishment satisfies interim war consistency.

As a corollary of this result, it is also possible under interim war consistency for a third party to use this same scheme to achieve peaceful outcomes that transfer all of the benefits of peace to one side in the conflict. In this case, the mechanism is biased in favor of one side over the other.

Corollary 1 *If f is uniformly distributed on W, then there then there exists a peaceful mechanism that is interim war consistent and ex post unsubsidized such that $v_i(w) = 1 - w_j$ for all $w \in W$ and for some $i \in \{1, 2\}$.*

To see why this result holds, suppose country 2 is to receive the benefits of peace. Define a peaceful mechanism as follows. If $w_1 + w_2 < 1$, then $v_1 = w_1$ and $v_2 = 1 - w_1$, and if $w_1 + w_2 \geq 1$, then $v_i = w_i - 1$ for $i = 1, 2$. Clearly, this mechanism is interim war consistent and ex post unsubsidized. It is incentive compatible for player 2 because her share does not depend on her report. Incentive compatibility for player 1 follows from Theorem 5 and so the corollary follows.

These results speak to the international relations literature on mediation, where there is some debate about the importance of impartial mediators, or "honest brokers," in international conflict (Bercovitch & Houston 2000, Kleiboer 1996, Kydd 2003). Our results suggest two things. First, an important overlooked aspect of mediation is the credibility of a mediator's sanctions when countries take unreasonable settlement positions. Second, while the mediator needs to be able to credibly punish both countries, it need not be impartial. That is, a variety of peaceful mechanisms exist for opportunistic mediators who want to extract the gains from peace for themselves or mediators who have preferences over how the benefits from peace are distributed between the relevant players. This is a very different view of how

the strategic incentives surrounding bargaining in the shadow of war shape the efficacy of mediation in dispute settlement.

Taken together, these results suggest an interesting new way to view the effect of the international system on the prospect for peace between countries. On the one hand, there is an extensive literature in international relations that views a country's war payoff as an unchangeable and ever-present reservation point for negotiations. In many circumstances this makes sense, given the lawless nature of the politics between states. On the other hand, it is certainly possible for a superpower, for example, to intervene in conflicts of other countries and influence their war payoffs. The reservation value for a war in the presence of this intervention may be quite different from the bilateral war that might otherwise occur. If this is true, then war consistency might not be the right participation constraint and a wider variety of mechanisms should be considered feasible.

4 Conclusion

Exploring the prospects of identifying peaceful mechanisms to settle international disputes is an important and interesting application of mechanism design. Moreover, there are aspects of this problem that differ from those that the mechanism design literature has traditionally focused on. Here, we analyze a piece of the puzzle: the existence of peaceful mechanisms when the countries' types are interdependent or correlated and where efficient settlements are not required but subsidies are unavailable.

In each case we examined, we find results that differ from standard mechanism design results that deal with contexts in which legally binding contracts can be signed. This illustrates how crucial it is to correctly incorporate the fact that countries can choose to go to war

at any time. On the other hand, limiting the autonomy of countries, perhaps due to threats by a more powerful country, leads to more encouraging results. These results suggest further work may profitably focus on the potential role of third party mediation and/or intervention in crisis bargaining. In particular, these results suggest it may be fruitful to analyze how the mechanism designer's interests affect the outcomes directly by including the designer as a player in the game.

In many ways, a mechanism design approach to studying incentives in international conflict is particularly appropriate because, unlike modeling elections, with campaigns followed by voting, or challenger entry in market or political competitions, there is no "natural" game form for crisis bargaining. The revelation principle then becomes a powerful tool for making progress in understand the incentives surrounding bargaining in the shadow of war. Indeed, our positive results that explicitly identify peaceful mechanisms should be viewed as outcomes that align the incentives of the countries engaged in crisis bargaining rather than as imposed structures that act to constrain the two sides.

On the other hand, while using the tools of mechanism design allows us to highlight the role that different kinds of uncertainty play in the prospects for conflict resolution, it is important to recognize that incomplete information is not the only rationalist explanation for war. Recently scholars such as Powell (2006) have argued that in some cases war is best understood as the result of a commitment problem. As Powell states, "the crucial issue in commitment problems is that ... states may be unable to commit themselves to follow through on an agreement and may also have incentives to renege on it." In this context, an international institution or mediator will be successful in resolving conflicts insofar as the institution or mediator is able to make credible commitments.

Viewed from this perspective, our findings are still pertinent because the war consistency

condition that we investigate can be interpreted as a commitment problem. As we have emphasized throughout the paper, war consistency captures the idea that a country always has the option of going to war, or, put differently, cannot commit to forgo the use of force when it is in its interest to do so. Thus, our results on the impossibility of peaceful mechanisms under war consistency can be viewed as another example of war as a commitment problem. Indeed, once we allow credible commitments by imposing interim war consistency, a variety of peaceful mechanisms are possible. In this way, our approach provides a framework that encompasses both of the major rationalist explanations of war: incomplete information and commitment.[12]

[12]Our thanks to Massimo Morelli for suggesting this broader perspective.

Appendix

Proof of Theorem 2: Suppose there is some mechanism that is war consistent, unsubsidized, and peaceful. Then, by the revelation principle, there is an incentive compatible direct mechanism yielding the same outcome. Since this outcome is peaceful, $\pi(t) = 0$, for all $t \in T$.

We begin by considering incentive compatibility. Because the mechanism is peaceful, the expected utility of reporting type $\hat{t}_1$ by country 1 with true type t_1 is just $U_1(\hat{t}_1 \mid t_1) = \int_{T_2} v_1(\hat{t}_1, y) dF_2(y)$. The incentive compatibility condition is then

$$U_1(t_1 \mid t_1) \geq U_1(\hat{t}_1 \mid t_1) \qquad \forall t_1, \hat{t}_1 \in T_1.$$

But as $U_1(\hat{t}_1 \mid t_1)$ does not depend on t_1, the only way this condition can be satisfied is if $U_1(\hat{t}_1 \mid t_1)$ is a constant, for all t_1 and $\hat{t}_1$. We write $\bar{U_1}$ for this constant. For later use, we note that this implies

$$\int_{T_1} U_1(t_1 \mid t_1) dF_1 = \int_{T_1} \int_{T_2} v_1(t_1, t_2) dF_2 dF_1 = \bar{U_1}.$$

Of course, a similar result holds for country 2. That is,

$$\int_{T_2} U_2(t_2 \mid t_2) dF_2 = \int_{T_2} \int_{T_1} v_2(t_1, t_2) dF_1 dF_2 = \bar{U_2}.$$

As $v_1(t_1, t_2) + v_2(t_1, t_2) \leq 1$, we conclude that $\bar{U_1} + \bar{U_2} \leq 1$.

Now we turn to the war consistency conditions (1) and (2) evaluated at the type pair $t_1 = \bar{t}_1$ and $t_2 = \bar{t}_2$:

$$v_1(\bar{t}_1, t_2) \geq \mathrm{E}[p(\bar{t}_1, t_2) \mid V_1(\bar{t}_1, v_1)] - c_1,$$

and

$$v_2(t_1, \bar{t}_2) \geq \mathrm{E}[1 - p(t_1, \bar{t}_2) \mid V_2(v_1, \bar{t}_2)] - c_2.$$

Taking expectations of both sides, we get

$$\mathrm{E}[v_1(\bar{t}_1, t_2)] \geq \mathrm{E}[\mathrm{E}[p(\bar{t}_1, t_2) \mid V_1(\bar{t}_1, v_1)]] - c_1,$$

and

$$\mathrm{E}[v_2(t_1, \bar{t}_2)] \geq \mathrm{E}[\mathrm{E}[1 - p(t_1, \bar{t}_2) \mid V_2(v_1, \bar{t}_2)]] - c_2.$$

Using the law of iterated expectations on the right hand side yields the following:

$$\int_{T_2} v_1(\bar{t}_1, t_2) dF_2 \geq \int_{T_2} p(\bar{t}_1, t_2) dF_2 - c_1,$$

and

$$\int_{T_1} v_2(t_1, \bar{t}_2) dF_1 \geq \int_{T_1} [1 - p(t_1, \bar{t}_2)] dF_1 - c_2.$$

The left hand side of the first of these two inequalities is equal to $\bar{U}_1$ and the left hand side of the second is equal to $\bar{U}_2$, so adding the two inequalities yields

$$\bar{U}_1 + \bar{U}_2 \geq \int_{T_2} p(\bar{t}_1, y) dF_2(y) - c_1 + 1 - \int_{T_1} p(x, \bar{t}_2) dF_1(x) - c_2, \tag{3}$$

$$c_1 + c_2 \geq [1 - (\bar{U}_1 + \bar{U}_2)] + P_1(\bar{t}_1) - P_2(\bar{t}_2) \tag{4}$$

$$c_1 + c_2 \geq [1 - (\bar{U}_1 + \bar{U}_2)] + \bar{c}. \tag{5}$$

In order to evaluate the right hand side of the last inequality, observe that monotonicity

of p implies that

$$p(\bar{t}_1, t_2) \geq p(t_1, t_2) \geq p(t_1, \bar{t}_2)$$
$$\int_{T_1} p(\bar{t}_1, t_2) dF_1(t_1) \geq \int_{T_1} p(t_1, t_2) dF_1(t_1) \geq \int_{T_1} p(t_1, \bar{t}_2) dF_1(t_1)$$
$$p(\bar{t}_1, t_2) \geq \int_{T_1} p(t_1, t_2) dF_1(t_1) \geq \int_{T_1} p(t_1, \bar{t}_2) dF_1(t_1)$$
$$\int_{T_2} p(\bar{t}_1, t_2) dF_2(t_2) \geq \int_{T_2} \int_{T_1} p(t_1, t_2) dF_1(t_1) dF_2(t_2) \geq \int_{T_2} \int_{T_1} p(t_1, \bar{t}_2) dF_1(t_1) dF_2(t_2)$$
$$\int_{T_2} p(\bar{t}_1, t_2) dF_2(t_2) \geq \int_{T_2} \int_{T_1} p(t_1, t_2) dF_1(t_1) dF_2(t_2) \geq \int_{T_1} p(t_1, \bar{t}_2) dF_1(t_1)$$

To finish, note that the assumption that p is not constant strengthens the last inequality into a strict inequality. That is,

$$\bar{c} = \int_{T_2} p(\bar{t}_1, y) dF_2(y) - \int_{T_1} p(x, \bar{t}_2) dF_1(x) > 0$$

must hold. This, along with the fact that $1 - (\bar{U}_1 + \bar{U}_2)$ is non-negative, establishes a contradiction between equation (5) and the assumption that $c_1 + c_2 < \bar{c}$. ∎

Proof of Theorem 4: Suppose there is some mechanism that is war consistent, unsubsidized, and peaceful. Then, by the revelation principle, there is an incentive compatible direct mechanism yielding the same outcome.

Because the mechanism is peaceful, we know that $\pi(w_1, w_2) = 0$ on W. We first prove that $\pi(w_1, w_2) = 1$ on $[0,1]^2 \setminus W$. Suppose not. That is, suppose there is a pair (w_1, w_2) such that $w_1 + w_2 > 1$ and $\pi(w_1, w_2) < 1$. In this case, war consistency requires $v_i(w_1, w_2) \geq w_i$ for both states and thus $v_1(w_1, w_2) + v_2(w_1, w_2) \geq w_1 + w_2 > 1$. But this violates unsubsidized. This contradiction proves that $\pi(w_1, w_2) = 1$ on $[0,1]^2 \setminus W$.

As f is uniform, the conditional distribution of w_2, given w_1, is uniform on $[0, w_1]$. Thus, for country 1 with a type w_1, the expected utility of truthfully reporting its type w_1 is

$$U_1(w_1) = \int_0^{1-w_1} v_1(w_1, y) \frac{dy}{1-w_1} = \frac{1}{1-w_1} \int_0^{1-w_1} v_1(w_1, y)\, dy.$$

To save on notation, we define

$$I_1(w_1) = \int_0^{1-w_1} v_1(w_1, y)\, dy = (1-w_1)U_1(w_1)$$

and note that $I_1(w_1)$ is absolutely continuous and therefore differentiable almost everywhere.

On the other hand, the expected utility of falsely reporting a type $\hat{w}_1 > w_1$ is

$$\begin{aligned} U_1(\hat{w}_1 \mid w_1) &= \int_0^{1-\hat{w}_1} v_1(\hat{w}_1, y) \frac{dy}{1-w_1} + \int_{1-\hat{w}_1}^{1-w_1} w_1 \frac{dy}{1-w_1} \\ &= \frac{1}{1-w_1} \frac{1-\hat{w}_1}{1-\hat{w}_1} \int_0^{1-\hat{w}_1} v_1(\hat{w}_1, y)\, dy + \frac{w_1}{1-w_1}(\hat{w}_1 - w_1) \\ &= \frac{1}{1-w_1}[(1-\hat{w}_1)U_1(\hat{w}_1) + w_1(\hat{w}_1 - w_1)] \\ &= \frac{1}{1-w_1}[I_1(\hat{w}_1) + w_1(\hat{w}_1 - w_1)]. \end{aligned}$$

Incentive compatibility requires that $U_1(w_1) \geq U_1(\hat{w}_1 \mid w_1)$, so

$$\begin{aligned} \frac{1}{1-w_1} I_1(w_1) - \frac{1}{1-w_1}[I_1(\hat{w}_1) + w_1(\hat{w}_1 - w_1)] &\geq 0 \\ I_1(w_1) - I_1(\hat{w}_1) - w_1(\hat{w}_1 - w_1)] &\geq 0 \\ I_1(\hat{w}_1) - I_1(w_1) + w_1(\hat{w}_1 - w_1)] &\leq 0 \\ \frac{I_1(\hat{w}_1) - I_1(w_1)}{\hat{w}_1 - w_1} &\leq -w_1. \end{aligned}$$

As I_1 is differentiable almost everywhere, taking limits on the last expression implies $I_1'(x) \leq -x$ holds almost everywhere. From this, it follows that

$$\int_{w_1}^{1} I_1'(x)\,dx \leq \int_{w_1}^{1} (-x)\,dx. \tag{6}$$

As I_1 is absolutely continuous, $\int_{w_1}^{1} I_1'(x)\,dx = I_1(1) - I_1(w_1)$. Noting that $I_1(1) = 0$ and evaluating the integral on the right side of (6) yields

$$\begin{aligned} -I_1(w_1) &\leq \frac{1}{2}(-1 + w_1^2) \\ (1 - w_1)U_1(w_1) &\geq \frac{1}{2}(1 - w_1^2) \\ U_1(w_1) &\geq \frac{1}{2}(1 + w_1). \end{aligned}$$

Of course, because the model is symmetric, an analogous calculation establishes that

$$I_2(w_2) \geq \frac{1}{2}(1 - w_2^2)$$

and that

$$U_2(w_2) \geq \frac{1}{2}(1 + w_2).$$

To finish the proof, we combine the conditions required by incentive compatibility as follows:

$$\begin{aligned}\int_W v_1(w_1,w_2)+v_2(w_1,w_2)\,dw &= \int_0^1\int_0^{1-w_1} v_1(w_1,w_2)\,dw_2\,dw_1 \\ &\qquad + \int_0^1\int_0^{1-w_2} v_2(w_1,w_2)\,dw_1\,dw_2 \\ &= \int_0^1 I_1(w_1)\,dw_1 + \int_0^1 I_2(w_2)\,dw_2 \\ &\geq \int_0^1 \frac{1}{2}(1-w_1^2)\,dw_1 + \int_0^1 \frac{1}{2}(1-w_2^2)\,dw_2 \\ &\geq \int_0^1 (1-x^2)\,dx = 2/3\end{aligned}$$

But as the mechanism is unsubsidized,

$$\int_W v_1(w_1,w_2)+v_2(w_1,w_2)\,dw \leq \int_W (1)\,dw = 1/2.$$

This contradiction establishes our result. ∎

References

Abbott, Kenneth W. & Duncan Snidal. 1998. "Why States Act through Formal International Organizations." *Journal of Conflict Resolution* 42(1):3–32.

Abbott, Kenneth W. & Duncan Snidal. 2000. "Hard and Soft Law in International Governance." *International Organization* 45(3).

Baron, David P. & Adam Meirowitz. 2006. "Fully-Revealing Equilibria of Multiple-Sender Signaling and Screening Games." *Social Choice and Welfare* 26(3):455–470.

Bercovitch, Jacob & Allison Houston. 2000. "Why Do They Do It? An Analysis of the Factors Influencing Mediation Behavior in International Conflicts." *Journal of Conflict Resolution* 44(2):170–202.

Bester, H. & K. Warneryd. 2006. "Conflict and the Social Contract." *Scandinavian Journal of Economics* 108(2):231–249.

Blainey, Geoffrey. 1988. *The Causes of War*. New York, NY: The Free Press.

Compte, Olivier & Philippe Jehiel. 2006. "Veto Constraint in Mechanism Design: Inefficiency with Correlated Types.".

Cramton, Peter C. & Thomas R. Palfrey. 1995. "Ratifiable Mechanisms: Learning from Disagreement." *Games and Economic Behavior* 10(2):255–283.

Crémer, Jacques & Richard P McLean. 1988. "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions." *Econometrica* 56(6):1247–57.

Dasgupta, Partha & Eric Maskin. 2000. "Efficient Auctions." *The Quarterly Journal of Economics* 115(2):341–388.

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

Fey, Mark & Kristopher W. Ramsay. 2007*a*. "Mutual Optimism and War." *American Journal of Political Science* 51(4).

Fey, Mark & Kristopher W. Ramsay. 2007*b*. "Uncertainty and Incentives in Crisis Bargaining Games." *typescript, University of Rochester* .

Fieseler, Karsten, Thomas Kittsteiner & Benny Moldovanu. 2003. "Partnerships, Lemons, and Efficient Trade." *Journal of Economic Theory* 113(2):223–234.

Forges, Francoise. 1999. Ex Post Individually Rational Trading Mechanisms. In *Current Trends in Economics: Theory and Applications*, ed. Charalambos Aliprantis, Nicholas Yannelis & Ahmet Alkan. Springer pp. 157–176.

Jehiel, Philippe & Benny Moldovanu. 2001. "Efficient Design with Interdependent Valuations." *Econometrica* 69(5):1237–1259.

Kleiboer, Marieke. 1996. "Understanding Success and Failure of International Mediation." *Journal of Conflict Resolution* 40(2):360–389.

Koremenos, Barbara, Charles Lipson & Duncan Snidal. 2001. "The Rational Design of International Institutions." *International Organization* 55(4):761–799.

Kydd, Andrew. 2003. "Which Side Are You On? Bias, Credibility, and Mediation." *American Journal of Political Science* 47(4):597–611.

Langlois, Catherine C. & Jean-Pierre P. Langlois. 2004. "Provisions for Noncompliance and Treaty Value: A Game Theoretic Perspective." *International Studies Quarterly* 48(2):383–408.

Manzini, Paola & Clara Ponsati. 2005. "Stakeholders in Bilateral Conflict." *Mathematical Social Sciences* 50:166–180.

Manzini, Paola & Clara Ponsati. 2006. "Stakeholders in Bargaining Games." *International Journal of Game Theory* 34:66–77.

Martin, Lisa L. & Beth Simmons. 1998. "Theories and Empirical Studies of International Institutions." *International Organization* 52(4):729–757.

Matthews, Steven A. & Andrew Postlewaite. 1989. "Pre-play communication in two-person sealed-bid double auctions." *Journal of Economic Theory* 48(1):238–263.

McAfee, R. Preston & Philip J. Reny. 1992. "Correlated Information and Mechanism Design." *Econometrica* 60(2):395–421.

Mezzetti, Claudio. 2004. "Mechanism Design with Interdependent Valuations: Efficiency." *Econometrica* 72(5):1617–1626.

Mezzetti, Claudio. 2007. "Mechanism Design with Interdependent Valuations: Surplus Extraction." *Economic Theory* 31:473–488.

Mitchell, Ronald B. 1994. "Regime Design Matters: Intentional Oil Pollution and Treaty Compliance." *International Organization* 48(3):425–458.

Myerson, Roger B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47(1):61–73.

Myerson, Roger B. & Mark A. Satterthwaite. 1983. "Efficient Mechansims for Bilateral Trading." *Journal of Economic Theory* 29(2):265–281.

Organski, A. F. K. & Jacek Kugler. 1980. *The War Ledger*. Chicago, IL: University of Chicago Press.

Pillar, Paul R. 1983. *Negotiating Peace: War Termination as a Bargaining Process*. Princeton, NJ: Princeton University Press.

Ponsati, Clara. 2006. "The Continuation of Bargaining by Other Means: an elementary theory of war." *Typescript, IAE-CSIC Barcelona* .

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

Powell, Robert. 2004. "Bargaining and Learning While Fighting." *American Journal of Political Science* 48(2):344–361.

Powell, Robert. 2006. "War as a Commitment Problem." *International Organization* 60(01):169–203.

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

Schultz, Kenneth. 1998. "Domestic Opposition and Signaling in International Crisis." *American Political Science Review* 94(4):829–844.

Slantchev, Branislav L. 2003. "The Principle of Convergence in Wartime Negotiation." *American Political Science Review* 47(4):621–632.

Slaughter, Anne-Marie, Andrew S. Tulemello & Stepan Wood. 1998. "International Law and International Relations Theory: A New Generation of Interdisciplinary Scholarship." *American Journal of International Law* 92(3):367–397.

Smith, Alastair & Allan C Stam. 2004. "Bargaining and the Nature of War." *Journal of Conflict Resolution* 48(6):783–813.

Smith, James McCall. 2000. "The Politics of Dispute Settlement Design: Explaining Legalism in Regional Trade Pacts." *International Organization* 54(1):137–180.

Wagner, R. Harrison. 2000. "Bargaining and War." *American Journal of Political Science* 44(3):469–484.

Wall, James A. & Ann Lynn. 1993. "Mediation: A Current Review." *Journal of Conflict Resolution* 37(1):160–194.

Waltz, Kenneth. 1959. *Man, the State and War: A Theoretical Analysis*. New York: Columbia University Press.

Waltz, Kenneth. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.

Wittman, Donald. 1979. "How Wars End." *Journal of Conflict Resolution* 23(4):743–763.