Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict*

Mark Fey†
Kristopher W. Ramsay‡

October, 2007

Abstract

The formal literature on international conflict has identified the combination of uncertainty and the incentive to misrepresent private information as a central cause of war. But there is a fundamental problem with using game-theoretic models to formulate general claims such as these—whether and to what extent a result that holds in a particular choice of game form continues to hold when different modeling choices are made is typically unknown. To address this concern, we present techniques from Bayesian mechanism design that allow us to establish general "game-free" results that must hold in *any* equilibrium of *any* game form in a broad class of crisis bargaining games. We focus on three different varieties of uncertainty that countries can face and establish general results about the relationship between these sources of uncertainty and the possibility of peaceful resolution of conflict. We find that in the most general setting of uncertainty about the value of war, there is no equilibrium of any possible crisis bargaining game form that allows the unilateral use of force that completely avoids the chance of costly war.

*We have benefited from comments received at the University of Rochester's Peter D. Watson Center and Princeton's International Relations Colloquium. We particularly thank Scott Ashworth, Dan Bernhardt, Josh Clinton, Tom Christensen, John Duggan, Jay Lyall, Adam Meirowitz, Andy Moravcsik, Tom Palfrey, Tom Romer, Curt Signorino, and Randy Stone. Any remaining errors are our own.

†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 international relations is understanding the obstacles that prevent countries from reaching mutually beneficial settlements in times of conflict. Why do countries engage in lengthly 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 and the resulting incentive to misrepresent private information together are a central cause of conflict among states (Waltz 1979, Wittman 1979, Blainey 1988, Fearon 1995). One central finding of this theoretical literature is that informational differences regarding aspects of the bargaining process play a key role in determining bargaining and war behavior (Fearon 1995, Schultz 1998, Powell 1999, Wagner 2000, Powell 2004, Slantchev 2003, Smith and Stam 2006). The common theme of this work is that, because of private information, the two sides in a conflict must learn about each other before they can identify suitable settlement terms. But because of incentives to misrepresent their private information, it may be that costly conflict is necessary to credibly signal their bargaining strength.

There is, however, a fundamental problem with using game-theoretic models like those found in the literature to formulate general claims about the role of uncertainty in international conflict. The root of this problem is the fact that the equilibria in any specific game are typically sensitive to the particular details of the game form. That is, it is typically not known how far a result that holds in a specific extensive form generalizes to different extensive forms. For example, recent results by Powell (2004) show that when countries fight and bargain simultaneously, it is optimal to sort the weak from the strong using both the bargaining and fighting processes, resulting in costly war. While insightful, these results—like the results in the sequential bargaining literature in economics—are known to be sensitive to the extensive form assumptions. (See, for example, a recent paper by Leventoglu and Tarar (2006)). Thus, while the requirement of formal models to explicitly include all relevant details, such as the kind and timing of available choices and the preferences and information of decision makers, is often a benefit because it disciplines our thinking, it can also be a hinderance to establishing general claims. Put another way, the fact that war is explained by private information in an particular bargaining game in principle tells us nothing about whether this is true in any of the infinite number of possible variations to the game.

This problem is magnified in the study of international conflict. Unlike the study of elections, say, where candidates must first declare their candidacy, then run their campaigns, which are followed by all voters voting simultaneously on election day, there is no "natural" game form for crisis bargaining. In a crisis, who gets to make proposals? Can a state start a war directly after their initial proposal is rejected? Can a state start a war even if their proposal was accepted? Is bargaining restricted to be bilateral or can mediators be used? Where would such a mediator fit into the process? For questions like these there are no clear answers and this lack of a "natural" game form undermines the applicability of results derived from any particular choice of game form.

In this paper, we take seriously the difficulty of writing down the "correct" model of crisis bargaining and, instead, give results that must hold in *any* equilibrium of *any* game where two states are deciding how to divide a benefit or surplus in the shadow of war.[1] On the face of it, this would seem to be a very difficult endeavor. After all, the set of possible game forms is unimaginably large, and each game form can have many equilibria. In order to overcome this analytical dilemma, we employ a methodological approach from economic theory called Bayesian mechanism design. This approach makes it possible to analyze the outcomes of bargaining games even when the precise procedures used by the parties are unclear.

By focusing on the basic incentive-compatibility constraints on decision-makers' strategies, and the constraints implied by a situation where the use of force is always an option, we find that there are some general facts about crisis bargaining that can be characterized without reference to a game's form. That is, we present results that are "game-free" (Banks 1990) and are not a consequence of particular modeling choices. In this way, our approach is a significant advance on traditional game-theoretic modeling in international relations.[2]

[1]Here we focus on bilateral conflicts, as they are the kind of conflict most often studied in the literature. Understanding bilateral conflict is also important because it is often the case that even wars that end up as multilateral conflicts often have as their source the inability of two particular states to settle their dispute peacefully.

[2]We should note, however, that not every strategic question relating war to the bargaining prob-

In this paper, we are interested in a game-free analysis of crisis bargaining between two countries that have a common interest in avoiding war but that are limited by the uncertainty that they may have over various aspects of the environment. As emphasized by Fearon (1995), just as important as this uncertainty is the resulting incentive for countries to misrepresent their private information and the effect that this incentive has on the prospects of peaceful settlement. In the mechanism design framework we employ, this concern is captured by an "incentive-compatibility" requirement that turns out to play a crucial role in our analysis. Put simply, our results tell us under what conditions private information and the incentive to misrepresent it together pose an *insurmountable* obstacle to peaceful settlement, no matter what bargaining procedure is used. Thus, the game-free approach we present in this paper is particularly well suited to address not only the effect of private information but also the consequences of the incentives to misrepresent this information in bargaining.

Our analysis shows that the link between uncertainty and war in the bargaining problem depends in important ways on the kind of uncertainty states face. For example, we first show that the interpretation of Fearon's (1995) result, linking private information about the cost of war and the positive probability of war, is overstated. In fact, the general link between uncertainty and war is shown to depend on intricate aspects of the nature of decision-makers' private information. That is, if states are uncertain about their opponent's costs of fighting, but there is no uncertainty about the probability of success in war, then there exist bargaining protocols that ensure war never occurs. On the other hand, if states are uncertain about war payoffs in general, then there is no bargaining protocol that eliminates war. In the the final case, uncertainty about relative power (p_i), but certainty about costs, the possibility of designing a bargaining procedure that ensures peace depends on the total (social) cost of war and the possibility of third party subsidies aimed at promoting a settlement.

In the literature on international conflict, the paper that is closest to ours is Banks (1990). This paper uses a game-free approach similar to ours but only considers the case of one-sided incomplete information. It shows that with this information

lem can be answered in a "game-free" fashion. But, given the interpretation of some results in the literature as general claims with universal applicability, we focus on what are, and are not, truly general results.

structure, all equilibria of all bargaining games are monotonic in the sense that the stronger the informed country is, the higher its payoff from peaceful settlement. Our results on correlated payoffs to war are also similar to results found by Compte and Jehiel (2006). There they also find that there is no mechanism that satisfies a veto constraint, is incentive-compatible, and efficient.

In the next section we outline the method we use to generate our results about the relationship between uncertainty and war. Section 3 details our model, defines what is meant by a peaceful mechanism, and explains how our analysis is influenced by anarchy. Section 4 contains our formal results. Section 5 concludes with a discussion of the implications of our findings for the theories of bargaining, war, and institutional design.

2 A Method For Game-Free Analysis

It is perhaps self-evident that in formal models, the results that are obtained depend on the assumptions that are made in formulating the model. While this fact is well understood, what is less recognized by practitioners of formal theory is how sensitive these results can be to the details of the assumed game form. Often the predictions of our models depend crucially on the precise specification of the game we choose (Banks 1990). In the game-theoretic literature on bargaining, for example, a number of variations of the standard alternating-offers model due to Rubinstein (1982) have been studied.[3] Taken together, these variations demonstrate that important features of the equilibrium outcome are often highly sensitive to the exact specification of the bargaining procedure. For example, a seemingly minor change in who makes the first offer can have a significant effect on the bargaining outcome. Other variations, including when disagreement leads to a costly inside option, when players can opt out of bargaining after their offer is rejected but before the counter-offer arrives, or when players cannot commit to not renegotiate their proposal after their opponent accepts, also can have significant effects on equilibrium outcomes.

Consequently, because equilibrium predictions of specific models often change

[3]For a survey, see the textbook by (Muthoo 1999).

drastically with changes in the specification of the interaction, we are forced to ask how broadly the results derived from a particular game form can be applied. Put simply, general results may not follow from specific models. As an example, consider the important paper of Fearon on rationalist explanations for war (Fearon 1995). To establish its general claim that war can be caused by "private information about relative capabilities or resolve and incentives to misrepresent such information," the paper contains an ultimatum offer game with one-sided incomplete information in which war occurs with positive probability. Left unaddressed, though, are questions such as: Can war occur when there is two-sided incomplete information? Can war occur if the bargaining structure is more complex? Although the answer to these questions may very well be "yes," *there is no way to answer these more general questions based on the results of a specific model.* However, in this paper, we show how these questions can, in fact, be answered for all imaginable models. The bottom line is that consumers of current theoretical models are necessarily left unsure about how robust existing findings are and how much our theoretical expectations depend on the analysis of a specific game form.[4] Indeed, (Powell 2004) emphasizes that importance of "the potential sensitivity of informational accounts of war to the bargaining environment—to the sources of uncertainty and the ability to resolve that uncertainty" and calls for "robustness checks for a particular formalization of the bargaining environment."

In order to overcome this analytical dilemma, we employ a methodological approach from economic theory called Bayesian mechanism design. This approach enables us to analyze the outcomes of bargaining games while leaving the precise procedures used by the parties unspecified. In particular, we may ask what possible outcomes could occur *for all possible bargaining procedures that could be used?* This seems, at first glance, to be an intractable question. It is not even apparent how one might categorize all the different kinds of bargaining procedures that could be used.

So how is Bayesian mechanism design able to generate game-free results? The answer is that through the use a powerful result known as the revelation principle, we are able to reduce the scope of our analysis from the class of all possible Bayesian

[4]Obviously, there is a trade-off between the ability to make specific predictions and the generality of results. For a discussion of this trade-off, see Banks (1990).

games to the much smaller class of "incentive-compatible direct mechanisms." In essence, the revelation principle allows us to include the strategic calculations and incentives to misrepresent of the bargainers as part of the direct mechanism. More specifically, the revelation principle states that the outcome of *any* equilibrium of *any* Bayesian game is also the outcome of some equivalent incentive-compatible direct mechanism.

The important implication of this observation is that *any* outcome achievable via *any* equilibrium, under *any* bargaining procedure, must be attainable as the equilibrium outcome of an "information revelation" game in which each player finds it optimal to truthfully reveal his information, given the conjecture that all other players will truthfully reveal their information as well. This is what is referred to as an incentive-compatible direct mechanism. The revelation principle thus implies that if all incentive-compatible direct mechanisms have some property, then every equilibria of every game form has this property. More importantly for our purposes, if *no* incentive-compatible direct mechanism has some property, then *no* equilibrium of *any* game form has this property.[5] In this way, the revelation principle enables us to use direct mechanisms as a powerful tool for analyzing strategic behavior in a wide variety of settings.

3 The Framework for Analysis

In this section we define the strategic problem that countries face in an international dispute and formalize the analytical approach we adopt to characterize the set of possible equilibrium outcomes for *any* game played by countries with certain kinds of private information.

As is customary within much of the conflict literature, consider a situation where two states 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

[5]We will invoke this version of the revelation principle to prove the impossibility of peaceful resolution of disputes, regardless of the game form.

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 $c_i \geq 0$ for country i fighting a war. Thus, if p_i is the probability that country i wins the war, the expected utility for country i of going to war is simply $w_i = p_i - c_i$.

The two countries can attempt to avoid war by resolving their dispute through some peaceful process, which may include direct negotiations, bargaining, threats, mediation by a third party, or some other interaction. Whatever settlement procedure is available in a given instance could then, in principle, be described (abstractly) by a *game form* G which is composed of a set of *actions* for each country, A_1 and A_2, and an *outcome function* $g(a_1, a_2)$ for $a_1 \in A_1$ and $a_2 \in A_2$. It is worth emphasizing that this game form can be anything from a simple strategic form game to an arbitrarily complicated extensive form. We denote a pair of actions (a_1, a_2) by $a \in A = A_1 \times A_2$. Thus, a game form defines the actions available to the countries (e.g., what negotiation tactic to use, etc.) and how those actions interact to determine outcomes. A *crisis bargaining game* is a game form in which the final outcome is either a peaceful settlement or an impasse that leads to war. Thus, we can decompose the outcome function $g(a)$ of a crisis bargaining game into two parts: the probability of war $g^w(a)$, and, in the case of a settlement, the value of the settlement to country 1, $g^v(a)$. We will assume that any potential settlement is efficient and therefore the value of the settlement to country 2 is given by $1 - g^v(a)$. We sometimes write $g_i^v(a)$ for the value of the settlement to country i. With this structure, it is easy to see that the payoff to country i of an action profile a is given by

$$g^w(a) \cdot w_i + (1 - g^w(a))g_i^v(a). \tag{1}$$

In words, the payoff to country i is the probability of a war times the payoff of war plus the probability of a peaceful settlement times the value of a peaceful settlement for an action profile a.

At the outset, each state has private information about their ability to contest a war. That is, each state has private information regarding their chance of prevailing in a war and/or the costs of conducting a military campaign. For example, a state could have unique 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 (reflected in the probability of victory p_i) or both. Formally, we think of each country as having a variety of possible types, where country i's *type* $t_i \in T_i$ represents its private information. The countries have a common prior about the joint distribution of types, given by $f(t)$ for type pair $t = (t_1, t_2) \in T = T_1 \times T_2$. As types can be correlated, it is conceivable that some type pairs cannot occur. To deal with this, we define the set of possible type pairs by T_p, the support of $f(t)$. In general, we will denote country i's war payoff for a type profile t by $w_i(t)$.

We reflect the fact that countries can condition their choice of action on their private information by defining a strategy for country i by a function $s_i : T_i \rightarrow A_i$. The set of all possible strategies for state i is S_i and we let $(s_1, s_2) = s \in S = S_1 \times S_2$. The equilibrium concept we employ is Bayesian-Nash equilibrium. In particular, a strategy profile s^* is a Bayesian-Nash equilibrium if each type of each player is playing a best response to the strategies used by the other players.

In the remainder of this section, we formalize the method of game-free analysis described in section 2 and discuss how to apply this method to crisis bargaining games. We begin by linking the game form and the information structure described above in the following way. Fix an equilibrium s^* to the overall game.[6] For a type pair $t = (t_1, t_2)$, this equilibrium generates an equilibrium probability of war $\pi(t) = g^w(s^*(t))$ and an equilibrium value of settlement to country i, $v_i(t) = g_i^v(s^*(t))$. As peaceful settlements are efficient, the equilibrium value of settlement to country 2 satisfies $v_2(t) = 1 - v_1(t)$. The functions $\pi(t)$ and $v_i(t)$ form what is called an *equivalent direct mechanism*, which can be understood as nothing more than a new game in which each country's action space is just its type space T_i and so the only decision of a country of type t_i is to whether or not to "mimic" some other type t_i'. If it is an equilibrium of the direct mechanism for all types to "tell the truth" and not mimic any other type, then we say that the direct mechanism is *incentive-compatible*.

With these definitions it is now possible to formally state the revelation principle.

Result 1 ((Myerson 1979)) *If s^* is a Bayesian Nash equilibrium of the crisis bargaining game form G, then there exists an incentive-compatible direct mechanism*

[6] For simplicity, we restrict attention to pure strategy equilibria, but our analysis extends to the case of mixed strategy equilibria in a straight-forward manner.

yielding the same outcome.

The intuition for this result is straightforward. Starting from a Bayesian Nash equilibrium in the crisis bargaining game form G, construct a direct mechanism so that truth-telling leads to the same allocation as this equilibrium strategy for every $t \in T$. This is accomplished via an outcome function for the direct mechanism equal to $g(s^*(t))$ for all $t \in T$. Now ask if there any incentives to lie? The answer is, an almost obvious, no. The reason is that if agent i lies, given type t_i, then the lie generates the same outcome that agent i could have gotten by playing some non-equilibrium strategy in the original game form G. Thus, if there is an incentive to lie in the direct mechanism it must be that there was also a profitable deviation for that player in the original game form G. This contradicts the initial supposition that $s^*(t)$ is an equilibrium strategy, proving the result.

By using the revelation principle, we can study equilibria that satisfy incentive-compatibility constraints in direct mechanisms and use our findings to establish general results about properties of equilibria in *all* possible crisis bargaining game forms. Before doing so, we discuss three important properties of crisis bargaining that play a significant role in our later analysis.

First, it is a fundamental feature of international politics that a country can, at any time, choose the use of force over further diplomacy. This threat of force is an unavoidable component of crisis bargaining and therefore must be a component of any game form that purports to model crisis bargaining. Formally, we suppose that a crisis bargaining game form *allows the unilateral use of force* if, for $i \neq j$, there exists an action $\tilde{a}_i \in A_i$ such that $g^w(\tilde{a}_i, a_j) = 1$ for all $a_j \in A_j$. In particular, in an anarchic international system, a country always has the option of rejecting a proposed settlement $g^v(a)$ if it thinks it will be better off by using force.

A second important observation about crisis bargaining is that the process of bargaining has the potential to reveal, to a greater or lesser extent, the private information of the bargainers. Of course, a country should incorporate this additional information into its decision whether to reject a proposed settlement in favor of using force. To capture this formally, we 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 is standard, we assume that this belief is formed via Bayes' Rule, whenever possible.[7]

Combining this updating with the our assumption on the unilateral use of force, it is easy to show the following result is a consequence of the revelation principle.[8]

Result 2 *Suppose that s^* is an equilibrium of a crisis bargaining game form that allows the unilateral use of force. Then there exists an incentive-compatible direct mechanism such that $v_i(t) \geq E[w_i(t) \mid \mu_i(v_i, t_i)]$ for all $t \in T$ such that $\pi(t) \neq 1$.*

The intuition for this result is straightforward. In a game form that allows the unilateral use of force, when faced with a settlement offer v_i, all types of country i have the option of playing $\tilde{a}_i$ and receiving a payoff of $E[w_i(t) \mid \mu_i(v_i, t_i)]$. Therefore if s^* is an equilibrium such that $\pi(t) = g^w(s^*(t)) < 1$, it must be that this deviation is not profitable, which is true if $v_i(t) \geq E[w_i(t) \mid \mu_i(v_i, t_i)]$.

In other words, if the unilateral use of force is always an option, 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 force, 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.

The third and final observation that we incorporate into our analysis is the simple fact that war is costly. Because of this, we are interested in whether private information make war unavoidable or whether there can be cases in which countries always arrive at peaceful settlements. Formally, an equilibrium s^* of a crisis bargaining game form is *always peaceful* if $g^w(s^*(t)) = 0$ for all $t \in T_p$. The revelation principle gives the following result.

Result 3 *Suppose that s^* is an equilibrium of a crisis bargaining game form that is always peaceful. Then there exists an incentive-compatible direct mechanism such that $\pi(t) = 0$ for all $t \in T_p$.*

[7]See Cramton and Palfrey (1995) for a similar treatment of mechanism design when actors can update their beliefs based on their earlier actions.

[8]This result contains our version of what is usually known as the individual rationality constraint or the participation constraint.

In other words, an always peaceful equilibrium of a game form is one in which no possible pair of types of the two countries ever end up abandoning a peaceful resolution of the dispute and resorting to force.

Combining Results 2 and 3 with our original statement of the revelation principle gives our final result.

Result 4 *Suppose that s^* is an always peaceful equilibrium of a crisis bargaining game form that allows the unilateral use of force. Then there exists an incentive-compatible direct mechanism such that $\pi(t) = 0$ and $v_i(t) \geq E[w_i(t) \mid \mu_i(v_i, t_i)]$ for all $t \in T_p$.*

The power of this result is that if we can show that for a given information structure there is no incentive-compatible direct mechanism with these properties, then *no* always peaceful equilibria exist in *any* crisis bargaining game form that allows the unilateral use of force.

4 Analysis

As noted above, the conflict literature has long held that uncertainty is a central cause of conflict among countries. What has been less clearly developed in this literature is what the source of the uncertainty faced by leaders is, how different sources of uncertainty influence the likelihood of conflict or how widely uncertainty and the incentive to misrepresent influence the probability of war. In terms of the source of the uncertainty, some scholars focus on uncertainty about the relative strength of the countries (Blainey 1988, Organski and Kugler 1980) while others concentrate on uncertainty about the costs of conflict or the resolve of countries (Morrow 1985, Kydd 2003, Schultz 1998, Ramsay 2004). In the notation of the previous section, uncertainty about relative strength is uncertainty about p, the probability of victory, and uncertainty about costs is uncertainty about c_i, the cost of conflict. In this section, we address the question of how these different sources of uncertainty influence the likelihood of conflict. We examine uncertainty about the costs of war and uncertainty about relative strength, as well as general uncertainty about the value of war.

4.1 Uncertainty about the Cost of Conflict

We first consider the case in which each country is uncertain about the other's cost of fighting. This is a situation with independent private values. That is, that the realization of one state's cost of war does not directly influence the utility of the the other. Formally, suppose that the probability that country 1 wins a war, p, is common knowledge but there is uncertainty regarding each country's cost for fighting c_i. In this setting, suppose country i's type, $c_i \in [0, \bar{c}_i] = C_i$, is their cost of war, which is distributed according to a cumulative distribution function $F_i(c_i)$, with support C_i.[9] Denote a pair of types $(c_1, c_2) = c \in C = C_1 \times C_2$.

With this information structure, we have the following simple result.

Theorem 1 *If costs c_i are private information, but each country's probability of winning a war is common knowledge, then there exists a crisis bargaining game form that allows the unilateral use of force in which an always peaceful equilibrium exists.*

Proof: To prove this result, it is enough to give an example of a game form that allows the unilateral use of force and that has an always peaceful equilibrium. The following very simple example shows that this is indeed the case. Consider a game form that allows the unilateral use of force such that $g^v(a) = p$ for all $a \in A$ and and $g^w(a^*) = 0$ for some $a^* \in A$. In this case, a strategy profile $s^*(c) = a^*$ is an equilibrium because if either side deviates and starts a war, then both sides are worse off and if either side deviates to a different peaceful action, then the settlement amount does not change. Moreover, this equilibrium is always peaceful by construction. ∎

This theorem shows that in the case where countries are only uncertain about each others costs of war, which is also often interpreted as a country's level of resolve (Schultz 2001, Smith 1998) or preference for fighting, then there exists at least one game form that allows the unilateral use of force and that can eliminate the possibility of war. How could such an institution operate? One possibility would be a direct "arbitration" mechanism. In this institution each country would privately report their costs to a mediator who would then present the two sides with the option of accepting

[9]Here, because the private information of country i is the cost of war, we use the notation c_i rather than t_i for the type of country i.

an agreement, where each side gets a share of the pie equal to their likelihood of success from war, or fighting in hopes of gaining the entire pie. Since this agreement would make both sides better off, regardless of their costs for fighting, it provides a rational and preferable alternative to war. In this sense, our result lends hope to institutional designers that the problem of war is solvable, at least in the case of uncertainty about costs or "resolve."

At some level, though, this is a weak result. It only tells us that a peaceful equilibrium is theoretically possible; it is not a general result about the possibility of war that applies to all possible crisis bargaining games. Put another way, Theorem 1 establishes that Fearon's claim that incomplete information about costs leads to war is not *completely* general, in that there is at least one instance where it does not hold. However, without a truly general result, we do not know if this case is just an isolated exception to Fearon's general claim or if peaceful equilibria are the norm and it is Fearon's example that is the exception.

In light of this, we next present a general characterization of peaceful equilibria in all possible crisis bargaining games that allow the unilateral use of force. For country i, define the expected settlement value from action a_i, given strategy s_j, by

$$Eg_i^v(a_i \mid s_j) = \int_{C_j} g_i^v(a_i, s_j(c_j))\, dF_j(c_j).$$

The following theorem gives a necessary and sufficient condition for the existence of an always peaceful equilibrium.

Theorem 2 *Suppose costs c_i are private information, but each country's probability of winning a war is common knowledge. Let G be any crisis bargaining game form that allows the unilateral use of force. Then an equilibrium s^* of G is always peaceful only if, for $i = 1, 2$, (1) $Eg_i^v(s_i^*(c_i) \mid s_j^*) = p_i$ for all $c_i \in C_i$, (2) $Eg_i^v(a_i \mid s_j^*) \leq p_i$ for all $a_i \in A_i$ and (3) $g_i^v(s^*(c)) \geq p_i - c_i$ for all $c \in C$, where $p_1 = p$ and $p_2 = 1 - p$.*

Proof: Suppose that s^* is an always peaceful equilibrium of a crisis bargaining game form that allows the unilateral use of force. Then by Result 4, there exists an incentive-compatible direct mechanism such that $\pi(c) = 0$ and $v_i(c) \geq p_i - c_i$ for all $c_i \in C_i$ and $i = 1, 2$. This direct mechanism is given by $\pi(c) = g^w(s^*(c))$ and $v_i(c) = g_i^v(s^*(c))$

For country i, incentive-compatibility requires that for every $c_i \in C_i$,

$$\int_{C_j} v_i(c_i, c_j)\, dF_j(c_j) \geq \int_{C_j} v_i(\hat{c}_i, c_j)\, dF_j(c_j), \qquad \text{for all } \hat{c}_i \neq c_i.$$

Clearly, this is only possible if $\int_{C_j} v_i(c_i, c_j) dF_j(c_j)$ is constant in c_i. In addition, because the constraint $v_i(c) \geq p_i - c_i$ must hold for $c_i = 0$, we have

$$\int_{C_j} v_i(c_i, c_j)\, dF_j(c_j) \geq p_i \qquad \text{for all } c_i \in C_i,\ i = 1, 2.$$

To show that this expression must hold with equality, suppose not. That is, suppose for some country i, $v_i > p_i$. Then

$$\begin{aligned} \int_{C_1}\int_{C_2} [v_1(c) + v_2(c)]\, dF_2(c_2)\, dF_1(c_1) &= \int_{C_2}\int_{C_1} v_1(c)\, dF_1(c_1)\, dF_2(c_2) \\ &\quad + \int_{C_1}\int_{C_2} v_2(c)\, dF_2(c_2)\, dF_1(c_1) \\ &> \int_{C_2} p\, dF_2(c_2) + \int_{C_1} (1-p)\, dF_1(c_1) \\ &> p + (1-p) = 1. \end{aligned}$$

But as peaceful settlements are efficient, $v_1(c) + v_2(c) = 1$ for all $c \in C$ and therefore we have a contradiction. This proves the theorem. ∎

The necessary condition given in the theorem for a peaceful equilibrium states that if s^* is an always peaceful equilibrium in a crisis bargaining game form that allows the unilateral use of force, then every type of both players must receive an expected settlement value equal to its probability of victory in a war. Thus, a peaceful equilibrium must, in expectation, be completely insensitive to the private information of the countries—high cost and low cost countries must receive the same expected settlement value. The most natural example of a strategy profile that would generate such insensitivity is a (completely) pooling strategy, in which all types of a country choose the same action. Such a characterization of types of equilibria to crisis bargaining games would be very useful, if completely pooling equilibria were the only type of equilibrium that had the property that, in expectation, the payoff to war was flat in the countries' reports. However, it can be shown that there also exist non-pooling strategies that satisfy the conditions given in the theorem.[10]

Even though our claim about the existence of peaceful mechanisms that solve the crisis bargaining problem extends beyond the set of games with only pooling equilibria, where in expectation each type gets expected settlement p, it does seem intuitively clear that many game forms will fail this necessary condition. In fact, the existence of an intuitive type of equilibrium can guarantee that a particular game form has equilibria with positive probability of war under the assumption of private information about the cost of fighting. If we define a *monotonic equilibrium* to be an equilibrium to a game where "stronger" (here having low costs) types get better outcomes from the negotiations, in expectation, then there must be positive probability of war in this equilibrium, and hence this game form.

4.2 Uncertainty about Relative Strength

While the previous section dealt with the case of uncertainty about the costs of conflict, in this section we deal with a second source of uncertainty that has received significant attention in the literature—uncertainty about the distribution of power and the likelihood of success in war. In this case, countries are assumed to be informed about their opponent's relative cost of fighting, but are uncertain about the likely outcome of conflict. In particular, countries have private information about their military's quality or their combat strategy that leads each side to hold private beliefs about what will happen as a result of fighting a war. This is a situation with interdependent values and uncorrelated types. That is, each country's utility for conflict is not only dependent on their own type, but it also depends on the type of their opponent. However, information is uncorrelated in that the realization of one country's type does not affect the likelihood of the other country's types.

We would note here that optimism—in the sense of Blainey (1988) and Wittman (1979)— can play an important role in this case, where countries' utilities for war

[10]Details of this example are available from the authors.

are interdependent. However, the situations coved by these results on interdependent types do not directly speak to the argument that war may result from mutual optimism. Central to the mutual optimism argument is the claim that *both* countries must want to fight for a war to occur, hence the name *mutual* optimism.[11] In the environment described in this section, however, any single country can choose to start a war rather than accept a settlement. Therefore, this setting can be interpreted as the compliment to the mutual optimism argument, i.e., the case where war may be the result of *unilateral* optimism. Below we will see that when any single country can start a war, optimism and the ability to unilaterally start a war can substantially limit the ability of countries to reach a peaceful agreement.

In this section we suppose the costs of engaging in a war, c_1 and c_2 are common knowledge, but the countries have private information regarding the probability of winning. We implement this in our framework by supposing that 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 prevails in a war, $p(t_1, t_2)$, is a function of both types. We assume that the types of players can be ordered such that the probability of victory is monotone in the countries' types. That is, higher types have a greater chance of winning, all other things being equal. Formally this assumption is that $t_1 > t_1'$ implies $p(t_1, t_2) \geq p(t_1', t_2)$, 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 way, the type t_i reflects the "strength" of country i and thus the probability of victory depends on the relative strength of the two combatants.

An important consequence of this focus on relative strength is that the process of bargaining can reveal important clues as to the likely strength of the two countries. In particular, when a country receives a settlement offer, it can update its prior about the private information of the opposing state by inferring what must be true of the other state in order to generate the received offer. Recall that $\mu_i(v_i, t_i)$ is country i's updated belief about the type of country j after observing the settlement offer v_i. Let $V_1(t_1, v) = \{t_2 \mid v_1(t_1, t_2) = v\}$; this is the set of possible types of country 2 that a given type t_1 of country 1 would think are possible after observing a settlement v.[12]

[11]For more on the issue of mutual optimism and assumptions about how countries end up at war, see Fey and Ramsay (2007).

[12]In general, this conditional expectation must be defined abstractly. But this abstract definition

In this setting, then,

$$E[w_1(t) \mid \mu_1(v_1, t_1)] = E[p(t_1, t_2) \mid V_1(t_1, v_1)] - c_1, \tag{2}$$

and a similar expression holds for country 2. The right hand side of this inequality is simply the updated expected utility for war incorporating the inference about the types of the other country from the observed settlement offer.

For convenience, we use the following notation in stating our results:

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

In words, $P_1(t_1)$ is the expected probability of winning a war for type t_1 of state 1 and $P_2(t_2)$ is the expected probability of *losing* a war for type t_2 of state 2.

Let $\bar{c} = P_1(\bar{t}_1) - P_2(\bar{t}_2)$. It follows from the monotonicity of p that $\bar{c} > 0$. Our first result shows that if the costs of war are less than $\bar{c}$, then no matter what bargaining procedure is used, there is a positive probability of war during a crisis.

Theorem 3 *Suppose costs are common knowledge but each country is uncertain about the probability of winning a war. If $c_1 + c_2 < \bar{c}$, then no always peaceful equilibrium exists in any crisis bargaining game form that allows the unilateral use of force.*

Proof: The method of proof is by contradiction. We begin by supposing there is an always peaceful equilibrium of a crisis bargaining game form that allows the unilateral use of force. By Result 4, there exists an incentive-compatible direct mechanism such that $\pi(t) = 0$ and $v_i(t) \geq E[w_i(t) \mid \mu_i(v_i, t_i)]$ for all t and $i = 1, 2$. Because $\pi(t) = 0$, the expected utility of country 1 with true type t_1 reporting type $\hat{t}_1$ is

$$U_1(\hat{t}_1 \mid t_1) = \int_{T_2} v_1(\hat{t}_1, y)\, dF_2(y). \tag{3}$$

---

simplifies in many cases. For example, if $V_1(t_1, v_1)$ is an interval, then

$$E[p(t_1, y) \mid V_1(t_1, v_1)] = \frac{\int_{V_1(t_1,v_1)} p(t_1, y)\, dF_2(y)}{\int_{V_1(t_1,v_1)} dF_2(y)} - c_1.$$

The incentive-compatibility condition is then

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

However, from equation (3), it is clear that $U_1(\hat{t}_1 \mid t_1)$ does not depend on t_1. Therefore the only way the incentive-compatibility condition can be satisfied for all t_1 and $\hat{t}_1$ 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.

Turning now to the condition that $v_i(t) \geq E[w_i(t) \mid \mu_i(v_i, t_i)]$ for all t and $i = 1, 2$, we use equation (2) evaluated at the type pair $t_1 = \bar{t}_1$ and $t_2 = \bar{t}_2$ to get

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

Taking expectations of both sides, we get

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

By the law of iterated expectations, this expression is equivalent to

$$\bar{U_1} = \int_{T_2} v_1(\bar{t}_1, t_2)\, dF_2(t_2) \geq \int_{T_2} p(\bar{t}_1, t_2)\, dF_2(t_2) - c_1. \tag{4}$$

By a similar argument, we can establish that

$$\bar{U_2} = \int_{T_1} v_2(t_1, \bar{t}_2)\, dF_1(t_1) \geq \int_{T_1} [1 - p(t_1, \bar{t}_2)]\, dF_1(t_1) - c_2. \tag{5}$$

We next show that $\bar{U_1} + \bar{U_2} = 1$. Starting with the fact that $v_1(t_1, t_2) + v_2(t_1, t_2) = 1$ for all pairs (t_1, t_2), it follows that

$$\begin{gathered}
\int_{T_1} \int_{T_2} [v_1(t) + v_2(t)]\, dF_2(t_2)\, dF_1(t_1) = 1 \\
\int_{T_1} \int_{T_2} v_1(t)\, dF_2(t_2)\, dF_1(t_1) + \int_{T_2} \int_{T_1} v_2(t)\, dF_1(t_1)\, dF_2(t_2) = 1 \\
\int_{T_1} \bar{U_1}\, dF_1(t_1) + \int_{T_2} \bar{U_2}\, dF_2(t_2) = 1 \\
\bar{U_1} + \bar{U_2} = 1.
\end{gathered}$$

Therefore, adding inequalities (4) and (5) yields

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

from which it follows that

$$c_1 + c_2 \geq P_1(\bar{t}_1) - P_2(\bar{t}_2) = \bar{c}.$$

This contradicts the supposition that $c_1 + c_2 < \bar{c}$ and thus proves the theorem. ∎

This theorem shows that there is always a range of costs such that the countries' private information about the probability of winning is always an obstacle to peace, no matter what the game form is of the interaction between the two countries. In this way, it is a sufficient condition for the impossibility of completely peaceful settlements. The next result shows that this condition is also necessary. In other words, if the total costs of war are greater than the critical value $\bar{c}$, it is possible to always avoid war.

Theorem 4 *Suppose costs are common knowledge but each country is uncertain about the probability of winning a war. If $c_1 + c_2 \geq \bar{c}$, then there exists a crisis bargaining game form that allows the unilateral use of force in which an always peaceful equilibrium exists.*

Proof: As with Theorem 1, it is enough to give an example of a crisis bargaining game form that allows the unilateral use of force which has an always peaceful equilibrium. Consider a crisis bargaining game form that allows the unilateral use of force such that $g^v(a) = P_1(\bar{t}_1) - c_1$ for all $a \in A$ and and $g^w(a^*) = 0$ for some $a^* \in A$. By construction, the strategy profile $s^*(t) = a^*$ is always peaceful. To show that it is an equilibrium, first note that if either side deviates to a different peaceful action, then the settlement amount does not change. If country 1 unilaterally starts a war, its expected payoff from war, given its type t_1, is

$$\int_{T_2} p(t_1, y)dF_2(y) - c_1 \leq \int_{T_2} p(\bar{t}_1, y)dF_2(y) - c_1 = P_1(\bar{t}_1) - c_1 = g_v(a^*),$$

where the first inequality follows from the monotonicity of p. Therefore, no type of country 1 will deviate. A similar argument shows that country 2 will not deviate and start a war. Therefore $s^*(t)$ is an equilibrium and the proof is complete. ∎

These two results give a condition for when a peaceful equilibrium is possible and when it is not. It is important to note that this condition is based solely on the information structure of the countries' uncertainty and their costs of war and is completely independent of the form of the bargaining process. In other words, no possible bargaining norm, protocol, or institution can eliminate all chance of war when the social cost of war is small and the relevant strategic uncertainty concerns the distribution of power. This is because, as demonstrated in Theorem 3, if $c_1 + c_2 < \bar{c}$, then the settlement amount is not large enough to satisfy the "strongest" type of both countries simultaneously. This, coupled with the incentive of weaker types to misrepresent and mimic this strongest type, precludes the possibility of always peaceful outcomes, no matter what bargaining process is used.

Another important feature of Theorem 4 is that it implies a trivial sufficient condition for the existence of peaceful game form is that $c_1 + c_2 \geq 1$. Yet, much of the time this condition is unlikely to hold, i.e., it is rarely the case that a real world conflict generates relative costs that are greater than the total value of the object of dispute. However, if there is a third party that is willing to provide a sufficiently large subsidy to the peace process, such as an international organization or a superpower, a peaceful settlement is possible. Permitting this possibility, we can establish the following corollary.

Corollary 1 *If a third party provides a subsidy*

$$\phi \geq P_1(\bar{t}_1) - P_2(\bar{t}_2) - (c_1 + c_2),$$

then there exists a crisis bargaining game form that allows the unilateral use of force in which an always peaceful equilibrium exists.

In this result, the subsidy amount is the minimum amount that will insure that there is an agreement that both sides will prefer to unilaterally starting a war. Thus, in a world with a large powerful country willing to provide sufficient subsidies, the occurrence of war as a consequence of private information about relative power can be avoided.[13]

4.3 General Uncertainty about the Value of War

Finally, we arrive at the case where there is general uncertainty about the two countries' values for war. In particular, we assume that w_i, the value of fighting for country i, is private information. Of course, it is commonly known by both sides that war is inefficient, so it must be that $w_1 + w_2 < 1$. It is then immediate that the countries' private information must be correlated. Although this formulation of the uncertainty facing countries has not been analyzed previously in the formal literature on international conflict, it may closely resemble what exists in the real world. That is, while countries may "know" that war is inefficient, they many not know much more than their own value for fighting. This is a situation with correlated values. That is, if one country has a high utility for war, the other country most likely has a low utility.

We implement this general uncertainty into our framework by defining the set of possible war values by

$$W = \{(w_1, w_2) \in [0, 1]^2 \mid w_1 + w_2 < 1\}.$$

This set is therefore the set of possible type combinations of the two countries, i.e. $T_p = W$.[14] In what follows we suppose the common prior f is uniform on W.

Our first result in this section shows that with general uncertainty about the value of war, there is always a positive probability of war, no matter what the details of the bargaining procedure are.

Theorem 5 *If f is uniformly distributed on W, then no always peaceful equilibrium exists in any crisis bargaining game form that allows the unilateral use of force.*

Proof: The method of proof is by contradiction. We begin by supposing there is an

[13]This corollary supports the argument that if there is a global hegemon, then the international system is likely to be more peaceful, given that the hegemon is willing to pay the cost (Kindleberger 1973, Gilpin 1981, Keohane 1984).

[14]Here, because the private information of country i is the value of war, we use the notation w_i rather than t_i for the type of country i.

always peaceful equilibrium of a crisis bargaining game form that allows the unilateral use of force. By Result 4, there exists an incentive-compatible direct mechanism such that $\pi(w) = 0$ and $v_i(w) \geq w_i$ for all $w \in W$ and $i = 1, 2$.

Because $\pi(w) = 0$, the expected utility of type w_1 of country 1 is given by

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

and a similar definition applies to country 2. Applying Lemma 1 in the Appendix with $\pi(w) = 0$, we have $\mathrm{E}\, U_1(w_1) = (1/2)(1 + w_1)$ and $\mathrm{E}\, U_2(w_2) = (1/2)(1 + w_2)$.

We finish the argument by recalling that $v_1(w) + v_2(w) = 1$ for all $w \in W$ and performing the following calculations:

$$\begin{aligned}
\int_W [v_1(w_1, w_2) + v_2(w_1, w_2)]\, dw &= \\
&\int_0^1 \int_0^{1-w_1} v_1(w_1, w_2)\, dw_2\, dw_1 + \int_0^1 \int_0^{1-w_2} v_2(w_1, w_2)\, dw_1\, dw_2 \\
\int_W [1]\, dw &= \int_0^1 (1 - w_1)\, \mathrm{E}\, U_1(w_1)\, dw_1 + \int_0^1 (1 - w_2)\, \mathrm{E}\, U_2(w_2)\, dw_2 \\
\frac{1}{2} &\geq \int_0^1 (1 - w_1)(1/2)(1 + w_1)\, dw_1 + \int_0^1 (1 - w_2)(1/2)(1 + w_2)\, dw_2 \\
\frac{1}{2} &\geq \int_0^1 (1 - x^2)\, dx \\
\frac{1}{2} &\geq \frac{2}{3}.
\end{aligned}$$

This contradiction establishes our result. ∎

The content of this theorem can be phrased as follows. Consider *any* bargaining procedure that may be used to moderate disputes between two countries. This includes, for example, any direct bargaining process in which the countries can make offers and counteroffers to each other, as well as any arbitration mechanism in which the parties communicate to a formal institution and this institution decides how the dispute will be settled. Then the theorem tells us that, under its assumptions, no such institution can have an that leads to a peaceful resolution of the dispute for all realizations of the countries' valuations of war. Here we have a robust result linking uncertainty and war. When war payoffs are correlated and each country can resort to the unilateral use of force, then there are no bargaining protocols, norms, or institutions that can eliminate inefficient war.

Given this general result linking uncertainty and war, it is natural to ask how much peace is possible. That is, of *all* equilibria of *all possible* settlement procedures, which is best in the sense of maximizing social welfare and what institution or bargaining process generates this outcome? In the next result we characterize all such a "second best" procedures.

Theorem 6 *If f is uniformly distributed on W, then any equilibrium outcome that maximizes social welfare among all crisis bargaining game forms that allow the unilateral use of force must satisfy*

$$\pi(w_1, w_2) = \begin{cases} 0 & \textit{if} \quad \min(w_1, w_2) < k[1 - \max(w_1, w_2)] \\ 1 & \textit{if} \quad \min(w_1, w_2) > k[1 - \max(w_1, w_2)], \end{cases}$$

where $k = (1/12)(\sqrt{97} - 5)$.

Proof: Consider some equilibrium of some crisis bargaining game form that allows the unilateral use of force. By Result 2, there exists an incentive-compatible direct mechanism $\Gamma = (\pi, v_1, v_2)$ yielding the same outcome that satisfies $v_i(w) \geq w_i$ for all $w \in W$ such that $\pi(w) \neq 1$ and $i = 1, 2$. We are interested in establishing which choice of Γ maximizes average social welfare, which is given by

$$\int_W (1)(1 - \pi(w_1, w_2)) + (w_1 + w_2)\pi(w_1, w_2)\, df.$$

Let Γ^* denote this optimal mechanism. To begin, it is easy to see that Γ^* must have $\pi^*(w_1, w_2) \in \{0, 1\}$.[15] Therefore, we can define a function $h(x)$ such that

$$\pi^*(w_1, w_2) = \begin{cases} 1 & \text{if } h(w_1) \leq w_2 \\ 0 & \text{if } h(w_1) > w_2. \end{cases}$$

[15]This is an application of the bang-bang principle of optimal control.

By the symmetry of the problem, we can assume that h is decreasing and symmetric about the 45° line. We can use this to write the social welfare of Γ^* as

$$\begin{aligned}
SW^* &= \int_0^1 \int_0^{1-w_1} \Big[(1-\pi^*(w_1,w_2)) + (w_1+w_2)\pi^*(w_1,w_2)\Big] \frac{dw_2\,dw_1}{1/2} \\
&= 2\int_0^1 \left[\int_0^{h(w_1)} (1)\,dw_2 + \int_{h(w_1)}^{1-w_1} (w_1+w_2)\,dw_2\right] dw_1 \\
&= 2\int_0^1 \left[h(w_1) + w_1(1-w_1-h(w_1)) + \frac{1}{2}((1-w_1)^2 - (h(w_1))^2)\right] dw_1.
\end{aligned}$$

As f is uniformly distributed on W, it follows from Lemma 1 that for Γ^*, the following expressions hold:

$$\mathrm{E}\,U_1(w_1) = \frac{1}{2}(1+w_1) - \frac{1}{1-w_1}\int_{w_1}^1 \int_0^{1-x} \pi^*(x,t)\,dt\,dx \tag{6}$$

and

$$\mathrm{E}\,U_2(w_2) = \frac{1}{2}(1+w_2) - \frac{1}{1-w_2}\int_{w_2}^1 \int_0^{1-y} \pi^*(t,y)\,dt\,dy. \tag{7}$$

In addition, Γ^* must be a solution to the problem of maximizing SW^* given that equations (6) and (7) hold and that the following inequality constraint holds:

$$\begin{aligned}
&\int_0^1 \int_0^{1-w_1} \Big[v_1^*(w_1,w_2)(1-\pi^*(w_1,w_2)) + w_1\pi^*(w_1,w_2)\Big] \frac{dw_2\,dw_1}{1/2} \\
&\quad + \int_0^1 \int_0^{1-w_2} \Big[v_2^*(w_1,w_2)(1-\pi^*(w_1,w_2)) + w_2\pi^*(w_1,w_2)\Big] \frac{dw_1\,dw_2}{1/2} \le SW^*
\end{aligned}$$

From Lemma 1, this inequality can be rewritten as follows:

$$\begin{aligned}
SW^* &\ge 2\int_0^1 \left[\frac{1}{2}(1-w_1^2) - \int_{w_1}^1 \int_0^{1-x} \pi^*(x,t)\,dt\,dx\right] dw_1 \\
&\qquad + 2\int_0^1 \left[\frac{1}{2}(1-w_2^2) - \int_{w_2}^1 \int_0^{1-y} \pi^*(t,y)\,dt\,dy\right] dw_2 \\
SW^* &\ge \int_0^1 (1-w_1^2)\,dw_1 - 2\int_0^1 \int_{w_1}^1 (1-r-h(r))\,dr\,dw_1
\end{aligned}$$

$$
\begin{aligned}
&\quad + \int_0^1 (1 - w_2^2)\, dw_2 - 2\int_0^1 \int_{w_2}^1 (1 - s - h^{-1}(s))\, ds\, dw_2 \\
SW^* &\geq \frac{2}{3} - 2\int_0^1 \int_w^1 (1 - r - h(r))\, dr\, dw \\
&\quad + \frac{2}{3} - 2\int_0^1 \int_w^1 (1 - s - h^{-1}(s))\, ds\, dw \\
SW^* &\geq \frac{4}{3} - 2\int_0^1 \int_w^1 (1 - t - h(t)) + (1 - t - h^{-1}(t))\, dt\, dw \\
SW^* &\geq \frac{4}{3} - 4\int_0^1 \int_w^1 [1 - t - h(t)]\, dt\, dw.
\end{aligned}
$$

The last step follows from h being symmetric about the 45° line.

Summarizing what we have done so far, we see that if we define

$$
L(w, h) = 2\left[h(w) + w(1 - w - h(w)) + \frac{1}{2}((1 - w)^2 - (h(w))^2)\right]
$$

and

$$
M(w, h) = 4\int_w^1 [1 - t - h(t)]\, dt + L(w, h),
$$

then Γ^* must be a solution to the problem of maximizing $\int_0^1 L(w, h)\, dw$ subject to $\int_0^1 M(w, h)\, dw \geq 4/3$. This is a calculus of variations problem and the solution must satisfy

$$
\begin{aligned}
\frac{\partial L}{\partial h} + \lambda \frac{\partial M}{\partial h} &= 0 \\
2[1 - w - 2h] + \lambda[-4(1 - w) + 2(1 - w - 2h)] &= 0 \\
(1 + \lambda)[1 - w - 2h] &= \lambda(2)(1 - w) \\
1 - w - 2h &= \frac{2\lambda}{1 + \lambda}(1 - w) \\
2h &= 1 - w - \frac{2\lambda}{1 + \lambda}(1 - w) \\
h(w) &= (1 - w)(\frac{1}{2} - \frac{\lambda}{1 + \lambda}).
\end{aligned}
$$

In order to solve for λ, we solve for the case in which the inequality constraint is

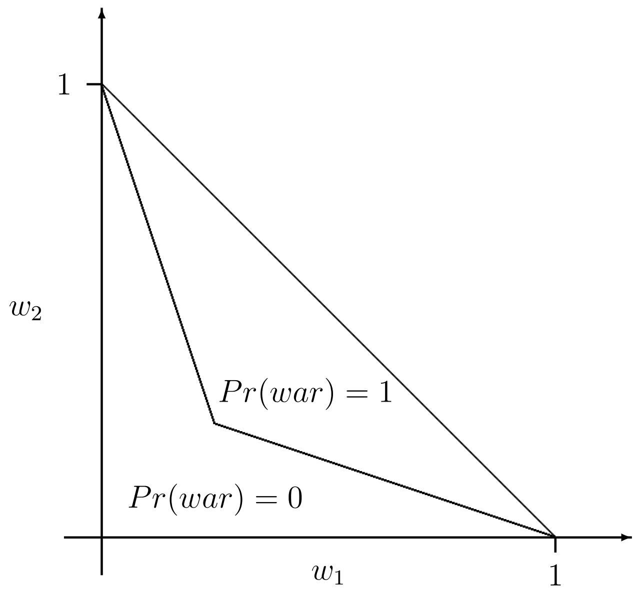

Figure 1: Optimal Mechanism: The boundary of the optimal mechanism is defined by the relation $\min(w_1, w_2) = k[1 - \max(w_1, w_2)]$. For $\min(w_1, w_2) < k[1 - \max(w_1, w_2)]$ a peaceful settlement is proposed (and accepted) and for $\min(w_1, w_2) > k[1 - \max(w_1, w_2)]$ war is induced.

binding. This gives the result. ∎

Remembering that a mechanism is a function that determines a settlement and a probability of war for each possible pair of types, the proof of this theorem involves solving a maximization problem defined over a set of functions. Specifically, the optimal mechanism is a function that must maximize social welfare subject to the incentive-compatibility conditions that are required in equilibrium and the constraint that the sum of the individual payoffs must not exceed total social welfare. This problem is solved via methods of the calculus of variations.

The result stated in Theorem 6 can perhaps best be understood pictorially. Figure 1 illustrates the regions in which war occurs and in which war is avoided in any optimal dispute settlement procedure. Specifically, any such procedure must have the cut-line shown in the figure such that for all type pairs (w_1, w_2) below the line, the procedure provides a settlement that both sides prefer to war, and for all type pairs above the line, war occurs with certainty. Our result is that this is the best that any institution, bargaining protocol, or mediation procedure can achieve.

The characterization given in Theorem 6 has several other interesting implications. First, it implies that, in any optimal procedure, for every type of country i, there is a positive probability of a peaceful settlement and a positive probability of a costly war. Second, this theorem implies that it is optimal to fight "small" wars (wars with low total costs) and settle more costly wars. Finally, from an institutional perspective, social welfare maximization implies an "all-or-nothing" approach to intervention and mediation. That is, in any optimal mechanism, the probability of war is either zero or one. Rather than working to have a low but non-zero probability of war everywhere, our result suggests that institutional peacemakers should act to resolve disputes when the social cost of war are high, but allow costly conflict when social costs are low.

5 Discussion

While the analysis above characterizes general consequences of strategic incentives in the crisis bargaining setting, it is important to note that the objective is not to generate an all inclusive model for war. Rather, the class of games explored is shaped by the types of bargaining situations discussed in the literature on bargaining and war. It is, therefore, worthwhile to take some time to note more clearly what sorts of extensions fit comfortably within the results stated above, and which do not.

Obviously, our results are not sensitive to refinements of the equilibrium set. That is, because the results hold for Baysian-Nash equilibria, they hold for all Perfect Baysian Equilibria, along with any off-the path belief refinements. Similarly, one might wonder whether or not our results speak to the new formal conflict literature on "war as a bargaining process." On the one hand, we model war as a costly lottery, and do not explicitly model the war fighting process. Clearly, this limits our ability to talk about this process. On the other hand, to intelligently (rationally) model the decision to start a war, or a war fighting process, the decision-maker needs to be able to evaluate the continuation values of each alternative, in particular, the continuation values for war and negotiations. In this sense, explicitly modeling the process that resolves the conflict inefficiently really isn't necessary for understanding the decision to go down that inefficient settlement path. We would note, however, that the mapping of war fighting processes into the different types of uncertainty is not a trivial task. For example, in the case where we consider uncertainty about the cost of war in the costly lottery, we are assuming that country two's costs do not influence country one's expected payoffs, and vice versa. If, however, the costs of country two determine how long they fight before they agree to a peaceful settlement, and the size of the acceptable settlement after some fighting depends on the realizations of the player's costs, the structure of the uncertainty created by the war subgame may be hard to characterize. In particular, it might not be a simple private value problem characterized by our costly lottery assumption in Theorems 1 and 2. This is a reason why it is important that we consider different types of uncertainty about war payoffs.

Another issue to consider is our assumption regrading how countries end up at war. There are many alternative assumptions one could make: both countries must choose to fight for war to occur, one country must choose to fight and the other must at least choose not to capitulate, or any single country can start a war unilaterally. The assumption chosen depends on the answer one prefers to the substantive question: can any single country start a war, or does it take to to have a fight? Fey and Ramsay (2007) consider the first case, in the context of mutual optimism arguments, assuming there both countries must choose to fight for a war to occur. He we assume that any country can start a war unilaterally. One might argue that an implicit assumption in our model is the that the target then resists. While a bit of a semantic issue, as a formal matter, our results are based on the assumption that a unilateral change in the status quo is costly and, given the state of information at the time of the decision, an uncertain endeavor. We interpret this cost as the cost of war, but whether or not the term "war" is equally appropriate for the conflict between Britain and Germany in 1914 and the invasions of Denmark and Luxembourg in 1940, is less important than the fact that a unilateral change to the status quo is costly.

Finally, there is the obvious question of the relative value of general results and specific predictions. The question of what modeling approach to take must depend on the question being asked. In particular, questions relating specific institutional forms, norms, or bargaining procedures to the probability of war and peace requires an analysis of the appropriate strategic environment, and call for an approach very different from the one we take here. On the other hand, as a theoretical matter, we may be interested in results that are not institution specific, or that apply to a wide class of different institutions and strategic settings. To the extent that answers to general questions about the relationship between uncertainty and war are the object of interest, no iteration of the process of analyzing specific protocols will reveal these general equilibrium phenomenon. For these questions, the "game-free" approach is a useful tool.

6 Conclusion

This paper seeks to establish general results about the fundamental incentives inherent in crisis bargaining. Rather than fix a particular model and derive results that are limited to this single extensive form, we seek to identify properties shared by all equilibria of all possible models of crisis bargaining. We develop these results across a range of information environments in order to identify the conditions under which the positive probability of war is an unavoidable consequence of private information. We will also characterize the nature of the "second-best" solution in the case where there is always a risk of war.

We have focused on three kinds of uncertainty: uncertainty about the costs of war, uncertainty about the distribution of power, and uncertainty about both costs and outcomes. These various sources of uncertainty correspond to significantly different informational structures with important implications for strategic interaction in a crisis. Uncertainty about costs implies that the underlying player "types" are independent, in the sense that one country's realized preference for fighting does not directly affect another's utility for fighting. Alternatively, the international system may produce uncertainty about the relative strength and the probability of victory. This outcome uncertainty implies a different strategic calculation on the behalf of states. In this situation, the countries' values for war are are interdependent, although information remains uncorrelated. That is, one state's utility for fighting is directly affected by the realization of the other's type. Finally, there is the possibility that the players are uncertain about both cost and the distribution of power. In such a situation, states only know that war is inefficient, i.e., state 1's utility for fighting plus state 2's utility for fighting must sum to less than 1. Given this form of uncertainty the players types are not necessarily interdependent, but they are correlated.

Given these alternative sources of uncertainty, we find that the link between incomplete information and positive probability of war depends in important ways on the types of uncertainty states face. First, we find that whether or not war occurs with some probability where there is uncertainty about the costs of war depends on details of the strategic interaction and the bargaining protocol. When there is uncertainty about the distribution of power, then whether or not there incomplete information is sufficient for a positive probability of war depends on the total social cost of conflict. This result is qualified by the fact that subsidies can decrease the range of parameter values where an institution may resolve a conflict.

Finally, we find that when there is uncertainty about countries' war payoffs, but certainty about inefficiency, there is always a positive probability of war in equilibrium, regardless of the bargaining environment. Our characterization of the second best mechanism shows that it will only be effective at the cost of allowing war when the social cost of conflict is low, while providing successful mediating when the social cost of war is high (similar to interdependent types). Yet, unlike the when there is uncertainty about the distribution of power, there are no circumstances where the second best institution is always peaceful, in the sense that the probability of war is zero for some given set of exogenous parameter values.

In the end, this paper has laid out a simple framework for analyzing some general questions about institutional design. We have begun to sketch out the situations where norms and institutions may be useful for resolving conflicts. In particular, we have focused on how various theoretical sources of uncertainty may influence the possibility for "well designed" institutions to eliminate unwanted, inefficient conflict. It is also true that institutions do not just arrive from nowhere, they are endogenous to the negotiation process between states. Therefore, another unresolved question is, when a peaceful institution exists, would it ever be possible for countries to agree to implement it before they know the particular circumstances under which it will be employed? Finally, we would note that our results and the method we use does not replace the need to better understand the logic of particular institutional arrangements. That is, in order to answer many important questions about international institutions we must depart from the abstract framework of mechanism design and explicitly characterize the actions, utilities, and information of states. So while this paper provides an outline of the possibilities, only careful analysis of particular strategic environments and incentive will explain why institutions in the real world may look and work as they do.

7 Appendix

In this appendix, we state and prove a lemma used to derive the main results in the text. Here, and in the text, we suppose that the functions we consider are measurable, but not necessarily continuous and thus we take all integrals to be Lebesgue integrals.

Lemma 1 *If f is uniformly distributed on W, then any equilibrium of any crisis bargaining game form that allows the unilateral use of force must satisfy*

$$\mathrm{E}\,U_1(w_1) = \frac{1}{2}(1+w_1) - \frac{1}{1-w_1}\int_{w_1}^{1}\int_{0}^{1-x}\pi(x,t)\,dt\,dx$$

and

$$\mathrm{E}\,U_2(w_2) = \frac{1}{2}(1+w_2) - \frac{1}{1-w_2}\int_{w_2}^{1}\int_{0}^{1-y}\pi(t,y)\,dt\,dy.$$

Proof: Suppose there is an equilibrium and suppose a direct mechanism…

In general, the payoff to country i, $i = 1, 2$, from a mechanism $\Gamma = (\pi, v_1, v_2)$, given values w_1 and w_2 is

$$\pi(w_1, w_2)\cdot w_i + (1 - \pi(w_1, w_2))v_i(w_1, w_2).$$

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, let $\mathrm{E}\,U_1(\tilde{w}_1 \mid w_1)$ denote the expected utility of falsely reporting a type $\tilde{w}_1$. This expected utility is given by

$$\mathrm{E}\,U_1(\tilde{w}_1 \mid w_1) = \frac{1}{1-w_1}\int_{0}^{1-w_1}\pi(\tilde{w}_1,t)w_1 + (1-\pi(\tilde{w}_1,t))v_1(\tilde{w}_1,t)\,dt.$$

We also define $\mathrm{E}\,U_1(w_1) = \mathrm{E}\,U_1(w_1 \mid w_1)$. To save on notation, we define

$$\begin{aligned} T_1(\tilde{w}_1 \mid w_1) =& (1-w_1)\,\mathrm{E}\,U_1(\tilde{w}_1 \mid w_1) \\ =& \int_{0}^{1-w_1}\pi(\tilde{w}_1,t)w_1 + (1-\pi(\tilde{w}_1,t))v_1(\tilde{w}_1,t)\,dt, \end{aligned}$$

and let

$$T_1(w_1) = T_1(w_1 \mid w_1) = \int_0^{1-w_1} \pi(w_1,t)w_1 + (1-\pi(w_1,t))v_1(w_1,t)\,dt \tag{8}$$

For later use, we note that $T_1(\tilde{w}_1 \mid w_1)$ and thus $T_1(w_1)$ are absolutely continuous and therefore differentiable almost everywhere.

For country 1 with a type w_1, incentive-compatibility requires that for all $w_1, \tilde{w}_1 \in [0,1]$, $\mathrm{E}\,U_1(w_1 \mid w_1) \geq \mathrm{E}\,U_1(\tilde{w}_1 \mid w_1)$, which is equivalent to

$$T_1(w_1 \mid w_1) \geq T_1(\tilde{w}_1 \mid w_1) \qquad \text{for all } w_1, \tilde{w}_1 \in [0,1]. \tag{9}$$

From this, it is clear that

$$T_1(w_1) = \max_{\tilde{w}_1 \in [0,1]} T_1(\tilde{w}_1 \mid w_1).$$

Thus, by the Envelope Theorem (Milgrom and Segal 2002) it follows that

$$T_1'(w_1) = \frac{\partial T_1(\tilde{w}_1 \mid w_1)}{\partial w_1}\Big|_{\tilde{w}_1 = w_1}. \tag{10}$$

The partial derivative on the right-hand side of this expression is

$$\begin{aligned}\frac{\partial T_1(\tilde{w}_1 \mid w_1)}{\partial w_1} = &\int_0^{1-w_1} \pi(\tilde{w}_1,t)\,dt \\ &- \Big[\pi(\tilde{w}_1, 1-w_1)w_1 + (1-\pi(\tilde{w}_1,1-w_1))v_1(\tilde{w}_1,1-w_1)\Big].\end{aligned}$$

By evaluating this partial derivative at $\tilde{w}_1 = w_1$ and noting that individual rationality (or war consistency) implies that $v_1(w_1, 1-w_1) = w_1$, we see that equation (10) yields the following expression:

$$T_1'(w_1) = \int_0^{1-w_1} \pi(w_1,t)\,dt - w_1.$$

As T_1 is absolutely continuous, it follows that

$$T_1(1) - T_1(w_1) = \int_{w_1}^{1} T_1'(x)\,dx = \int_{w_1}^{1} \int_{0}^{1-x} \pi(x,t)\,dt - x\,dx$$

Noting from equation (8) that $T_1(1) = 0$, we get

$$T_1(w_1) = \frac{1}{2}(1 - w_1^2) - \int_{w_1}^{1} \int_{0}^{1-x} \pi(x,t)\,dt\,dx.$$

Finally, recalling that $T_1(w_1) = (1 - w_1)\,\mathrm{E}\,U_1(w_1)$, we get

$$\mathrm{E}\,U_1(w_1) = \frac{1}{2}(1 + w_1) - \frac{1}{1 - w_1} \int_{w_1}^{1} \int_{0}^{1-x} \pi(x,t)\,dt\,dx,$$

which establishes the result for player 1. An analogous calculation establishes the result for player 2. ∎

References

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

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

Compte, Olivier and Philippe Jehiel. 2006. "Veto Constraint in Mechanism Design: Inefficiency with Correlated Types." *Typescript, Paris-Jourdan Sciences Économiques, Paris* .

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

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

Fey, Mark and Kristopher W. Ramsay. 2007. "Mutual Optimism and War." *American Journal of Political Science* 51(4):738–754.

Gilpin, Robert. 1981. *War and Change in World Poltics*. New York, NY: Cambridge University Press.

Keohane, Robert. 1984. *After Hegemony*. Princeton, NJ: Princeton University Press.

Kindleberger, Charles P. 1973. *The World in Depression:1929–1939*. Berkeley, CA: University California Press.

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

Leventoglu, Bahar and Ahmer Tarar. 2006. "War and Incomplete Information." *mimeo* .

Milgrom, Paul and Ilya Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica* 70(2):583–601.

Morrow, James D. 1985. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33(4):941–972.

Muthoo, Abhinay. 1999. *Bargaining Theory with Applications*. New York: Cambridge University Press.

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

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

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.

Ramsay, Kristopher W. 2004. "Politics at the Water's Edge: Crisis Bargaining and Electoral Competition." *Jounral of Conflict Resolution* 48(4):459–486.

Rubinstein, Ariel. 1982. "Perfect equilibrium in a bargaining model." *Econometrica* 50:97–109.

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

Schultz, Kenneth. 2001. *Democracy and Coercive Diplomacy*. New York: Cambridge University Press.

Slantchev, Branislav L. 2003. "The Principle of Convergence in Wartime Negotiation." *American Journal of Political Science* 97(4):621–632.

Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92(3):623–638.

Smith, Alastair and Allan C Stam. 2006. "Divergent Beliefs in 'Bargaining and the Nature of War'." *Journal of Conflict Resolution* 50(4):614–618.

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

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.