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

Mark Fey†
Kristopher W. Ramsay‡

April 2009

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. Whether and to what extent results derived from these a particular game forms continue to hold when different modeling choices are made is typically unknown. To address this concern, we establish general "game-free" results that must hold in *any* equilibrium of 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. Among these results, we show that under some weak conditions, there is no equilibrium of any crisis bargaining game that has voluntary agreements and zero probability of costly war. Moreover, we show that while uncertainty about the other side's cost of war may be relatively benign in peace negotiations, uncertainty about the other side's strength in war makes it much more difficult to guarantee peaceful outcomes.

*We have benefited from comments received at the University of Rochester's Peter D. Watson Center, Princeton's International Relations Colloquium, the UCSD IR Seminar, and participants at various professional conferences. 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, Randy Stone, and Johannes Urpelainen. 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 & Stam 2006). The common theme of this work is that, because of private information, the two sides in a conflict may be unable to identify suitable settlement terms without learning more about each other. But because of incentives to misrepresent their private information, there is no way to credibly signal this information short of war.

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 Leventoğlu & Tarar (2008)). 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 limits the applicability of results derived from any particular choice of game form.

The approach we employ in our analysis takes 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. Indeed, by using a tool known as the revelation principle along with the constraints implied by a situation where any agreements must be voluntary, we find that there are some general facts about crisis bargaining that can be characterized without reference to a specific game form. That is, we present results that are "game-free" in the sense of Banks (1990) and are not a consequence of particular modeling choices.

We focus on the fundamental role that private information plays in explaining why crisis bargaining can break down even when countries have a common interest in avoiding war. As emphasized by Fearon (1995), just as important as this uncertainty is the resulting incentive for countries to misrepresent their private information. 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 pose an insurmountable obstacle to peaceful settlement, no matter what bargaining procedure is used.

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. We first consider the case in which states are uncertain about their opponent's costs of fighting, but there is no uncertainty about the probability of success in war. For this commonly assumed type of uncertainty, we show that in any equilibrium of any crisis bargaining game both the expected probability of war and the expected utility of each side are weakly decreasing in the cost of war. We then show that there exist crisis bargaining games in which war never occurs and such games have a payoff structure that is necessarily insensitive to the private

[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.

information of each country. The second kind of uncertainty we investigate is one in which the two sides' values for war are interdependent. Such interdependence arises naturally in situations where uncertainty is about relative power, but not costs. We show that when these costs are low, private information and the incentive to misrepresent always generate a positive probability of war, no matter what form the crisis bargaining game takes. We then describe how external subsidies from a third country or international organization can be used to peacefully resolve such a conflict and again offer a characterization of peaceful game forms. This implies that while uncertainty about the other side's cost of war may be relatively benign in peace negotiations, uncertainty about the other side's strength in war makes it much more difficult to guarantee peaceful outcomes.

The final specification of uncertainty that we examine is general uncertainty about the value of war. Here we assume that states know only that war is inefficient and thus the their values for war must be correlated. In this case, we show that every crisis bargaining games necessarily involves a positive probability of war in equilibrium. Given this negative result, we then ask what is the best that can be done? That is, what outcome gives the highest social welfare among those outcomes that can be achieved in some crisis bargaining game? Our characterization of such a "second-best" outcome involves the peaceful resolution of wars with high social cost, at the expense of allowing disputes with low social costs to be settled on the battlefield.

The game-free approach that we employ to establish these results builds on the mechanism design literature in economics, particularly on the classic paper of Myerson & Satterthwaite (1983) on bilateral bargaining.[2] However, our application to international conflict differs from the standard economic literature in several important respects. 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 2001). That is, a country can, at any time, choose to use force and there is no way a country can commit to not do so. We incorporate this in our formal analysis by requiring that any agreements must be voluntary. Thus, in the international arena, 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. To be precise, a standard assumption in the mechanism design literature is that agents cannot back out of a contract that they previously agreed to. Thus, participation decisions are generally assumed to occur at the interim stage, before the terms of the agreement are announced. In contrast, because the lack of enforceable contracts is a fundamental fact in

[2]Several recent surveys of the vast literature on mechanism design in economics include Jackson (2003), Myerson (2008), and Baliga & Sjöström (2008).

international relations, we suppose that either side can veto a potential agreement at any stage.[3]

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 a country's final decision about whether or not to use force should incorporate whatever additional information it has inferred about the opposing country through the negotiation process.[4] A third important difference from the mechanism design literature in economics is that war is always costly and it is therefore common knowledge that there is some peaceful settlement available that both sides would strictly prefer. This is unlike the standard model of bilateral trade due to Myerson & Satterthwaite (1983) in which trade is sometimes inefficient.[5]

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 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. We show that these results only partially extend to two-sided uncertainty. Finally, our results on correlated payoffs to war are also similar to results found independently by Compte & Jehiel (2009). While they also find that there is no mechanism that satisfies a veto constraint, is incentive-compatible, and efficient, we go further and characterize the best possible procedure for resolving disputes.

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 crisis bargaining game, and explains how our analysis uses the revelation principle. Section 4 contains our formal results. Section 5 provides a discussion of the implications of our findings for theories of bargaining, war, and institutional design and the final section concludes.

---

[3] However, see Cramton & Palfrey (1995) for a model in which players can opt out of mechanism. This paper differs from our project, though, because we permit actors to opt out *after* the settlement is generated, instead of only beforehand.

[4] In the mechanism design literature, similar ideas have been discussed by Matthews & Postlewaite (1989), Forges (1999), and Compte & Jehiel (2009). Our construction differs conceptually from ex post individual rationality because the payoff of the outside option in our case, namely war, is not constant but rather depends on the private information of the two sides.

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

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 often discussed by practitioners of formal theory is how sensitive these results can be to the details of the assumed game form. Unfortunately, often the predictions of our models depend crucially on the precise specification of the game we choose. 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.[6] 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 distributional outcome. Other variations, including when disagreement leads to a costly inside option, when players can opt out of bargaining, 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 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). In this paper, Fearon makes a general claim that war can be caused by "private information about relative capabilities or resolve and incentives to misrepresent such information." To be precise, Fearon states "under very broad conditions, if [state] A cannot learn [state] B's private information and if A's own costs are not too large, then state A's optimal grab produces a positive chance of war." (p. 394) Thus, the paper makes the general claim that, when costs are small, private information is a sufficient condition for the risk of war.[7] To establish this claim, the paper shows that this result occurs in an ultimatum offer game with one-sided incomplete information. But would the same outcome occur in other models with alternative bargaining procedures or different assumptions about information? There is no way to answer these questions based on the results of a specific model. However, in this paper, we show how these questions can, in fact, be answered for a broad class of crisis bargaining 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

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

[7] To be clear, this is a sufficient condition for a positive probability of war, not the certain occurrence of war.

depend on the analysis of a specific game form.[8] Indeed, Powell (2004) emphasizes the 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 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.[9]

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.[10] 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.

[8]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).

[9]A formal description of and intuition for the revelation principle are given below. See Chwe (1999) and Baron (2000) for other applications of the revelation principle in political science.

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

3 Formalizing the Approach to Crisis Bargaining

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 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 $w_i = p_i - c_i$.

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* $t_i \in T_i \subseteq \mathbb{R}$ 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 $w_i(t)$.

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 $\pi^g(a)$, and, in the case of a settlement, the value of the settlement to country 1, $v^g(a)$. We will assume that any potential settlement is efficient and therefore the value of the settlement to country 2 is given by $1 - v^g(a)$. We sometimes write $v_i^g(a)$ 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

$$u_i(a,t) = \pi^g(a) \cdot w_i(t) + (1 - \pi^g(a))v_i^g(a). \tag{1}$$

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. As in the canonical take-it-or-leave-it bargaining game studied by Fearon (1995), the private information of the two sides only affects payoffs in the case of war. That is, the terms of a peaceful settlement are determined by the actions of the two sides in the bargaining game and, in the case of peace, it is only these terms that determine the countries' payoffs and not their private information about their war-fighting ability.

Of course, a country's private information can affect its behavior at the bargaining table. Formally, the actions taken in the game can depend on the countries' types, and we reflect this fact by defining a strategy for country i by a function $s_i : T_i \to A_i$. 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. For a given equilibrium s^* of G, define $U_i(t_i)$ to be the expected utility of this equilibrium for a type t_i of country i. Analytically, we let

$$U_i(t_i) = \int_{T_j} u_i(s^*(t_i, t_j), t)\, dF(t_j \mid t_i), \tag{2}$$

where $i \neq j$ and $u_i(a,t)$ is given by equation (1).

In the remainder of this section, we formalize the method of game-free analysis 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. For a type pair $t = (t_1, t_2)$, this equilibrium generates an equilibrium probability of war $\pi(t) = \pi^g(s^*(t))$ and an equilibrium value of settlement to country i, $v_i(t) = v_i^g(s^*(t))$. 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. If it is then an equilibrium of the direct mechanism for all types to "tell the truth," thus choosing the action equal to their 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 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 is any incentive to lie about one's type? 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 qualitative properties of crisis bargaining that play a significant role in our later analysis.

First, given the anarchic nature of the international system, any peaceful agreement reached must be voluntary. In particular, we suppose that a country always has the option of rejecting a proposed settlement $v^g(a)$ if it thinks it will be better off by resolving the crisis by force. The form of this rejection can literally be the use of military force or it can be an escalation of the crisis which either leads to war or to the other side capitulating. Whichever is true, this feature of the bargaining context insures each side can act in such a way that the outcome is either war or a voluntary settlement that that side prefers to war. Formally, we say a crisis bargaining game form has *voluntary agreements* if, for $i \neq j$, there exists an action $\tilde{a}_i \in A_i$ such that either $\pi^g(\tilde{a}_i, a_j) = 1$ for all $a_j \in A_j$ or for all $t \in T_p$, $v_i^g(\tilde{a}_i, a_j) \geq w_i(t)$.[11] Put simply, in a crisis bargaining game form with voluntary agreements there is no way to force a country to accept an agreement that makes it worse off than it would expect to be by going to war.

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

[11] For an example of such an action, consider the take-it-or-leave-it bargaining game of Fearon (1995). For country 1, the action $\tilde{a_1}$ consists of demanding the whole pie. This demand will either be rejected, leading to war, or accepted, which results in a payoff that is higher than country 1's war payoff. Likewise, for country 2 in this game, rejecting any offer leads to war which thus satisfies the definition. So this game has voluntary agreements.

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.[12] As is standard, we assume that this belief is formed via Bayes' Rule, whenever possible.

Combining this updating with the our assumption on voluntary agreements, it is easy to show the following result is a consequence of the revelation principle.[13]

Result 2 *Suppose that s^* is an equilibrium of a crisis bargaining game form that has voluntary agreements. 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 $t \in T_p$ such that $\pi(t) \neq 1$.*

The intuition for this result is straightforward. Referring back to equation (1), we see that in a game form that has voluntary agreements, if $\pi^g(s^*(t)) \neq 1$, then this condition simplifies to the requirement that, the equilibrium settlement value to country i must be at least as big as the expected war payoff to country i. That is, 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 and $\pi(t) = \pi^g(s^*(t)) < 1$ for some $t = (t_1, t_2)$, it must be that this deviation is not profitable, which is true only if $v_i(t) \geq E[w_i(t) \mid \mu_i(v_i, t_i)]$. In other words, if agreements are voluntary and 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.

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 makes 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 $\pi^g(s^*(t)) = 0$ for all $t \in T_p$. In other words, an always peaceful equilibrium of a game form is one in which no possible pair of types ever end up abandoning a peaceful resolution of the dispute and resorting to force. Adding this requirement to Result 2 gives our final result.

Result 3 *Suppose that s^* is an always peaceful equilibrium of a crisis bargaining game form that has voluntary agreements. Then there exists an incentive-compatible direct mechanism such that for all $t \in T_p$, $\pi(t) = 0$ and $v_i(t) \geq E[w_i(t) \mid \mu_i(v_i, t_i)]$.*

[12] Although additional information may be revealed by the specific actions taken by countries in a specific game, we focus on the information revealed by the final settlement offer because this information will be available in all crisis bargaining games. Incorporating the additional information possible revealed in a specific game would only strengthen our results.

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

The importance 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 has voluntary agreements.

4 Game-Free Results

In this section we establish several general results that hold for *every* crisis bargaining game that has voluntary agreements. Before doing so, though, we must identify the appropriate information structure for our analysis. Historically, the conflict literature has long held that uncertainty is a central cause of conflict among countries, but has varied with respect to the type of uncertainty it views as important. Some scholars focus on uncertainty about the relative strength of the countries (Blainey 1988, Organski & Kugler 1980) while others concentrate on uncertainty about the costs of conflict or the resolve of countries (Kydd 2003, Morrow 1985, Schultz 1998, Ramsay 2004).[14]

In particular, we focus 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 values of war to each country 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. In this situation, the countries' values for war are interdependent, although information remains uncorrelated. Finally, there is the possibility that the two sides are uncertain about both cost and the distribution of power. In such a situation, states only know that war is inefficient. In this case, war payoffs are not necessarily interdependent, but information is correlated.

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, which is also often interpreted as a country's level of resolve (Schultz 2001, Smith 1998) or preference for fighting. This is a situation with independent private values. That is, the realization of one state's cost of war does not directly influence the utility of the other

[14]It is often the case that uncertainty about costs is said to also model uncertainty about "resolve," where resolve is loosely defined as how much one side "cares" about the issue or item of dispute. The equivalence comes from the fact that, if costs are known, but the value of the prize is unknown, we can normalize the value of the prize generating a new "relative cost" c/v_i which is a privately known value.

side. 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.[15] Denote a pair of types $(c_1, c_2) = c \in C = C_1 \times C_2$.

We begin our analysis of this information structure by identifying some qualitative features of equilibrium play in crisis bargaining games with uncertain costs. Often, such qualitative features are central components in the analysis of a particular model. For example, in Fearon (1995)'s classic model of take-it-or-leave-it international bargaining, it is possible to show that the expected probability of war is weakly decreasing in the cost of war for both countries. The following results demonstrates that this qualitative feature of Fearon's model is completely general—it will hold in any equilibrium of any crisis bargaining game with uncertainty about the cost of conflict. The result also shows that a similar conclusion can be reached about the overall expected utility of each country. For country i, define the expected probability of war, given strategy profile s, by

$$\Pi(c_i) = E\pi^g(c_i) = \int_{C_j} \pi^g(s(c_i, c_j))\, dF_j(c_j).$$

Also, recall that the expected utility of this strategy profile is given by

$$U_i(t_i) = \int_{C_j} u_i(s(c_i, c_j), c_i)\, dF_j(c_j).$$

We then have the following theorem.

Theorem 1 *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 and let s^* be any equilibrium of G. Then $\Pi(c_i)$ and $U_i(c_i)$ are both weakly decreasing in c_i, for $i = 1, 2$.*

Proof: Suppose that s^* is an equilibrium of a crisis bargaining game form. Then by Result 1, there exists an incentive-compatible direct mechanism yielding the same outcome as s^*. This direct mechanism is given by $\pi(c) = \pi^g(s^*(c))$ and $v_i(c) = v_i^g(s^*(c))$. Therefore the expected utility of type c_i from equation (2) can be written

$$U_i(c_i) = \int_{C_j} \pi(c_i, c_j)(p_i - c_i) + (1 - \pi(c_i, c_j))v_i(c_i, c_j)\, dF_j(c_j)$$

[15]Because the private information of country i is the cost of war, here we use the notation c_i rather than t_i for the type of country i.

and the expected utility of falsely reporting a type $\tilde{c}_i$ is

$$U_i(\tilde{c}_i \mid c_i) = \int_{C_j} \pi(\tilde{c}_i, c_j)(p_i - c_i) + (1 - \pi(\tilde{c}_i, c_j))v_i(\tilde{c}_i, c_j)\, dF_j(c_j).$$

Incentive-compatibility requires that for all $c_i, \tilde{c}_i \in C_i$,

$$U_i(c_i \mid c_i) \geq U_i(\tilde{c}_i \mid c_i) \qquad \text{and} \qquad U_i(\tilde{c}_i \mid \tilde{c}_i) \geq U_i(c_i \mid \tilde{c}_i).$$

Adding these two inequalities and simplifying yields

$$\int_{C_j} \pi(c_i, c_j)(p_i - c_i) - \int_{C_j} \pi(c_i, c_j)(p_i - \tilde{c}_i) \geq \int_{C_j} \pi(\tilde{c}_i, c_j)(p_i - c_i) - \int_{C_j} \pi(\tilde{c}_i, c_j)(p_i - \tilde{c}_i)$$
$$\int_{C_j} \pi(c_i, c_j)(\tilde{c}_i - c_i)\, dF_j(c_j) \geq \int_{C_j} \pi(\tilde{c}_i, c_j)(\tilde{c}_i - c_i)\, dF_j(c_j).$$

From this, it follows that if $\tilde{c}_i > c_i$, then $\Pi(\tilde{c}_i) \leq \Pi(c_i)$ and so $\Pi(c_i)$ is weakly decreasing.

We now show that $U_i(c_i)$ is also weakly decreasing. Applying Lemma 1 in the Appendix and recalling that $w_i = p_i - c_i$, we get

$$\frac{dU_i(c_i)}{dc_i} = \int_{C_j} \pi(c)(-1)\, dF_j(c_j).$$

As $\pi(c)$ is always nonnegative, the result follows. ∎

Thus, monotonicity of utility and the probability of war are general qualitative characteristics of any equilibrium of any game with uncertainty about the costs of war.[16] These types of comparative statics represent testable predictions that are robust to the particular formalization of crisis bargaining, as advocated by Powell (2004). Moreover, as this result depends only on the incentive compatibility constraints inherent in the revelation principle, it holds for all crisis bargaining games, even those without voluntary agreements. In this way, we see that the monotonicities identified in the theorem are purely a consequence of the incentives to misrepresent private information that are fundamental to the crisis bargaining environment.

Of course, in a peaceful equilibrium the probability of war is constant at zero which satisfies Theorem 1 in a trivial way. We now turn to inquiring about the characteristics of such equilibria. We begin with a simple existence result.

[16]See Banks (1990) for additional monotonicity results that hold in Bayesian games with one-sided incomplete information. Unfortunately, these additional results do not extend to the case of two-sided incomplete information studied here.

Theorem 2 *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 has voluntary agreements in which an always peaceful equilibrium exists.*

Proof: To prove this result, it is enough to give an example of a game form that has voluntary agreements and that has an always peaceful equilibrium. The following very simple example shows that this is indeed the case. Consider a game form that has voluntary agreements such that $g_i^v(a) = p_i$ for all $a \in A$ and and $\pi^g(a^*) = 0$ for some $a^* \in A$. In this case, the 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 there exists at least one game form that has voluntary agreements and that can eliminate the possibility of war. The game form described in the proof has a particularly simple payoff structure in which the potential settlement is always equal to a country's (commonly known) likelihood of success from war. One example of such a game form would be a direct "arbitration" mechanism in which an arbitrator would present both sides with a take-it-or-leave-it offer. Since this agreement would make both sides better off, regardless of their costs for fighting, it would provide 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, the mere existence of a game form with a peaceful equilibrium is not a very satisfying 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 2 establishes that Fearon's claim that incomplete information about costs leads to the risk of war is not always true; there exist instances where it does not hold. However, without a better characterization of equilibrium incentives, we do not know if this case is an isolated exception to Fearon's claim or if peaceful equilibria are the norm and it is Fearon's example that is the exception.

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

$$Ev_i^g(a_i \mid s_j) = \int_{C_j} v_i^g(a_i, s_j(c_j))\, dF_j(c_j).$$

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

Theorem 3 *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 has voluntary agreements. Then an equilibrium s^* of G is always peaceful only if, for $i = 1, 2$, (1) $Ev_i^g(s_i^*(c_i) \mid s_j^*) = p_i$ for all $c_i \in C_i$, (2) $Ev_i^g(a_i \mid s_j^*) \leq p_i$ for all $a_i \in A_i$ and (3) $v_i^g(s^*(c)) \geq p_i - c_i$ for all $c \in C$.*

Proof: Suppose that s^* is an always peaceful equilibrium of a crisis bargaining game form that has voluntary agreements. Then by Result 3, there exists an incentive-compatible direct mechanism such that $\pi(c) = 0$ and

$$v_i(c) \geq E[w_i(c) \mid \mu_i(v_i, c_i)] = p_i - c_i$$

for all $c_i \in C_i$ and $i = 1, 2$. Here, $E[w_i(c) \mid \mu_i(v_i, c_i)] = p_i - c_i$ because the value of war to country i does not depend on the value of c_j. This direct mechanism is given by $\pi(c) = \pi^g(s^*(c))$ and $v_i(c) = v_i^g(s^*(c))$.

Because $\pi(c) = 0$, applying Lemma 1 in the Appendix yields $U_i'(c_i) = 0$ and therefore $U_i(c_i) = \int_{C_j} v_i(c_i, c_j) dF_j(c_j)$ 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

$$\int_{C_j} v_i(c_i, c_j)\, dF_j(c_j) = \int_{C_j} v_i(0, c_j)\, dF_j(c_j) \geq p_i,$$

for all $c_i \in C_i$, $i = 1, 2$. To show that this expression must hold with equality, suppose not. Then

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

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 statement (1) of the theorem. The remaining two statements follow directly from the assumption that s^* is an equilibrium and thus no profitable deviation is possible. ∎

This theorem shows that the simple payoff structure used in the proof of Theorem 2 is, in fact, a general property of peaceful equilibria. Every peaceful equilibrium must have this same simple payoff structure in which the terms that a country accepts are completely insensitive to the costs or resolve of the country. High cost and low cost countries (and all types in between) 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.[17] A general lesson from this result, then, is that peaceful equilibria must be "simple" equilibria that do not depend on the private information of the countries. That is, an always peaceful agreement necessarily depends only on publicly observable information and cannot vary in response to the privately known costs or resolve of countries.

We can understand the general idea of this characterization of peaceful equilibria by observing that it is the generalized expression of Fearon's "incentive to misrepresent." Fearon showed that this incentive renders preplay communication non-credible. Theorem 3 shows that this same incentive forces peaceful equilibria to be completely insensitive to the private information that countries possess. This follows because if a peaceful equilibria did provide different expected settlements to different types, the type getting the worse settlement would have an incentive to mimic the behavior of the type getting the better settlement. Such a deviation is profitable because in a peaceful equilibrium the discipline generated by the risk of war does not exist, which means we could not have had an equilibrium that responds to the countries' private information.

It is immediately clear from this general lesson that because many examples of crisis bargaining games in the literature have equilibria that do not conform to this simple payoff structure, these examples support the claim that incomplete information about costs leads to war. Indeed, given the severity of the necessary condition that strong types receive the same expected settlement as weak types, it does seem intuitively clear that many game forms will fail this necessary condition. In fact, we can strengthen this observation into the following necessary and sufficient condition:

Corollary 1 *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 has voluntary agreements and let s^* be any equilibrium of G. Then $U_i(c_i) \neq U_i(\tilde{c}_i)$ for some $c_i, \tilde{c}_i \in C_i$ if and only if there is a positive probability of war in equilibrium.*

Proof: The "if" direction follows directly from Theorem 3. To show the other direction, suppose there is a positive probability of war. This means that $\pi^g(s^*(c)) > 0$ on a set of

[17]It would be very useful if this result implied that completely pooling equilibria were the only type of peaceful equilibrium possible. However, it can be shown that there also exist non-pooling strategies with the property that, in expectation, the payoff to settlement is constant and that therefore satisfy the conditions given in the theorem. Details of this example are available from the authors.

positive measure and by Lemma 1 in the Appendix, $U_i'(c_i) > 0$ on this set. The result follows. ∎

In other words, this corollary establishes that two types can have different expected utilities if and only if war occurs with positive probability. If we interpret private information about costs as a country's "resolve," then this result can be stated more clearly in the following form. Crisis bargaining games in which more resolved countries expect to be better off than less resolved countries in equilibrium are those with a positive probability of war. Put another way, this result shows that the *only* way that countries stand to gain from their resolve is through running the risk of war, as in the well-known "risk-reward" tradeoff.

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 the quality of their military 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.

To model this we assume 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 independently 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 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.

Note that we do not assume that the function $p(t_1, t_2)$ is necessarily continuous. [18] If we

[18] But $p(t)$ is integrable because it is monotone in each argument.

imposed the additional assumption that $p(t)$ is continuously differentiable in each argument, then Lemma 1 in the Appendix implies that the expected utility of country i, $U_i(t_i)$, is weakly increasing in t_i. In other words, regardless of the game form, a stronger type of a country is never worse off than a weaker type of that country. Thus the result for expected utility given in Theorem 1 with uncertainty about costs continues to hold with uncertainty about relative power.

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 $V_1(t_1, v) = \{t_2 \mid v_1(t_1, t_2) = v\}$ be 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.[19] In this setting, then,

$$E[w_1(t) \mid \mu_1(v_1, t_1)] = E[p(t_1, t_2) \mid V_1(t_1, v_1)] - c_1, \tag{3}$$

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 4 *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*

[19] 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

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

crisis bargaining game form that has voluntary agreements.

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 has voluntary agreements. By Result 3, 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

$$U_1(\hat{t}_1 \mid t_1) = \int_{T_2} v_1(\hat{t}_1, y)\, dF_2(y). \tag{4}$$

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 (4), 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 $\hat{t}_1$.[20] 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 (3) evaluated at the type pair $t_1 = \bar{t}_1$ and $t_2 = \bar{t}_2$ to get

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

Taking expectations of both sides, we get

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

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

$$\bar{U}_1 = \int_{T_2} v_1(\bar{t}_1, t_2)\, dF_2(t_2) \geq \int_{T_2} p(\bar{t}_1, t_2)\, dF_2(t_2) - c_1. \tag{5}$$

By a similar argument, it follows that

$$\bar{U}_2 = \int_{T_1} v_2(t_1, \bar{t}_2)\, dF_1(t_1) \geq \int_{T_1} [1 - p(t_1, \bar{t}_2)]\, dF_1(t_1) - c_2. \tag{6}$$

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

[20] This result would follow immediately from Lemma 1 in the Appendix if we made the additional assumption that $p_i(t)$ was continuously differentiable in t_i.

all pairs (t_1, t_2), it follows that

$$\int_{T_1}\int_{T_2} [v_1(t) + v_2(t)]\, dF_2(t_2)\, dF_1(t_1) = 1$$

$$\int_{T_1}\int_{T_2} v_1(t)\, dF_2(t_2)\, dF_1(t_1) + \int_{T_2}\int_{T_1} v_2(t)\, dF_1(t_1)\, dF_2(t_2) = 1$$

$$\int_{T_1} \bar{U}_1\, dF_1(t_1) + \int_{T_2} \bar{U}_2\, dF_2(t_2) = 1$$

$$\bar{U}_1 + \bar{U}_2 = 1.$$

Therefore, adding inequalities (5) and (6) 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 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 form of interaction between the two countries. In other words, no possible bargaining norm, protocol, or institution can eliminate war when the social cost of war is small and the relevant strategic uncertainty concerns the distribution of power.

Phrased as a general result, Theorem 4 shows that small costs of war is a sufficient condition for the impossibility of completely peaceful settlements. We thus can view this result as extending the claim of Fearon (1995) that when costs are low private information leads to a positive probability of war from the case of uncertainty about costs to the case of uncertainty about relative power. The reason that the war problem is particularly hard to solve in games with interdependent types is because, in such games, the strongest type of a given country knows that every type of their opponent cannot be a stronger adversary. So, from the strongest type's perspective, war is an relatively attractive option and therefore in order to persuade the strongest type to forgo this option, such types must get relatively generous terms from peaceful settlements. This in turn creates an incentive for weaker types to pretend that they are strong in order to secure this high payoff. The only way a peaceful settlement can exist with these incentives is if the settlement gives *every* type of a country the same payoff as the *strongest* type of the country. But when the total costs of war are less than the critical value $\bar{c}$, such settlements are impossible.

This intuition suggests that when costs are large enough, peaceful settlements are possible. This is exactly what we show in the next result. Specifically, we show that if the total costs of war are greater than the critical value $\bar{c}$, it is possible to always avoid war.

Theorem 5 *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 has voluntary agreements in which an always peaceful equilibrium exists.*

Proof: As with Theorem 2, it is enough to give an example of a crisis bargaining game form that has voluntary agreements which has an always peaceful equilibrium. Consider a crisis bargaining game form that has voluntary agreements such that $v^g(a) = P_1(\bar{t}_1) - c_1$ for all $a \in A$ and and $\pi^g(a^*) = 0$ 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 takes an action that leads to 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. When the total costs of war are less than $\bar{c}$, peaceful equilibria are impossible, but when total costs exceed $\bar{c}$, they are not. It is important to note that the value of $\bar{c}$ depends only on the distributions F_1 and F_2 and the function $p(t_1, t_2)$. Thus, in any crisis bargaining game with voluntary agreements, whether or not it is possible to always avoid war depends on the form of the two sides' uncertainty and how relative power affects the outcome of war, but not on the details of the game form itself!

Another important feature of Theorem 5 is that it implies that 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 2 *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 has voluntary agreements 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.[21]

We conclude our analysis of this case by presenting a characterization of peaceful equilibrium that is similar to the results in the previous section.

Theorem 6 *Suppose costs are common knowledge but each country is uncertain about the probability of winning a war. Suppose also that $p(t)$ is continuously differentiable in t_1 and t_2. Let G be any crisis bargaining game form that has voluntary agreements and let s^* be any equilibrium of G. Then $U_i(t_i) \neq U_i(\tilde{t}_i)$ for some $t_i, \tilde{t}_i \in T_i$ if and only if there is a positive probability of war in equilibrium.*

Proof: Suppose s^* is an always peaceful equilibrium of a crisis bargaining game form that has voluntary agreements. Then it follows from Lemma 1 in the Appendix that $U_i(t_i)$ is constant in t_i, which establishes one direction of the theorem. The reverse direction follows from the same argument as the proof of Corollary 1. ∎

This theorem shows that our previous results for uncertainty about costs continue to hold in this information environment as well. Moreover, an implication of this result is that, as in the previous section, every peaceful equilibrium of every crisis bargaining game with uncertainty about relative power must give every type the same expected settlement. Therefore, the same conclusions drawn in the previous section continue to apply to this uncertainty condition.

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

[21] 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).

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. $T_p = W$.[22] In what follows we suppose the common prior f is uniform on W.

We begin by showing that, as with uncertainty about costs, in any crisis bargaining game with this information structure the equilibrium probability of war and expected utilities are monotonic in the war payoff of each country. Here the result is that each is increasing in w_i, which is the same as having decreasing comparative statics in the case of costs. For country i, define the expected probability of war, given strategy profile s, by

$$\Pi(w_i) = E[\pi^g(w_i)] = \int_{W_j} \pi^g(s(w_i, w_j))\, dF_j(w_j)$$

and $U_i(w_i)$ as in equation (2). Then we have the following result.

Theorem 7 *Suppose war payoffs w_i are private information and f is uniform on W. Let G be any crisis bargaining game form and let s^* be any equilibrium of G. Then $\Pi(w_i)$ and $U_i(w_i)$ are both weakly increasing in w_i, for $i = 1, 2$.*

Proof: The proof for the probability of war closely parallels the argument in the proof of Theorem 1. The proof for expected utility follows directly from differentiating the expressions for $U_i(w_i)$ given in Lemma 2. ∎

Our next 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.

[22]Because the private information of country i is the value of war, here we use the notation w_i rather than t_i for the type of country i.

Theorem 8 *If war payoffs w_i are private information and f is uniform on W, then no always peaceful equilibrium exists in any crisis bargaining game form that has voluntary agreements.*

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 has voluntary agreements. By Result 3, 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

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

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

To finish the argument, recall that $v_1(w) + v_2(w) = 1$ for all $w \in W$ and perform the following calculations:

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

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. Then the theorem tells us that, under its assumptions, no such procedure can exist 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.

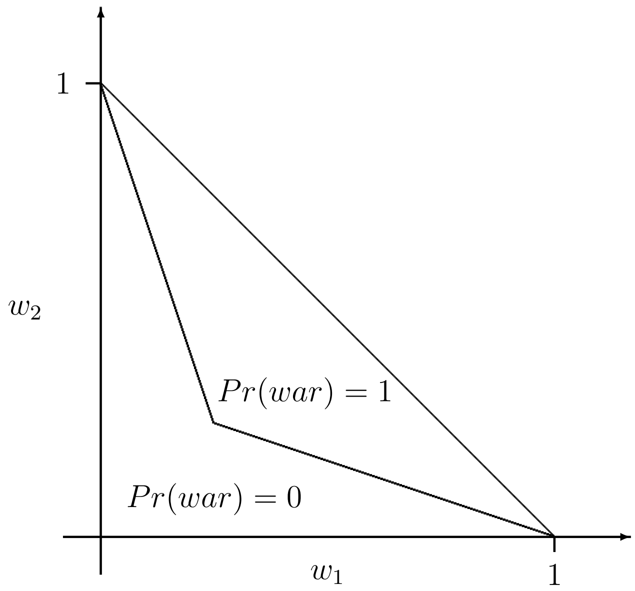

Figure 1: Equilibrium outcome of any optimal settlement procedure

Because no peaceful equilibria exist in any crisis bargaining game with this information structure, it is of course not possible to provide a characterization of peaceful equilibria as we have done in the previous sections. Instead, 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 such "second best" procedures.[23]

Theorem 9 *If f is uniformly distributed on W, then any equilibrium outcome that maximizes social welfare among all crisis bargaining game forms that have voluntary agreements must satisfy*

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

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

The proof of this theorem is in the appendix. 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

[23] Andersen, Harr & Tarp (2006) address a similar question in the specific context of IMF loan allocations by the United States.

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 9 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 9 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. 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

As we have emphasized, the strength of the game-free method of analysis presented here is the ability it provides to establish results that hold whatever the form of the strategic interaction between countries. That is, our results apply to 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. In this way, we have been able to give a clear picture of the general consequences of private information and the incentives to misrepresent this information on the possibility and form of war and peace.

Using our techniques, we have established several general results. The first group of such results concerns the monotonicity properties of equilibria. We show that crisis bargaining games with two-sided private information have some, but not all, of the monotonicity properties described by Banks (1990) for games with one-sided private information. Theorems 1 shows that countries' expected probability of war and expected utility are both decreasing in their costs. So the general intuition that low cost types have higher expected utilities from crisis bargaining, but also face greater risk of war, carries over from the setting with one-sided incomplete information. Unlike Banks (1990), however, we are unable to guarantee monotonicity in expected settlements with two-sided incomplete information about costs. Similarly, Theorem 7 shows that for correlated values these monotonicity results for the probability of war and utility also hold in every crisis bargaining game. In fact, Lemmas 1 and 2 imply monotonicity of utilities holds in each of the kinds of uncertainty that we consider. Thus, this finding is truly general.

A second interesting general result occurs in the interdependent case, where countries are uncertain about relative power. Here we find that the conditions that determine whether or not peaceful equilibria exist do not depend on any aspect of the game form. As seen in Theorems 4 and 5, the existence of peaceful equilibria depends only on the costs of fighting and the form of the uncertainty about relative power, and not on the details of the bargaining. When the conditions for peaceful equilibria exist, then a simple mechanism that gives both sides at least the best they could hope for from war is a simple solution. Otherwise, when such an arrangement is not possible, there is no bargaining process, no matter how complicated, that ensures peace.

A final interesting result of our analysis is that the effect of private information and the incentive to misrepresent varies significantly as we vary the structure of the decision-maker's uncertainty. This is true even though in each framework we consider there is common knowledge that a settlement exists that both sides prefer to war. The fact that in each framework the conditions on the game form that lead to war are different makes clear that what kind of uncertainty exists in the international environment leads to important strategic differences in the incentives of countries operating in such an environment. At the same time, our focus on the informational roots of conflict is not meant to suggest that this is the only potential cause of war. To be clear, this paper does not claim to present an all-inclusive model for war. Rather, we have attempted to provide game form free results for an important class of strategic situations discussed in the literature on private information and war.

One basic assumption that we have used to define this class of games is that each side can always take some action that guarantees their expected war payoff. While our view is that this assumption is a natural consequence of the anarchic nature of the international arena, there may remain some question as to how best to describe the process leading countries to war. As discussed earlier, our assumption that agreements are voluntary explicitly includes models in which a country can capitulate in order to avoid or bring an end to fighting. One alternative assumption is that war occurs only if both countries agree to fight and otherwise the status quo or an imposed settlement prevails. This assumption has been used to consider the idea of mutual optimism as a cause of war. Central to this literature (Wittman 1979, Blainey 1988) is the claim that *both* countries must want to fight for a war to occur, hence the name *mutual* optimism. Not surprisingly, this alternative assumption yields different conclusions about the possibility of war. Thus our results for the case of voluntary agreements highlight the role of private information and the incentives to misrepresent and can be interpreted as the complement to the mutual optimism argument.[24]

To amplify this point, the assumption of voluntary agreements that we make is intended to be very general. We require that there be an action that allows decision-makers to reject settlements that give them less than they might expect from fighting, but we do not require that this action leads to a war lottery or even immediate fighting. This flexibility implies that our results hold in many different kinds of games. For example, the fact that our results hold in game forms that include capitulation as a possible outcome distinct from war and peace also aids in clarifying the application of our model to several historical cases. While there are many clear examples of the direct, unilateral use of force such as the Japanese attack on Pearl Harbor and the American invasion of Iraq in 2003, there are other historical cases that could be better categorized as capitulation, such as the invasions of Denmark and Luxembourg in 1940. Thus, our results apply to both kinds of circumstances, including those that involve armed aggression but that ultimately do not involve significant armed conflict.

The other basic assumption we use to define the class of crisis bargaining games is that whatever form the interaction of the two sides takes, the end result is always either a negotiated settlement or war. Dividing outcomes into these two categories seems to us to be quite natural. At the same time, because we make the common assumption that the war outcome is a game-ending costly lottery, we should consider how our results speak to the recent formal conflict literature on "war as a bargaining process" (Filson & Werner 2002, Slantchev 2003, Powell 2004). In a manner similar to Powell (2004), we can relax the assumption that fighting is a game-ending move by supposing that fighting creates a chance of a militarily decisive outcome but can also be indecisive and lead to further negotiations or further fighting. We can fit such a model into our theoretical framework by identifying all terminal nodes with a decisive military outcome as "war" outcomes and all other terminal nodes as "peaceful settlement" outcomes. However, as a number of rounds of fighting can precede any such outcome, we must make changes to two of the assumptions we make about our class of crisis bargaining games. First, we must modify the assumption that peaceful settlement are efficient because if the countries fight for some time before reaching a peaceful settlement, the payoffs to the two countries for such a strategy profile will sum to less than one. Second, for a similar reason we must modify our assumption that a country's war payoff does not depend on the actions taken in the game. The longer the countries

[24]For more on the issue of mutual optimism and assumptions about how countries end up in war, see Fey & Ramsay (2007) and Fey & Ramsay (2008).

choose to fight before a decisive military outcome results, the lower is the expected value of this outcome.

How do our results fare if we make these two changes to our framework in order to accommodate models of war as a bargaining process? Reassuringly, it is possible to show that all of our main results continue to hold. In the case of the first change, we show in a companion paper (citation omitted) that permitting settlement outcomes to be inefficient does not change any of the qualitative features of our results. This is also true with the second change, permitting the war payoff to vary with the actions take in the game. Rather than provide formal proofs here, instead we briefly explain why our results continue to hold.[25] Our reasoning rests on the fact that even if a country's war payoff depends on its actions, as its equilibrium actions depend on its type, we can use the revelation principle to find a direct mechanism in which a country's war payoff depends directly on its type. In this way, Lemma 1 continues to hold. As this lemma is a key ingredient in our proofs, this result, coupled with the fact that the actions of a country can only change its payoff by incurring the costs of multiple rounds of fighting, yields the same characterizations as Theorems 1 through 6.

6 Conclusion

This paper has set out 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 develop a method to identify properties shared by all equilibria of all possible models of crisis bargaining. We have developed these results across a range of information environments in order to identify conditions under which the positive probability of war is an unavoidable consequence of private information and conditions under which peaceful resolution of conflicts is possible. We have also characterized the nature of such peaceful resolutions and given a "second-best" solution in the general case in which there is always a risk of war. In short, we have found that while it is possible to give general results that support the rationalist explanation of war as a consequence of private information and the incentive to misrepresent this information, the link between private information and war depends in important ways on the types of uncertainty that states face.

Throughout this paper, we have emphasized the generality of our approach and the necessity of utilizing such a general approach to justify broad claims about the causes of war. Although we are able to give results that identify specific relationships between information and outcomes, as in Theorem 1, there is an obvious tradeoff between general results and

[25] Details are presented in the Reviewer's Appendix.

specific predictions. The relative value of the two must depend on the question being asked. In particular, questions regarding particular behaviors in a specific institutional setting 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, independent of institutional specifics, 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. As a guiding principle, we endorse the "two-step" approach outlined in Banks (1990) in which general results are established first using the approach developed here and then additional behavioral predictions are generated from a particular game form that captures specific institutional features of interest.

In the end, while this paper has laid out a framework for analyzing some general questions about war and peace, several additional aspects of this approach remain to be explored. First, while 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. It is natural, then, to consider the process by which countries bargain over possible game forms that then govern their interaction. It should be possible to use our approach to deal with such a situation by viewing the institution selection process as itself part of a more broadly defined game and applying the revelation principle.[26] Second, although each of the three kinds of uncertainty that we have considered in this paper have involved one-dimensional types, it should also be possible to generalize our approach to allow for multiple-dimensional types. One example of this kind of uncertainty would be a case in which countries possessed private information about both their own cost of conflict as well as the relative balance of power. We leave the analysis of this case as a topic for further research.

[26]See Holmstrom & Myerson (1983) for more on this approach.

Appendix

In this appendix, we state and prove two lemmas used to derive the main results in the text and give the proof of Theorem 9. 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.

Our first lemma applies to the case of uncorrelated types. Using the notation in Section 3, we say that types are uncorrelated if there exist distributions F_1 and F_2 on T_1 and T_2, respectively, such that $F(t) = F_1(t_1)F_2(t_2)$ for every $t \in T$.

Lemma 1 *Suppose types are uncorrelated and $T_i = [\underline{t}_i, \bar{t}_i]$. Suppose also that $w_i(t)$ is continuously differentiable in t_1 and t_2. Let G be any crisis bargaining game form and let s^* be any equilibrium of G. Then*

$$\frac{dU_i(t_i)}{dt_i} = \int_{T_j} \pi(t) \frac{\partial w_i}{\partial t_i}(t)\, dF_j(t_j),$$

where $\pi(t) = \pi^g(s^*(t))$.

Proof: Suppose that s^* is an equilibrium of a crisis bargaining game form. Then by Result 1, there exists an incentive-compatible direct mechanism yielding the same outcome as s^*. This direct mechanism is given by $\pi(t) = \pi^g(s^*(t))$ and $v_i(t) = v_i^g(s^*(t))$. Therefore the expected utility of type t_i from equation (2) can be written

$$U_i(t_i) = \int_{T_j} \pi(t_i, t_j) w_i(t_i, t_j) + (1 - \pi(t_i, t_j)) v_i(t_i, t_j)\, dF_j(t_j)$$

and the expected utility of falsely reporting a type $\tilde{t}_i$ is

$$U_i(\tilde{t}_i \mid t_i) = \int_{T_j} \pi(\tilde{t}_i, t_j) w_i(t_i, t_j) + (1 - \pi(\tilde{t}_i, t_j)) v_i(\tilde{t}_i, t_j)\, dF_j(t_j).$$

Clearly, $U_i(t_i) = U_i(t_i \mid t_i)$. Moreover, incentive-compatibility requires that for all $\tilde{t}_i \in T_i$,

$$U_i(t_i \mid t_i) \geq U_i(\tilde{t}_i \mid t_i),$$

which is the same as

$$U_i(t_i) = U_i(t_i \mid t_i) = \max_{\tilde{t}_i \in T_i} U_i(\tilde{t}_i \mid t_i).$$

Now, viewing $U_i(\tilde{t}_i \mid t_i)$ solely as a function of t_i, because w_i is differentiable in t_i, it follows that $U_i(\tilde{t}_i \mid t_i)$ is differentiable in t_i and because w_i is continuously differentiable on a compact set, this derivative is bounded. This implies that $U_i(\tilde{t}_i \mid t_i)$ is absolutely continuous in t_i. Therefore, we can apply the Envelope Theorem (Milgrom & Segal 2002, Theorem 2) to get

$$\begin{aligned} \frac{dU_i(t_i)}{dt_i} &= \frac{\partial U_i(\tilde{t}_i \mid t_i)}{\partial t_i}\Big|_{\tilde{t}_i = t_i} \\ &= \int_{T_j} \pi(\tilde{t}_i, t_j) \frac{\partial w_i}{\partial t_i}(t_i, t_j)\, dF_j(t_j)\Big|_{\tilde{t}_i = t_i}. \end{aligned}$$

Evaluating this expression at $\tilde{t}_i = t_i$ yields the desired result. ∎

Our second lemma deals with the case of correlated types.

Lemma 2 *If f is uniformly distributed on W, then any equilibrium of any crisis bargaining game form that has voluntary agreements must satisfy*

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

and

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

Proof: Suppose there is an equilibrium of a crisis bargaining game form that has voluntary agreements. By Result 2, there exists an incentive-compatible direct mechanism such that $v_i(w) \geq w_i$ for all $w \in W$ such that $\pi(w) \neq 1$ and $i = 1, 2$. This direct mechanism is given by $\pi(w) = \pi^g(s^*(w))$ and $v_i(w) = v_i^g(s^*(w))$.

As f is uniform, the conditional distribution of w_2, given w_1, is uniform on $[0, w_1]$. Thus, the expected utility for country 1 with a type w_1 from equation (2) can be written

$$U_1(w_1) = \frac{1}{1 - w_1} \int_{0}^{1-w_1} \pi(w_1, t) w_1 + (1 - \pi(w_1, t)) v_1(w_1, t)\, dt.$$

and the expected utility of falsely reporting a type $\tilde{w}_1$ is

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

To save on notation, we define

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

and let

$$T_1(w_1) = T_1(w_1 \mid w_1) = \int_0^{1-w_1} \pi(w_1, t)w_1 + (1 - \pi(w_1, t))v_1(w_1, t)\, dt \tag{7}$$

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.

As in the proof of Lemma 1, we note that $U_1(w_1) = U_1(w_1 \mid w_1)$ and observe that incentive-compatibility implies that

$$U_1(w_1) = \max_{\tilde{w}_1 \in [0,1]} U_1(\tilde{w}_1 \mid w_1).$$

From this it follows that

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

and by the Envelope Theorem we have

$$T_1'(w_1) = \frac{\partial T_1(\tilde{w}_1 \mid w_1)}{\partial w_1}\Big|_{\tilde{w}_1 = w_1}. \tag{8}$$

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

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

By evaluating this partial derivative at $\tilde{w}_1 = w_1$ and noting that $v_i(w) \geq w_i$ for all $w \in W$ such that $\pi(w) \neq 1$ and $i = 1, 2$ implies that $v_1(w_1, 1 - w_1) = w_1$ if $\pi(w) \neq 1$, we see that equation (8) 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 (7) 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 $T_1(w_1) = (1-w_1)U_1(w_1)$, we get

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

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

Proof of Theorem 9

Proof: Consider some equilibrium of some crisis bargaining game form that has voluntary agreements. 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 $\pi^*(w_1, w_2) \in \{0, 1\}$.[27] Therefore, we can define a function $h(x)$ such that

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

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 2 that for Γ^*, the following expressions hold:

$$U_1(w_1) = \frac{1}{2}(1+w_1) - \frac{1}{1-w_1}\int_{w_1}^{1}\int_0^{1-x} \pi^*(x,t)\,dt\,dx \tag{9}$$

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

and

$$U_2(w_2) = \frac{1}{2}(1+w_2) - \frac{1}{1-w_2}\int_{w_2}^{1}\int_{0}^{1-y} \pi^*(t,y)\,dt\,dy. \tag{10}$$

In addition, Γ^* must be a solution to the problem of maximizing SW^* given that equations (9) and (10) hold and that the following inequality constraint holds:

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

From Lemma 2, this inequality can be rewritten as follows:

$$\begin{aligned}
SW^* &\geq 2\int_0^1\left[\frac{1}{2}(1-w_1^2) - \int_{w_1}^1\int_0^{1-x}\pi^*(x,t)\,dt\,dx\right]dw_1\\
&\quad + 2\int_0^1\left[\frac{1}{2}(1-w_2^2) - \int_{w_2}^1\int_0^{1-y}\pi^*(t,y)\,dt\,dy\right]dw_2\\
SW^* &\geq \int_0^1(1-w_1^2)\,dw_1 - 2\int_0^1\int_{w_1}^1(1-r-h(r))\,dr\,dw_1\\
&\quad + \int_0^1(1-w_2^2)\,dw_2 - 2\int_0^1\int_{w_2}^1(1-s-h^{-1}(s))\,ds\,dw_2\\
SW^* &\geq \frac{2}{3} - 2\int_0^1\int_w^1(1-r-h(r))\,dr\,dw\\
&\quad + \frac{2}{3} - 2\int_0^1\int_w^1(1-s-h^{-1}(s))\,ds\,dw\\
SW^* &\geq \frac{4}{3} - 2\int_0^1\int_w^1(1-t-h(t)) + (1-t-h^{-1}(t))\,dt\,dw\\
SW^* &\geq \frac{4}{3} - 4\int_0^1\int_w^1[1-t-h(t)]\,dt\,dw.
\end{aligned}$$

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 the constraint $\int_0^1 M(w,h)\,dw \geq 4/3$. 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 binding. This gives the result. ∎

References

Andersen, Thomas Barnebeck, Thomas Harr & Finn Tarp. 2006. "On US politics and IMF lending." *European Economic Review* 50(7):1843–1862.

Baliga, Sandeep & Tomas Sjöström. 2008. Mechanism Design: New Developments. In *The New Palgrave Dictionary of Economics*, ed. Steven N. Durlauf & Lawrence E. Blume. Second edition ed. Palgrave Macmillan.

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

Baron, David P. 2000. "Legislative Organization with Informational Committees." *American Journal of Political Science* 44(3):485–505.

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

Chwe, Michael Suk-Young. 1999. "Minority Voting Rights Can Maximize Majority Welfare." *The American Political Science Review* 93(1):85–97.

Compte, Olivier & Philippe Jehiel. 2009. "Veto Constraint in Mechanism Design: Inefficiency with Correlated Types." *American Economic Journal: Microeconomics* 1(1):182–206.

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

Fey, Mark & Kristopher W. Ramsay. 2008. "Mutual Optimism and Unilateral War." Typescript.

Filson, Darren & Suzanne Werner. 2002. "A Bargaining Model of War and Peace." *American Journal of Political Science* 46:819–838.

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.

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

Holmstrom, Bengt & Roger B. Myerson. 1983. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* 51(6):1799–1819.

Jackson, Matthew O. 2003. Mechanism Theory. In *Encyclopedia of Life Support Systems*, ed. Ulrich Derigs. EOLSS Publishers.

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.

Leventoğlu, Bahar & Ahmer Tarar. 2008. "Does Private Information Lead to Delay or War in Crisis Bargaining?" *International Studies Quarterly* . forthcoming.

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

Milgrom, Paul & 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.

Myerson, Roger B. 2008. Mechanism Design. In *The New Palgrave Dictionary of Economics*, ed. Steven N. Durlauf & Lawrence E. Blume. Second edition ed. Palgrave Macmillan.

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.

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.

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.

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 & 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.

Waltz, Kenneth. 2001. *Man, the State and War: A Theoretical Analysis.* New York: Columbia University Press.

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