Crisis Bargaining without Mediation

Bahar Leventoğlu*

First version: September 2011, This version: March 2012

Abstract

A large body of game-theoretic work examines the process by which uncertainty can lead to inefficient war. In a typical crisis bargaining model, players negotiate according to a pre-specified bargaining protocol and no player has the ability to change the rules of the game. However, when one of the parties has full bargaining power and is able to set the bargaining protocol on her own, the protocol itself becomes an endogenous decision variable. I formulate this problem in a principal-agent framework. I show that both the likelihood of costly war and the exact mechanism that yields it depend on the nature of the informational problem as well as the identity of the informed player.

Word Count: 8,353

*Duke University, Department of Political Science, Durham, NC 27708. E-mail: bahar.leventoglu@duke.edu

1 Introduction

The international relations literature has long argued that some kind of incomplete information, uncertainty and misperception between more-or-less rationally led states make them go into war (e.g. Blainey 1988, Fearon 1995, Jervis 1976, Wittman 1979, Van Evera 1999). In particular, there is a large body of game-theoretic work that examines the process by which uncertainty can lead to inefficient war. One central result is that incentives to misrepresent privately held information play a key role in shaping the behavior of the participants and thereby the likelihood of war and the nature of a peaceful settlement (Fearon 1995, Powell 1996a, 1999, Schultz 1998, Slantchev 2003, Smith and Stam 2006, Wagner 2000).

A typical crisis bargaining model starts with ad hoc assumptions on the bargaining protocol. A protocol or a game form may reflect the institutional setting for a particular crisis bargaining situation. For example, an international body, such as the UN, or any other mediator may impose and enforce a certain protocol among negotiating parties. A model with a specific bargaining protocol provides precise predictions in such cases.

However, if one of the negotiating parties has the power to reject any such mediation and has the ability to set the bargaining protocol on her own, then the protocol itself becomes a decision variable. I refer to this problem as *crisis bargaining without mediation* and ask the following questions: Which bargaining protocol will emerge endogenously in such a crisis situation, what

bargaining outcome will prevail, and what kind of mechanism will lead to inefficient war?

When two states are engaged in crisis bargaining without mediation, unlike the previous models, there is no bargaining protocol given a priori. Instead one of the states (she), which I will refer to as the powerful state, has the ability to set and enforce the bargaining protocol, and her opponent (he) will follow.

Rationality implies that the powerful state will choose a protocol that will benefit her the most. But there are infinitely many protocols to choose from. The main methodological challenge here is to represent all possible protocols in a coherent and analytically tractable way. I overcome this challenge by formulating the problem as a principal-agent problem. The principal-agent framework deals with information and incentive issues between a principal and an agent when the principal has full bargaining power.

Informational asymmetry in crisis bargaining models typically concerns either the cost of war players incur (Fearon 1994, Powell 1996a, Schultz 1999) or the power distribution between them, i.e. the probability of victory in a war (Reed 2003, Smith and Stam 2006, Wagner 2000, Wittman 1979). I study these two special types of private information and refer to a player's private information as that player's type.

These two types of informational problems are fundamentally different. A player's cost of war determines only his or her war payoff, but it does not say anything about the adversary's war payoff. This is known as the private

values case. In contrast, private information about the probability of victory in a war determines one's as well as her adversary's war payoffs. This is known as the common values case.

When it is the adversary that holds some private information, the standard principal-agent framework (Salanie 1999, Bolton and Dewatripont 2005) applies. If the powerful state has private information, then the problem turns into a principal-agent problem with informed principal (Maskin and Tirole 1992), because the powerful state's choice of bargaining protocol may signal her private information to the adversary, which in turn may affect the adversary's payoffs.

Surprisingly, depending on who holds private information, the private values and common values cases produce fundamentally different predictions about the mechanism that yields inefficient fighting. If a player's private information concerns his or her cost of war, then the standard principal-agent framework applies and a risk-return trade-off (Fearon 1995) turns out to be a robust prediction as a cause of war regardless of whether the powerful state or her adversary or both sides hold private information. Accordingly, the powerful state can always avoid war by making a generous offer to her opponent. But she may also prefer to resolve the conflict by making a smaller offer that carries some risk of war.

In contrast, if private information concerns the likelihood of victory, then the identity of the informed party matters. This is because such information determines war payoffs for both sides and the choice of protocol may reveal

information about the power distribution to the adversary when the powerful state is informed. This turns the game into a signaling game where the standard principal-agent framework does not apply. Unlike standard signaling games in which the game form is exogenously given, the contract (or the game form) is endogenously determined in this case. Therefore, the standard framework of signaling games does not apply, either. I apply the principal-agent framework with informed principal (Maskin and Tirole 1992) to this case.

When the powerful state is informed about power distribution, there is neither risk nor a risk-return trade-off. However, war may still break out because risk of war may signal the powerful state's high resolve. This is a novel mechanism through which information asymmetry may cause inefficient fighting. The conflict may also be resolved peacefully where the powerful state prefers not to signal her high resolve. Since no information is transmitted to the adversary, a militarily strong adversary accepts offers that he would not have accepted if he were informed of the power distribution. On the other hand, if it is the adversary that holds private information about power distribution, then war breaks out due to the risk-return trade-off that the powerful state faces. The model predicts that the mechanism for inefficient fighting, the likelihood or war and the nature of a peaceful settlement all depend on the identity of the informed party when private information concerns power distribution.

The principal-agent methodology contrasts with two earlier modeling

practices in the formal international relations literature in an interesting manner.

First, in his seminal work, Fearon (1995) models crisis bargaining process as a take-it-or-leave-it offer game: One of the parties makes an offer, which her adversary can either accept of reject. If the adversary accepts the offer, the game ends peacefully and the parties share the pie accordingly. If he rejects the offer, war breaks out. Although the take-it-or-leave-it game seems to model full bargaining power, it is not a complete description. For example, a bluffing equilibrium (Bueno de Mesquita and Lalman 1992) may arise in a take-it-or-leave-it offer game, but no such outcome prevails in a principal-agent model (See section 5 for an extensive discussion).

Second, as pointed out by Banks (1990) and Fey and Ramsay (2011), the formal crisis bargaining literature hardly agrees on the practice of game-theoretic modeling. For example, who will make the first offer? Will it be a take-it-or-leave-it offer model, or can a second offer be made when the first one is rejected? In the latter case, will one side make all the offers or will it be an alternating offers model? When, in the game, do players make an offer and when do they fight?

Such modeling choices have significant implications on predictions. For example, Fearon (1995) identifies a risk-return trade-off in a take-it-or-leave-it bargaining game with uncertainty about the opponent's resolve. Powell (1996a) generalizes this risk-return trade-off argument in an alternating offers model in which a player may reject an offer to make a counter offer.

Leventoğlu and Tarar (2008) show that a small and rather intuitive change to the timing for when players can engage in war in Powell's model can generate a peaceful outcome with no risk-return trade-off. Then which prediction will prevail?

Banks (1990) and Fey and Ramsay (2011) take a mechanism design approach (Myerson 1979) to find predictions that are robust to changes in the bargaining protocol. Although the mechanism design approach provides the set of robust predictions, it does not tell which prediction will prevail. In contrast, when one of the parties has full bargaining power, the rationality assumption and the principal-agent framework provide unique predictions. For example, Fey and Ramsay (2011) find peace as one of the robust predictions when private information is about cost of war. They show that there always exists a game form in which parties peacefully settle and sharing resources proportional to their respective military strengths is a necessary condition for a game form to always have a peaceful equilibrium. In contrast, I show that, when one of the parties has full bargaining power, Fearon's risk-return trade-off argument prevails as the unique prediction.

I contribute to the literature in several important ways. I identify the problem of crisis bargaining without mediation and adopt the principal-agent framework for its analysis. I formally show that both the likelihood of costly fight and the exact mechanism that yields it depend on the nature of the informational problem as well as the identity of the informed player. When the bargaining protocol is determined endogenously, the principal-agent ap-

proach pins down the exact mechanism that arises endogenously. Thus, I provide new predictions on the nature of costly conflict.

The paper proceeds as follows: Section 2 introduces and studies the crisis bargaining game without mediation under private information with uncertainty regarding the opponent's cost of war (private values). Section 3 studies the same model with uncertainty regarding distribution of power (common values). Section 4 discusses the results and concludes. I defer all technical proofs to an appendix.

2 The Model and the methodology

Two states, D (he) and S (she) have a dispute over a divisible good of size 1. D and S's status quo shares of the disputed good are q and $1-q$, respectively, where $0 \leq q \leq 1$. They can live with status quo, reach a peaceful agreement to reallocate the good between themselves, or they can go to war. If parties go to war, the state that wins the war obtains the entire good. D wins the war with probability p and S wins the war with probability $1-p$. Fighting is costly, D and S pay a cost of $c_D, c_S > 0$, respectively if they go to war. The players are risk neutral. Therefore, the expected payoff from war to D is $EU_D(war) = p1 + (1-p)0 - c_D = p - c_D$ and the expected payoff from war to S is $EU_S(war) = (1-p)1 + p0 - c_S = 1 - p - c_S$. If the two states do not reallocate the good peacefully and do not go to war, then the status quo prevails.

A state is "satisfied" if it receives from its status quo share as much utility as it receives from war and it is "dissatisfied" if it strictly prefers to go to war rather than living with its status quo share (Powell 1996a, 1996b, 1999). At most one state can be dissatisfied. I assume that S is satisfied and D is dissatisfied. That is, $1 - q > 1 - p - c_S$ and $p - c_D > q$ for all values of p, c_S and c_D. Next, I will introduce informational asymmetry concerning the values of p, c_S and c_D.

These assumptions, the information structure and a bargaining protocol for a peaceful resolution of the conflict complete the standard crisis bargaining model. I depart from the literature by not assuming any specific bargaining protocol here. Instead, I assume that S has full bargaining power and the ability to set the bargaining protocol on her own. Therefore, the protocol is determined endogenously in the model.

I discuss the principal-agent approach in the following section before applying it to crisis bargaining without mediation.

2.1 Choosing a bargaining protocol: The principal-agent problem

Rationality implies that S will choose the bargaining protocol that benefits her the most. However, there are infinitely many types of protocols that S can choose from: It can be a take-it-or-leave-it offer (Fearon 1995), an alternating offers game (Powell 1996a, Leventoglu and Tarar 2008); S may

choose to have crisis bargaining as a war of attrition (Fearon 1994), or she may use crisis bargaining with entry (Schultz 1999). If S chooses a multiple offers protocol, she may end the bargaining after a number of offers are rejected, the number of which may be determined randomly or set a priori; S may condition the offers or the identity of the next proposer on the history of the interaction.

The main methodological challenge here is to represent all possible protocols in a coherent and analytically tractable way. We can overcome this challenge by formulating the problem in a principal-agent framework. The principal-agent framework deals with information and incentive issues between a principal and an agent where the principal has full bargaining power.

As in the mechanism design approach (Banks 1990; Fey and Ramsay 2011), in a principal-agent problem, it is sufficient for the principal to focus on the following type of bargaining protocol, which is referred to as a contract in the contract theory literature (Salanie 1999, Bolton and Dewatripont 2005): A contract asks for players' private information and determines the outcome according to players' reports. A contract is individually rational for a player if the contract guarantees him a payoff that is at least as high as his outside option, which corresponds to his or her war payoff in this model. A contract is incentive compatible if the player finds it optimal to report his private information truthfully. Then rationality implies that the principal chooses an individually rational and incentive compatible contract (Myerson and Satterthwaite, 1983) that provides her with the largest expected payoff.

Implementation of this general methodology depends on the informational problem at hand.

First, the nature of information matters. The informational assumptions of the crisis bargaining literature fall into two broad categories: When the opponent's private information concerns only his own payoff, it is called private values. For example, private information about cost of war is in this category. In this case, both players know the true value of p, but each player privately knows its cost of war. Then D's private information does not affect S's war payoff $1-p-c_S$ and vice versa. In contrast, when a player privately knows the true value of p, then both players' war payoffs are determined by this privately held information. This case is called common values.

Second, the identity of the player that holds private information matters. If D has private information, the nature of the information does not matter for S. Regardless of whether there is a private values or a common values situation, S has no information that might affect D's payoffs, so her choice of the contract will not have any impact on the set of individually rational and incentive compatible contracts. This is a standard principal-agent problem (Salanie 1999, Bolton and Dewatripont 2005).

However, if S holds private information, her choice of contract can potentially signal her private information to D, and thereby affect the terms of a peaceful bargain. In effect, the problem turns into a signaling game and the contract that S chooses becomes a signal for D. This is the principal-agent problem with informed principal (Maskin and Tirole 1990, 1992). The

optimal contract may separate different types of principals where different types of principals offer different contracts and have no incentives to mimic each other. Different types of S may also find it optimal to pool together by offering the same contract. In that case, the pooling contract must provide every type of S with a payoff higher than her payoff from the optimal separating contract. Pooling may benefit all types of S because no information is revealed in a pooling contract and that relaxes the individual rationality constraint for D.

I will discuss these issues in more detail below.

3 Private Values: Cost of War

Assume that D's cost of war is his private information. It is common knowledge that it is either low $c_D = c_l$ with probability π, or high $c_D = c_h > c_l$ with probability $1 - \pi$. Both types of D are dissatisfied with the status quo. That is, $p - c_l > p - c_h > q$. I will refer to D with c_h as "high-cost" or "low-resolve" type and to D with c_l as "low-cost" or "high-resolve" type.

For the time being, I will assume that S's cost of war is common knowledge. However, I will argue later that the analysis and the results remain the same when S's cost of war is her private information. Therefore, the identity of the informed player does not affect either the methodology or the results in the case of private values.

I will explicitly discuss below how to set up individually rational and

incentive compatible contracts. A familiar reader may skip the following section.

3.1 Bargaining Protocols

Since S is able to set the bargaining protocol, she will choose a game form that provides her with the highest equilibrium payoff. If D rejects to play the game form S has selected, the players end up in war and receive their war payoffs.

A bargaining protocol or a crisis bargaining game $< A_S, A_D, g >$ is a game form that specifies the action spaces A_S and A_D for S and D, respectively and the outcome function $g(a) = (\pi^g(a), t^g(a))$ of $a \in A_S \times A_D$, where $\pi^g(a)$ is the probability that the players go into war, and $t^g(a)$ is the share D receives in case of a peaceful settlement. The bargaining protocol $< A_S, A_D, g >$ induces a game between the players, given the players' costs are private information and drawn from a commonly known joint distribution. A strategy profile constitutes a Bayesian Nash equilibrium if each type of a player is playing a best response to the strategies used by the other players.

There exist infinitely many forms of bargaining protocols that S can choose from. Contract theory tells us that we only need to consider the following type of bargaining protocols or contracts: D submits a report $\hat{c}$ about his type. If he reports his type as $\hat{c} = c_h$, then the players go into war with probability α_h, and D's share is revised to t_h peacefully with probability $1 - \alpha_h$. If D reports his type as $\hat{c} = c_l$, then the players go into war

with probability α_l, and D's share is revised to t_l peacefully with probability $1-\alpha_l$. Therefore, a contract is characterized by four numbers $\{\alpha_h, t_h, \alpha_l, t_l\}$. The contract also satisfies individual rationality and incentive compatibility constraints, as described below.

Since D can unilaterally guarantee his war payoff, individual rationality requires that each type of D is offered at least his outside option in a peaceful deal:

$$D_l - IR : t_l \geq p - c_l, \text{ and} \tag{1}$$

$$D_h - IR : t_h \geq p - c_h \tag{2}$$

This condition also ensures that it is ex ante optimal for each type of D to accept the terms of the contract. That is, $t_\tau \geq p - c_\tau$ implies

$$\alpha_\tau(p - c_\tau) + (1 - \alpha_\tau)t_\tau \geq p - c_\tau$$

for each $\tau \in \{h, l\}$.

Incentive compatibility ensures that D will report his type truthfully:

$$D_l - IC : \alpha_l(p - c_l) + (1 - \alpha_l)t_l \geq \alpha_h(p - c_l) + (1 - \alpha_h)\max\{t_h, p - c_l\}$$

$$D_h - IC : \alpha_h(p - c_h) + (1 - \alpha_h)t_h \geq \alpha_l(p - c_h) + (1 - \alpha_l)\max\{t_l, p - c_h\}.$$

The left hand side of $D_l - IC$ is the expected payoff of D with $c_D = c_l$ in case

he reports c_D truthfully and the right hand side is his payoff if he reports $c_D = c_h$. In the latter case, war will break out and he will collect his war payoff $p - c_l$ with probability α_h, and he will be offered t_h with probability $1 - \alpha_h$. If he is offered t_h, he will accept the offer if t_h is at least as good as his war payoff, i.e. $t_h \geq p - c_l$. So, his payoff will be the maximum of t_h and $p - c_l$ in that case.

Note that $D_l - IR$ ensures that $t_l \geq p - c_l > p - c_h$ so that $t_l = \max\{t_l, p - c_h\}$ in $D_h - IC$.

The following result by Fey and Ramsay (2011) obtains by invoking the revelation principle (Myerson 1979): Consider any Bayesian Nash equilibrium of any bargaining protocol. Then there exists an individually rational and incentive compatible contract yielding the same outcome. Thus, S needs to consider only the set of individually rational and incentive compatible contracts.

3.2 The Principal-Agent Problem

The next step is to formulate S's problem. When S offers an individually rational and incentive compatible contract $\{\alpha_h, t_h, \alpha_l, t_l\}$, both types of D will accept the contract by individual rationality and report their types truthfully by incentive compatibility.

If D turns out to have $c_D = c_\tau$, $\tau \in \{h, l\}$, then the two states will go to war with probability α_τ, and they will reach the peaceful settlement t_τ with probability $1 - \alpha_\tau$ where S will receive $1 - t_\tau$. Here, S will achieve an

expected payoff of

$$\alpha_\tau(1-p-c_S)+(1-\alpha_\tau)(1-t_\tau)$$

S does not know D's type when she offers the contract $\{\alpha_h, t_h, \alpha_l, t_l\}$ but she knows that D is a high-resolve type with probability π, so S's expected payoff from offering $\{\alpha_h, t_h, \alpha_l, t_l\}$ is given by

$$\begin{aligned} V(\{\alpha_h, t_h, \alpha_l, t_l\}) = \pi\left[\alpha_l(1-p-c_S)+(1-\alpha_l)(1-t_l)\right] \\ +(1-\pi)\left[\alpha_h(1-p-c_S)+(1-\alpha_h)(1-t_h)\right] \end{aligned}$$

Given that S only needs to consider the set of individually rational and incentive compatible contracts, she chooses a contract that solves the following maximization problem

$$\max_{\{\alpha_h, t_h, \alpha_l, t_l\}} V(\{\alpha_h, t_h, \alpha_l, t_l\}) \tag{P1}$$

subject to

$$D_l - IR,\ D_h - IR,\ D_l - IC,\ D_h - IC$$

and $\alpha_h, \alpha_l, t_h, t_l \in [0,1]$

where $\alpha_h, \alpha_l, t_h, t_l \in [0,1]$ are the usual feasibility constraints on probabilities and shares.

I discuss the solution of this problem in the following section and defer

the full analysis to the appendix.

3.3 The optimal bargaining protocol

The first result below states that only two of the four constraints will restrict S's decision:

Result 1: The individual rationality constraint for the high-resolve type D and the incentive compatibility constraint for the low-resolve type D are the only binding constraints at the optimal solution. We can ignore the individual rationality constraint for the low-resolve type D and the incentive compatibility constraint for the high resolve type D.

If S knew that D were of type τ, then S could make a take-it-or-leave-it offer $t_\tau = p - c_\tau$, which D of type $\tau \in \{h, l\}$ would accept. In that case, the payoff of the high-resolve type D would be higher: $p - c_l > p - c_h$. Thus, the low-resolve type D has the incentive to mimic the high-resolve type. Result 1 states that the optimal contract makes the low-resolve type D just indifferent between revealing his type and mimicking the high-resolve type D. The optimal contract achieves this by making an offer of $p - c_l$ to the high-resolve type D that may also carry some risk of war, and by making an offer that is potentially larger than $p - c_h$ to the low-resolve type D. As a result, the individual rationality constraint for the high-resolve type D is binding and the individual rationality constraint for the low-resolve type D may be slack.

A slack individual rationality constraint for the low-resolve type D means

that the low-resolve type may be made an offer that is larger than his war payoff $p - c_h$, which he accepts. That gain is due to the private information he holds. This is known as *information rent* in contract theory (Salanie, 1999). S may find it optimal to pay some rent to the low resolve type D to extract his private information.

In contrast, a binding individual rationality constraint for the high-resolve type D implies that the high-resolve type is always offered his war payoff $p-c_l$, so he never collects a payoff higher than his war payoff under the optimal contract.

The next result states that inefficient fighting may occur only when S faces a high-resolve type D.

Result 2: S does not fight with a low-resolve type D. War breaks out with positive probability only if S faces a high-resolve type D.

Incentive compatibility of the contract ensures that D will reveal his type to S. When D reveals himself as a low-resolve type, S does not fight with D. If D reveals himself as a high-resolve type, S offers him his war payoff (Result 1). If S wants to separate the low-resolve type from the high-resolve type, then she has to fight with the high-resolve type with positive probability. Otherwise, the low-resolve type would mimic the high resolve type to secure a large offer from S.

Finally, the following result tells how S will solve her risk-return trade-off.

Result 3: Let $\pi^* = \frac{c_h - c_l}{c_h + c_S} \in (0, 1)$. If $\pi \geq \pi^*$, then S offers $p - c_l$ to both

types of D, and both types of D accept the offer. If $\pi < \pi^*$, then if D reports his type as high-resolve, then S fights with D with probability 1. If D reports his type as low-resolve, then S does not fight with D and offers him $p - c_h$, which he accepts.

The optimal contract is given by the following numbers:

$$\text{if } \pi \geq \pi^* \text{ then } \alpha_l = \alpha_h = 0 \text{ and } t_h = t_l = p - c_l$$

$$\text{if } \pi < \pi^* \text{ then } \alpha_l = 1, \alpha_h = 0, \, t_l = p - c_l \text{ and } t_h = p - c_h$$

This is effectively equivalent to Fearon's optimal take-it-or-leave-it offer. If S is sufficiently confident that she is likely to face a high-resolve type D, $\pi \geq \pi^*$, she solves her risk-return trade-off by offering $p - c_l$ to D and thereby avoids war. Otherwise, she takes the risk of war against the high-resolve type by making a low offer.[1]

In summary, Fearon's risk-return trade-off argument is a robust prediction when parties negotiate without mediation under private values and S sets the bargaining protocol. As I explain below, this result is independent of S's private information so that the risk-return trade-off prediction is robust to the identity of the player that holds private information under private values.

[1]It is easily verified in this and following problems that the optimal contracts are individually rational for S as well.

3.4 Private Values with Informed Principal

S's cost of war may also be her private information. In this case, S's private information does not affect D's payoff, and this constitutes a private values case.

Maskin and Tirole (1990) develop a method to solve the problem of an informed principal in the case of private values. They assume that some types of principals may violate the individual rationality constraint for the agent, however the individual rationality constraint has to hold in expectations (see their problem F_*^i on page 392). In contrast, a fundamental assumption in crisis bargaining is that no player can be forced to accept any deal that is worse than its war payoff. That is, the individual rationality constraint cannot be violated by any type of principal.[2] Therefore, their approach reduces to the standard principal-agent framework in my crisis bargaining model.

Since S's private information does not affect D's payoff, it does not affect the set of individually rational and incentive compatible contracts, either. Thus, the analysis remains the same and the identity of the player that holds private information does not matter when the informational problem is that of private values.

[2] Technically, this means that the only feasible value for r^i is zero in their problems F_*^i and V_I^i on page 392.

4 Common Values: Distribution of Power

In this section, I study the informational problem that concerns distribution of power between players. Assume that c_D and c_S are now common knowledge, but p is equal to p_h with probability π and to $p_l < p_h$ with probability $1 - \pi$. Although the distribution of p is common knowledge, only one of the players knows the true value of p. The identity of the informed player matters in this case. First, I consider the scenario that D holds private information.

4.1 Common Values with Uninformed Principal

Assume that D privately knows the true value of p. I will refer to D with $p = p_h$ as the high-resolve type and with $p = p_l$ as the low-resolve type. This case is similar to the previous one because S's choice of contract does not transmit information from S to D, and so S's problem is set up as in (P1). S chooses an individually rational and incentive compatible contract $\{\alpha_h, t_h, \alpha_l, t_l\}$ that maximizes her expected payoff among all individually rational and incentive compatible contracts. Her problem is formulated as follows:

$$\max_{\alpha_l, t_l, \alpha_h, t_h} V(\alpha_l, t_l, \alpha_h, t_h) = \pi \left[\alpha_h(1 - p_h - c_S) + (1 - \alpha_h)(1 - t_h)\right] \qquad \text{(P2)}$$
$$+ (1 - \pi) \left[\alpha_l(1 - p_l - c_S) + (1 - \alpha_l)(1 - t_l)\right]$$

subject to

$$D_l - IR : t_l \geq p_l - c_D,$$

$$D_h - IR : t_h \geq p_h - c_D,$$

$$D_l - IC : \alpha_l(p_l - c_D) + (1 - \alpha_l)t_l \geq \alpha_h(p_l - c_D) + (1 - \alpha_h)\max\{t_h, p_l - c_D\}$$

$$D_h - IC : \alpha_h(p_h - c_D) + (1 - \alpha_h)t_h \geq \alpha_l(p_h - c_D) + (1 - \alpha_l)\max\{t_l, p_h - c_D\}$$

and the usual feasibility constraints $\alpha_i \in [0, 1]$ and $t_i \in [0, 1]$.

I will discuss the solution to (P2) below.

4.1.1 The best contract

The solution to (P2) and its interpretation mirror those of (P1).

Result 4: The individual rationality constraint for the high-resolve type D and the incentive compatibility constraint for the low-resolve type D are binding at the optimal solution. We can ignore the individual rationality constraint for the low-resolve type D and the incentive compatibility constraint for the high-resolve type D.

Result 5: S does not fight with a low-resolve type D. War breaks out with positive probability only if S faces a high-resolve type D.

Result 6: Let $\pi^{**} = \frac{p_h - p_l}{(p_h - p_l) + (c_S + c_D)} \in (0, 1)$. If $\pi \geq \pi^{**}$, then S offers $p_h - c_D$, which both types of D accept. If $\pi < \pi^{**}$, then S does not fight with the low-resolve type D, and offers him $p_l - c_D$, which he accepts; S fights with the high-resolve type D.

The optimal contract is characterized by

$$\text{if } \pi \geq \pi^{**} \text{ then } \alpha_l = \alpha_h = 0 \text{ and } t_h = t_l = p_h - c_D$$

$$\text{if } \pi < \pi^{**} \text{ then } \alpha_l = 0, \alpha_h = 1,\ t_l = p_l - c_D \text{ and } t_h = p_h - c_D$$

This contract is effectively equivalent to a take-it-or-leave-it offer. If S is sufficiently confident that she is likely to face a high-resolve type D, $\pi \geq \pi^{**}$, she solves her risk-return trade-off by offering $p_l - c_D$ to D and thereby avoids war. Otherwise, she takes the risk of war against the high-resolve type D by making a low offer. Fearon's prediction of risk-return trade-off arises in this case as well.

4.2 Common values with Informed Principal

Assume that S privately knows the true value of p. Recall that p is the probability with which D wins the war. Thus, I will refer to S with $p = p_h$ as the low-resolve type and to S with $p = p_l$ as the high-resolve type.

In this case, S's choice of contract can signal her private information to D. Such information transmission changes D's payoffs so S's problem turns into a signaling game without a pre-specified game form. Neither the standard principal-agent framework nor the standard signaling games applies in this case. The solution of S's problem obtains by applying Maskin and Tirole (1992). The first step is to find an optimal separating contract. The second step looks for a pooling contract which is better than the optimal separating

contract for both types of S.

4.2.1 Optimal Separating Contracts

Let the low-resolve type S choose the contract (α_h, t_h) and the high-resolve type S choose the contract (α_l, t_l) in a separating equilibrium. Since D will find out about the true value of p when he observes S's choice of contract, each of these contracts must satisfy the associated individual rationality constraint for D:

$$t_\tau \geq p_\tau - c_D \text{ for } \tau \in \{h, l\}.$$

Since D does not hold private information, S is not constrained by incentive compatibility for D. However, each type of S must make sure that, given the contract for the other type of S, she chooses a contract that will separate her from the other type. That induces an incentive compatibility constraint for the other type. For example, a low-resolve type S collects an expected payoff of

$$\alpha_h(1 - p_h - c_S) + (1 - \alpha_h)(1 - t_h)$$

from her own contract and

$$\alpha_l(1 - p_h - c_S) + (1 - \alpha_l)(1 - t_l)$$

from choosing the contract of the high-resolve type. The former must be at least as big as the latter for the low-resolve type not to imitate the high-

resolve type by choosing the contract of the high-resolve type.

A separating contract is optimal if, given the expected payoff for a type, it maximizes the expected payoff for the other type. Thus, a pair of optimal separating contracts $(\alpha_h^{sep}, t_h^{sep})$ and $(\alpha_l^{sep}, t_l^{sep})$ is a solution to the following two problems:

The problem of the high-resolve type S: Given $(\alpha_h^{sep}, t_h^{sep})$,

$$(\alpha_l^{sep}, t_l^{sep}) \text{ solves } \max_{(\alpha_l, t_l)} \ \alpha_l(1 - p_l - c_S) + (1 - \alpha_l)(1 - t_l)$$

subject to

$$D_l - IR : t_l \geq p_l - c_D$$

$$S_h - IC : \alpha_h^{sep}(1 - p_h - c_S) + (1 - \alpha_h^{sep})(1 - t_h^{sep}) \geq$$

$$\alpha_l(1 - p_h - c_S) + (1 - \alpha_l)(1 - t_l)$$

The problem of the low-resolve type S: Given $(\alpha_l^{sep}, t_l^{sep})$,

$$(\alpha_h^{sep}, t_h^{sep}) \text{ solves } \max_{(\alpha_h, t_h)} \ \alpha_h(1 - p_h - c_S) + (1 - \alpha_h)(1 - t_h)$$

subject to

$$D_h - IR : t_h \geq p_l - c_D$$

$$S_l - IC : \alpha_l^{sep}(1 - p_l - c_S) + (1 - \alpha_l^{sep})(1 - t_l^{sep}) \geq$$

$$\alpha_h(1 - p_l - c_S) + (1 - \alpha_h)(1 - t_h)$$

$S_h - IC$ ensures that the "low resolve" type S will not have an incentive

to choose the contract of the high resolve type. This is the incentive compatibility constraint that the "high resolve" type S is subject to when she is choosing her contract. Similarly, the low resolve type S is constrained by $S_l - IC$ when she chooses her own contract.

The following result states that only the high resolve type S is constrained by incentive compatibility at the optimal separating contract.

Result 7: $S_l - IC$ is slack and $S_h - IC$ is binding at the optimal solution.

That is, at the optimal solution, the high resolve type S separates herself from the low resolve type by choosing a contract that does not give the low resolve type S any incentive to mimic her. Therefore, the high resolve type S pays the cost of separation at the optimal solution.

The following summarizes the unique optimal separating contract.

Result 8: The pair of optimal separating contracts is unique and is given by $(\alpha_h^{sep} = 0, t_h^{sep} = p_h - c_D)$ and $(\alpha_l^{sep} = \pi^{**}, t_l^{sep} = p_l - c_D)$.

Substantive interpretation of this finding is quite interesting. The high resolve type S separates herself from the low resolve type by being aggressive. She commits to fighting with positive probability in order to secure a bigger share of the pie in peace time.

The expected payoffs for the two types of S are given as follows.

$$V_h^{sep} = 1 - p_h + c_D$$

$$V_l^{sep} = 1 - p_l + c_D - \pi^{**}(c_D + c_S)$$

These payoffs set the lower bound for what each type of S can achieve in crisis bargaining with D. The next step searches for contracts that provide each type of S with a better payoff.

4.2.2 The best contract

The second step of Maskin and Tirole (1992) derives several constraints to characterize the set of optimal contracts. Now, I derive these constraints for the crisis bargaining game.

Let V_h^* and V_l^* be the expected payoffs from an optimal pair of contracts (α_h^*, t_h^*) and (α_l^*, t_l^*). For (α_h^*, t_h^*) and (α_l^*, t_l^*) to be optimal, they must provide payoffs that are at least as good as the payoffs from the optimal separating contracts, that is

$$(Better) : V_h^* \geq V_h^{sep} \text{ and } V_l^* \geq V_l^{sep}$$

must hold. Also, $S_h - IC$ and $S_l - IC$ must hold with (α_h^*, t_h^*) and (α_l^*, t_l^*).

Finally, an individual rationality constraint for D must hold. In a crisis bargaining game, no player can be forced to accept a deal that is worse than his or her expected war payoff. This fundamental assumption requires some modification in Maskin and Tirole's individual rationality constraint.

In particular, Maskin and Tirole (1992) assume that once players sign the contract, they commit to the terms of the contract. That is, in Maskin and Tirole (1992), both types of S offer the same pair of contracts $\{(\alpha_h^*, t_h^*),(\alpha_l^*, t_l^*)\}$,

and if D agrees, and only after he agrees, S reveals her type and both D and S have to go with the terms of the contract associated with S's type. This induces the following individual rationality constraint for D:

$$D - IR^{MaskinTirole} : \pi \left[\alpha_l^*(p_l - c_D) + (1 - \alpha_l^*)t_l^*\right] + (1 - \pi)\left[\alpha_h^*(p_h - c_D) + (1 - \alpha_h^*)t_h^*\right]$$
$$\geq \pi(p_l - c_D) + (1 - \pi)(p_h - c_D)$$

When D decides to accept or reject the contract pair, he does not know the type of S he faces. He will learn S's type only after accepting the contract pair. Then both players will play the contract associated with S's type. The left hand side of $D - IR^{MaskinTirole}$ is D's expected payoff from accepting and committing to the terms of these contracts. Alternatively, he can reject the offer and fight, which provides him with the expected payoff on the right hand side. Applying Maskin and Tirole (1992) requires the constraints $(Better)$, $S_h - IC$, $S_l - IC$ and $D - IR^{MaskinTirole}$.

However, notice that $D - IR^{MaskinTirole}$ does not ensure $t_\tau \geq p_\tau - c_D$. In Maskin and Tirole (1992), D commits to living with t_τ even if $t_\tau < p_\tau - c_D$. In other words, the individual rationality constraint is relaxed and that is how (α_h^*, t_h^*) and (α_l^*, t_l^*) can potentially provide better payoffs than the optimal separating contracts in Maskin and Tirole (1992).

In our crisis bargaining model, once D finds out about the true value of p, he can not be forced to live with $t_\tau < p_\tau - c_D$. So, if $(\alpha_h^*, t_h^*) \neq (\alpha_l^*, t_l^*)$, then D will learn S's type and S will not be able to force D to accept anything

less than his war payoff. Then $\{(\alpha_h^*, t_h^*), (\alpha_l^*, t_l^*)\}$ will constitute a separating equilibrium with the usual individual rationality constraints for D. Since the optimal separating contract is already characterized with these rationality constraints and it is unique, $\{(\alpha_h^*, t_h^*), (\alpha_l^*, t_l^*)\}$ can not be strictly better for S than the optimal separating contract.

This implies that the only other potentially better contract is a pooling contract that does not reveal any information, that is $(\alpha_h^*, t_h^*) = (\alpha_l^*, t_l^*) = (\alpha^*, t^*)$. Then $S_h - IC$ and $S_l - IC$ trivially hold and the individual rationality constraint becomes

$$D - IR^{Pooling} : t^* \geq \pi(p_l - c_D) + (1 - \pi)(p_h - c_D)$$

Also

$$V_h^* = \alpha^*(1 - p_h - c_S) + (1 - \alpha^*)(1 - t^*) \text{ and}$$
$$V_l^* = \alpha^*(1 - p_l - c_S) + (1 - \alpha^*)(1 - t^*)$$

I summarize this result in the following:

Result 9: If the pair of optimal separating contracts is not the best choice for S, then the optimal contract is a pooling contract and it satisfies $D - IR^{Pooling}$, $V_h^* \geq V_h^{sep}$ and $V_l^* \geq V_l^{sep}$.

The constraints $V_h^* \geq V_h^{sep}$ and $V_l^* \geq V_l^{sep}$ are required for optimality. If either of them is violated, then the associated type will have an incentive to

separate herself by offering the optimal separating equilibrium contract.

The individual rationality constraint $D-IR^{Pooling}$ has an important substantive interpretation. In a separating contract, the individual rationality constraint holds for each type: $t_h^{sep} \geq p_h - c_D$ and $t_l^{sep} \geq p_l - c_D$, which also imply $D-IR^{Pooling}$. However, $D-IR^{Pooling}$ does not imply the former. In other words, $D-IR^{Pooling}$ relaxes the individual rationality constraint imposed by a separating contract on S. This is because when D is offered a pooling contract, he does not find out about the true value of p so his individual rationality constraint must only hold in expectation. This relaxation opens up the possibility of better bargains for S.

Then the solution for the best contract boils down to comparing the optimal pooling contract with the pair of optimal separating contracts. Optimality and $D-IR^{Pooling}$ imply that

$$\alpha^* = 0 \text{ and } t^* = \pi(p_l - c_D) + (1-\pi)(p_h - c_D)$$

must hold with an optimal pooling contract. That is, there is no fighting at the optimal pooling contract, which yields

$$V_h^* = V_l^* = 1 - t^* = 1 - p_h + c_D + \pi(p_h - p_l)$$

Finally, $V_h^* \geq V_h^{sep}$ holds and

$$V_l^* \geq V_l^{sep} \Leftrightarrow \pi \geq \pi^{**} = \frac{p_h - p_l}{(p_h - p_l) + (c_D + c_S)}$$

so the following main result follows:

Theorem 1 *Assume that S privately knows the true value of p.*

(i) If $\pi \geq \pi^{**}$, then the conflict is resolved efficiently at the optimal pooling equilibrium: Both types of S offer $t^* = \pi(p_l - c_D) + (1-\pi)(p_h - c_D)$ without risking war and D accepts the offer.

(ii) If $\pi < \pi^{**}$, then the conflict is resolved inefficiently at the optimal separating equilibrium: The low-resolve type S offers $p_h - c_D$ and avoids war; high-resolve type S fights with probability π^{**} and offers $p_l - c_D$ otherwise. Thus, the ex ante probability of war is $\pi\pi^{**}$.

Now let us contrast the optimal contract of an informed principal to that of an uninformed principal. The following theorem summarizes Result 6 for the uninformed principal case:

Theorem 2 *Assume that D privately knows the true value of p.*

(i) If $\pi \geq \pi^{**}$, then the conflict is resolved efficiently: S offers $p_h - c_D$ without risking war and D accepts the offer.

(ii) If $\pi < \pi^{**}$, then the conflict is resolved inefficiently: S offers $p_l - c_D$, which the low-resolve type D accepts and the high-resolve type D rejects and fights. So, the ex ante probability of war is π.

I discuss the substantive implications of these results in the next section.

5 Discussion

Risk-return trade-off vs Separating-Pooling trade-off

When one of the parties has full bargaining power, inefficient war may break out for two reasons: The first one is Fearon's celebrated risk-return trade-off explanation (Fearon, 1995). Accordingly, the powerful party may risk war in order to obtain a bigger share of the pie in a peaceful settlement. The findings above imply that this prediction is robust to the nature of informational problem when the powerful state is not informed. That is, regardless of whether the informational asymmetry is about parties' individual costs of war or power distribution between them, war may break out as a consequence of a risk-return trade-off calculation by the powerful party when she is not informed.

In contrast, the analysis predicts a novel and different type of trade-off when the powerful state holds private information regarding power distribution. If the powerful state knows that power distribution favors her, then she can signal her high-resolve by aggression. That is, in order to separate herself from a low-resolve type, she can make an offer that carries risk of war. That can happen when the ex ante probability of her being a high-resolve type is low. In this case, a low-resolve type avoids war by offering her adversary a high payoff of $p_h - c_D$. The high-resolve type fights with her adversary with positive probability, otherwise offers him a low payoff of $p_l - c_D$ for a peaceful settlement and the adversary accepts.

The powerful state knows the type of her adversary, and so she does not face any risk regarding her adversary's type, nor does she face any risk-return trade-off. Instead, she makes use of war to signal her high resolve. This is a novel prediction of the model and it offers a substantially different mechanism for inefficient fighting based on signaling.

When the probability of being a high-resolve type is high for the powerful state, then she can do better by avoiding risk of war and pooling with the low-resolve type. In this case, D does not find out about the true value of p after receiving a "pooling" offer. Thus, both types of S make a moderate offer between $p_l - c_D$ and $p_h - c_D$ and D's individual rationality constraint only holds in expectations.

Regardless of whether D or S knows the true value of p, they settle peacefully if $\pi \geq \pi^{**}$ and inefficient fighting occurs if $\pi < \pi^{**}$. In the former case, if S knows the true value of p, she can convince D for a peaceful settlement with a smaller offer. In the latter case, the ex ante probability of fighting is smaller when S is the informed player. In other words, both the likelihood of war and the players' shares in a peaceful settlement depend on the identity of the informed player.

Other types of rational behavior, such as bluffing (Bueno de Mesquita and Lalman 1992), can also explain war. In a bluffing equilibrium, a low resolve type may rationally imitate a high resolve type (see below). Such bluffing may increase the likelihood of war as well as the low-resolve type's share from a peaceful settlement. However, my findings imply that such rational

behavior may emerge only if neither party possesses full bargaining power.

Take-it-or-leave-it offer

In his seminal paper, Fearon (1995) models crisis bargaining process as a take-it-or-leave-it offer game: One of the parties makes an offer, which her adversary can either accept of reject. If the adversary accepts the offer, the game ends peacefully and the parties share the pie accordingly. If he rejects the offer, war breaks out.

Although this take-it-or-leave-it offer game provides most of the ingredients for modeling full bargaining power, it is still not a complete description. For example, the following bluffing equilibrium may also arise in a crisis bargaining game with take-it-or-leave-it offer (I provide the equilibrium analysis in the appendix): Suppose that the party who makes the take-it-or-leave-it offer knows about the power distribution and her adversary cannot observe this private information. In equilibrium, the high-resolve type always makes a low offer. The low resolve type bluffs by making the same low offer with positive probability and makes a higher offer otherwise. The adversary accepts the higher offer. However, when he receives the low offer, he knows that it may be a bluff and he fights with positive probability. There may be many bluffing equilibria described by the probability of bluffing. There also exists a separating equilibrium without bluffing.

Therefore, in contrast to the principal-agent approach, in a take-it-or leave-it crisis bargaining game, one needs to make additional assumptions to

select from the set of multiple equilibria.[3]

Mechanism Design vs Principal-Agent Framework

Banks (1990) and Fey and Ramsay (2011) address the following question: Given that there are many game forms and bargaining protocols, which predictions are robust to variations in the underlying game structure? These two works adopt a mechanism design approach and provide "game-free" results without reference to a specific game form.

Their question is fundamentally different from the question I address in this paper. In particular, mechanism design theory does not provide an answer for the question about which crisis bargaining game is going to be played. Moreover, mechanism design theory does not tell which prediction will survive when there are multiple robust predictions regarding the outcome. When one of the players can set the rules of the crisis bargaining game, the bargaining protocol itself becomes endogenous. In this case, the rationality assumption provides unique answers for both: The player chooses the protocol that benefits her the most, which induces the best outcome for her.

To illustrate, consider Fey and Ramsay's findings that if negotiating parties know each other's military strength but each party privately knows its cost of war, there always exists a game form in which parties peacefully settle (Proposition 2, Fey and Ramsay 2011) and sharing resources proportional to

[3]Cho and Kreps' (1987) intuitive criterion does not eliminate any bluffing equilibrium. See Appendix.

their military strengths is a necessary condition for a game form to always have a peaceful equilibrium (Proposition 3, Fey and Ramsay 2011). That is, peace is a robust prediction according to Fey and Ramsay (2011).

But this does not mean that peace will prevail. For example, consider two countries D and S engaged in a crisis bargaining over a pie of size 20. If they fight, the victor is determined with equal probability and obtains the entire pie. They also pay a cost for fighting. D's cost of war is either 5 or 1 with equal probability and S's cost of war is 2. D knows his cost but S cannot observe it and S's cost is commonly known.

If the players agree to play the game form with a peaceful equilibrium, each collects a payoff of 10. If they fight, the expected payoffs of a high-cost (low-resolve type) D, a low-cost (high-resolve type) D and S are given by 5, 9 and 8 respectively.

If S has the power to choose the bargaining protocol, then she can make a take-it-or-leave it offer of 6 to D. A high-cost D accepts this offer since it is higher than his expected war payoff of 5, and S collects a payoff of 14 (= 20 - 6). On the other hand, a low-cost D rejects it, because it is lower than his war payoff of 8, and S collects her war payoff of 8 as a result. S's expected payoff from this risky deal is 0.5x14 + 0.5x8 = 11 which is greater than her peace payoff of 10. Thus, S would risk war if she could make that offer.

In other words, existence of a game form with a peaceful equilibrium, even when its existence is robustly predicted as in Fey and Ramsay (2011), does not mean that peace will prevail. When one of the players has full bargaining

power, the principal-agent approach predicts the bargaining protocol that emerges endogenously and its outcome.

When D also knows something about p

The true value of p may be determined by private information of both S and D (Fey and Ramsay 2011). Maskin and Tirole (1992) can be applied to this case as well. This requires two types of changes.

The first one concerns D's incentives to reveal his private information, so the associated problems are appended with incentive compatibility constraints for D. In the problem of optimal separating contract, D will learn S's type. Then the incentive compatibility constraint will hold for each type of D given that he knows S's type. In contrast, in a pooling contract, D will not learn S's type so the incentive compatibility constraint will only hold in expectations for each type of D.

The second concerns S's payoffs from contracts. Since S will not learn D's type until after offering a contract, S's payoff will be in expectations.

I conjecture that both risk-return trade-off and separating-pooling trade-off will play a role in inefficient fighting in this case. I leave this for future research.

6 Conclusion

The formal international conflict literature makes, but does not explicitly state, an important assumption: Players negotiate according to a pre-specified

bargaining protocol in a crisis bargaining game and no player has the ability to change the rules of the game. This is a harmless assumption if we are ready to assume that a third party, e.g. a mediator, enforces the bargaining protocol. However, this assumption is hardly satisfied in crisis situations where one of the parties has the ability to refuse any kind of mediation and set the bargaining protocol on her own. In this paper, I identify the problem of crisis bargaining without mediation and formulate it in a principal-agent framework.

If one party in a crisis bargaining situation has full bargaining power, then she chooses the bargaining protocol that would benefit her the most. However, there are infinitely many forms of bargaining protocols and moreover a particular protocol may give rise to multiple equilibria. Then what particular bargaining protocol and which equilibrium of this protocol will predict the outcome of such a crisis bargaining game? Game theory is silent on this question except when one of the negotiating parties has full bargaining power. I produce unique predictions for this empirically plausible crisis bargaining scenario by formulating it as a principal-agent problem.

The methodology and the predictions depend crucially on the nature of the informational problem and the identity of the informed party. The two classical assumptions in the existing literature regarding the nature of the informational problem fall into two broad categories: When the informational asymmetry is about a player's cost of war, then a player's private information determines only his own payoff from war. We refer to that case as private

values case and the standard principal-agent framework applies. On the other hand, when a player's private information concerns power distribution between players, this information determines his as well as his opponent's payoffs. We refer to that case as common values and the problem becomes a principal-agent problem with informed principal.

In the analysis, Fearon's celebrated risk-return trade-off argument arises as a robust prediction when the party that has full bargaining power is uninformed. This finding is independent of the nature of the informational problem. That is, regardless of whether the informational asymmetry concerns the cost of war or power distribution, if the powerful state is not informed, then her risk-return trade-off may cause inefficient fighting. On the contrary, if the informational asymmetry concerns power distribution between parties and the powerful state is informed, then war may break out not as a consequence of a risk-return trade-off but because the possibility of war signals the type of the high-resolve type principal. This is a novel prediction of the model.

The standard principal-agent framework is applied widely in social sciences. However only few applications of the principal-agent problem with informed principal exist. Crisis bargaining without mediation is a natural application for both the standard principal-agent framework and the one with informed principal. It is likely that the principal-agent framework with or without an informed principal will find more applications in political science.

References

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

[2] Blainey, Geoffrey. 1988. *The Causes of War*. New York: The Free Press.

[3] Bolton, Patrick and Mathias Dewatripont. 2005. *Contract Theory.* Cambridge, MA: The MIT Press.

[4] Bueno de Mesquita, Bruce and David Lalman. 1992. *War and Reason.* New Haven, CT: Yale University Press.

[5] Cho, In-Koo and David M Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102(2): 179-221.

[6] Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(September): 577-592.

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

[8] Fey, Mark and Kristopher W. Ramsay. 2011. "Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict." *American Journal of Political Science* 55(1): 149-169.

[9] Jervis, Robert. 1976. *Perception and Misperception in International Politics.* Princeton, NJ: Princeton University Press.

[10] Leventoglu, Bahar and Ahmer Tarar. 2008. "Does Private Information Lead to Delay or War in Crisis Bargaining?" *International Studies Quarterly* 52: 533-553.

[11] Maskin, Eric S. and Jean Tirole. 1990. "The Principal-Agent Relationship with an Informed Principal, I: Private Values." *Econometrica* 58: 379-410.

[12] Maskin, Eric S. and Jean Tirole. 1992. "The Principal-Agent Relationship with an Informed Principal, II: Common Values." *Econometrica* 60: 1-42.

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

[14] Myerson, Roger B. and Mark Satterthwaite. 1983. "Efficient Mechanisms for Bilateral Trading." *Journal of Economic Theory.* 29(2): 265-281.

[15] Powell, Robert. 1996a. "Bargaining in the Shadow of Power." *Games and Economic Behavior* 15(2): 255-289.

[16] Powell, Robert. 1996b. "Stability and the Distribution of Power." *World Politics* 48(January): 239-267.

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

[18] Reed, William. 2003. "Information, Power and War." *American Political Science Review* 97(4): 633-41.

[19] Salanie, Bernard. 1999. *The Economics of Contracts.* A Primer. Cambridge, MA. The MIT Press.

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

[21] Schultz, Kenneth. 1999. "Do Domestic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53(2): 233-266.

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

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

[24] Van Evera, Stephen. 1999. *Causes of War: Power and the Roots of Conflict.* Ithaca, NY: Cornell University Press.

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

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

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

A The solution of S's problem with uncertainty about costs

I will prove the results in the main text by characterizing the solution to (P1) in several steps. The order of the steps is different than the order of the results in the main text. However I will mention the corresponding results in my analysis below. Suppose that the contract $\{\alpha_l, t_l, \alpha_h, t_h\}$ is a solution to (P1).

Lemma 3 $t_h \leq p - c_l$

Proof. To the contrary, suppose that $t_h > p - c_l$. Also $t_l \geq p - c_l$ since $\{\alpha_l, t_l, \alpha_h, t_h\}$ satisfies $D_l - IR$. Consider the following alternative contract: $t'_h = t'_l = p - c_l$ and $\alpha'_l = \alpha'_h = 0$. This alternative contract also satisfies all the constraints $D_l - IR$, $D_h - IR$, $D_l - IC$ and $D_h - IC$, it does not lead to war and provides higher payoffs for S under peace, so $V(\alpha'_l, t'_l, \alpha'_h, t'_h) > V(\alpha_l, t_l, \alpha_h, t_h)$, which contradicts the optimality of $\{\alpha_l, t_l, \alpha_h, t_h\}$. So $t_h \leq p - c_l$. ■

Then $\max\{t_h, p - c_l\} = p - c_l$ in $D_l - IC$. Also, $\max\{t_l, p - c_h\} = t_l$ in $D_h - IC$ since $t_l \geq p - c_l > p - c_h$. We can now rewrite the problem (P1) as follows:

$$\max_{\alpha_l, t_l, \alpha_h, t_h} V(\alpha_l, t_l, \alpha_h, t_h) = \pi \left[\alpha_l(1 - p - c_S) + (1 - \alpha_l)(1 - t_l)\right] + (1 - \pi)\left[\alpha_h(1 - p - c_S) + (1 - \alpha_h)(1 - t_h)\right]$$

subject to

$$D_l - IR : t_l \geq p - c_l,$$

$$D_h - IR : t_h \geq p - c_h,$$

$$D_l - IC : \alpha_l(p - c_l) + (1 - \alpha_l)t_l \geq p - c_l$$

$$D_h - IC : \alpha_h(p - c_h) + (1 - \alpha_h)t_h \geq \alpha_l(p - c_h) + (1 - \alpha_l)t_l$$

and the usual feasibility constraints $\alpha_i \in [0, 1]$ and $t_i \in [0, 1]$.

Lemma 4 *$\alpha_l \geq \alpha_h$ for any contract $\{\alpha_l, t_l, \alpha_h, t_h\}$ that satisfies the incentive compatibility constraints.*

Proof. By $D_l - IC$ and Lemma 3,

$$\alpha_l(p - c_l) + (1 - \alpha_l)t_l \geq p - c_l \geq \alpha_h(p - c_l) + (1 - \alpha_h)t_h$$

so

$$\alpha_l(p - c_l) + (1 - \alpha_l)t_l \geq \alpha_h(p - c_l) + (1 - \alpha_h)t_h$$

Sum this inequality up with

$$D_h - IC : \alpha_h(p - c_h) + (1 - \alpha_h)t_h \geq \alpha_l(p - c_h) + (1 - \alpha_l)t_l$$

to obtain

$$(\alpha_l - \alpha_h)(c_h - c_l) \geq 0.$$

Since $c_h > c_l$, this implies $\alpha_l \geq \alpha_h$. ■

Lemma 5 $\alpha_h = 0$

Proof. To the contrary, suppose that $\alpha_h > 0$. Consider decreasing α_h by some small $\delta > 0$. Such a decrease does not affect the individual rationality constraints $D_h - IR$, $D_l - IR$ and $D_l - IC$. It does not change the right hand side of $D_h - IC$, and it does not decrease the left hand side of $D_h - IC$ because $t_h \geq p - c_h$ by $D_h - IR$. Thus, the resulting contract is feasible for (P1) and it provides S with a higher expected payoff, which contradicts the optimality of $\{\alpha_l, t_l, \alpha_h, t_h\}$. ■

Lemma 5 proves Result 2 in the text.

Lemma 6 *$D_l - IR$ implies $D_l - IC$ so we can ignore $D_l - IC$.*

Proof. Substitute $t_h \leq p - c_l$ and $\alpha_h = 0$ in $D_l - IC$, then $D_l - IC$ becomes

$$\alpha_l(p - c_l) + (1 - \alpha_l)t_l \geq p - c_l \Leftrightarrow$$
$$(1 - \alpha_l)t_l \geq (1 - \alpha_l)(p - c_l)$$

If $\alpha_l < 1$, then the last inequality is implied by $D_l - IR$. Otherwise the first inequality becomes $p - c_l \geq p - c_l$, which is true. So $D_l - IR$ implies $D_l - IC$ and we can ignore $D_l - IC$. ■

This proves the part of Result 1 that states that we can ignore the incentive compatibility constraint for the high-resolve type.

Lemma 7 *$D_l - IR$ is binding, that is, $t_l = p - c_l$*

Proof. To the contrary suppose that $t_l > p - c_l$. Consider an alternative contract by decreasing t_l by some small $\epsilon > 0$ without changing t_h, α_h and α_l. This alternative contract satisfies all the constraints because (i) $D_l - IR$ continues to hold if ϵ is small enough, (ii) $D_l - IR$ implies $D_l - IC$ (iii) if $\alpha_l > 0$ then $D_h - IC$ becomes slack, otherwise $D_h - IC$ is not affected and (iv) $D_h - IR$ is not affected . The new contract provides a higher expected payoff for S, which is a contradiction, and so $t_l = p - c_l$. ■

This proves the part of Result 1 that states that the individual rationality constraint for the high-resolve type is binding.

Lemma 8 *If $\alpha_l < 1$, then $D_h - IR$ is slack so we can ignore it.*

Proof. Since $\alpha_h = 0$ by Lemma (5), $D_h - IC$ becomes

$$t_h \geq \alpha_l(p - c_h) + (1 - \alpha_l)t_l$$

Also $t_l \geq p - c_l$ by $D_l - IR$ and $p - c_l > p - c_h$, so that

$$\alpha_l(p - c_h) + (1 - \alpha_l)t_l > p - c_h$$

The two inequalities imply that $t_h > p - c_h$ and so we can ignore it when $\alpha_l < 1$. ■

This proves the part of Result 1 that states that we can ignore the individual rationality constraint for the low-resolve type.

Lemma 9 *$D_h - IC$ is binding.*

Proof. Suppose that $\alpha_l < 1$ and $D_h - IC$ is slack. Since $\alpha_h = 0$ by Lemma (5), a slack $D_h - IC$ becomes

$$t_h > \alpha_l(p - c_h) + (1 - \alpha_l)t_l$$

Since $\alpha_l < 1$ implies that $D_h - IR$ is slack by Lemma (8), that is $t_h > p - c_h$, we can decrease t_h by some small $\epsilon > 0$ without violating $D_h - IR$ and $D_h - IC$. $D_l - IR$ is not affected so that $D_l - IC$ also continues to hold by Lemma 6. This new contract provides a higher expected payoff for S, which is a contradiction and so $D_h - IC$ is binding when $\alpha_l < 1$.

If $\alpha_l = 1$, then a slack $D_h - IC$ becomes

$$t_h > p - c_h$$

Then we obtain the same contradiction as above. ■

This proves the part of Result 1 that states that the incentive compatibility constraint for the low-resolve type is binding.

These results imply that either $\alpha_l = 1$ so that $\alpha_h = 0$, $t_l = p - c_l$ and $D_h - IC$ and optimality of the contract imply that $t_h = p - c_h$; or $\alpha_l < 1$.

When $\alpha_l < 1$, (P1) becomes

$$\max_{\alpha_l, t_h} \pi \left[\alpha_l(1-p-c_S) + (1-\alpha_l)(1-p+c_l)\right] + (1-\pi)(1-t_h)$$
$$\text{subject to } t_h = \alpha_l(p-c_h) + (1-\alpha_l)(p-c_l)$$

Substituting $t_h = \alpha_l(p-c_h) + (1-\alpha_l)(p-c_l) = \alpha_l(c_l-c_h) + (p-c_l)$ in the objective, the problem becomes

$$\max_{\alpha_l \in [0,1]} -\alpha_l \left[\pi(c_S+c_l) + (1-\pi)(c_h-c_l)\right] + \pi(1-p-c_S) + (1-\pi)(1-p+c_l)$$

Since the coefficient of α_l is negative, $\alpha_l = 0$ solves the problem, which implies $t_h = p - c_l$.

Thus, the optimal contract is either $C = \{\alpha_l = 1, t_l = p - c_l, \alpha_h = 0, t_h = p - c_h\}$ or $C' = \{\alpha'_l = 0, t'_l = t'_h = p - c_l, \alpha'_h = 0\}$. The former provides S with an expected payoff of

$$\pi(1-p-c_S) + (1-\pi)(1-p+c_h)$$

and the latter provides S with an expected payoff of

$$1 - p + c_l$$

Then

$$1 - p + c_l \geq \pi(1-p-c_S) + (1-\pi)(1-p+c_h)$$

if and only if

$$\pi \geq \pi^* = \frac{c_h - c_l}{c_h + c_S} \in (0, 1)$$

So if $\pi \geq \pi^*$, that is S is confident that he is likely to meet a high-resolve type D, she avoids war by offering $\alpha_l = \alpha_h = 0$ and $t_h = t_l = p - c_l$. Otherwise, she offers $p - c_H$ to a low-resolve type D and she fights with a high-resolve type D. This proves result 3.

Note that I have so far ignored S's individual rationality constraints. Therefore, the solution is a solution to a relaxed problem. If the solution satisfies S's individual rationality constraints, then it solves the more constrained problem. Check that

$$S - IR_h : 1 - t_h \geq 1 - p - c_S \text{ and}$$
$$S - IR_l : 1 - t_l \geq 1 - p - c_S$$

are equivalent to $t_h \leq p + c_S$ and $t_l \leq p + c_S$, which are satisfied. That is, our solution also solves the more constrained problem.

Also, the analysis and the solution will be the same whether or not S privately knows her cost of war.

B The solution of S's problem with uncertainty about distribution of power when D holds private information

I will characterize the solution to (P2) in several steps. The analysis will follow similar steps. Suppose that the contract $\{\alpha_l, t_l, \alpha_h, t_h\}$ is a solution to (P2).

Lemma 10 $t_l \leq p_h - c_D$

Proof. To the contrary, suppose that $t_l > p_h - c_D$. Also $t_h \geq p_h - c_D$ since $\{\alpha_l, t_l, \alpha_h, t_h\}$ satisfies $D_h - IR$. Consider the following alternative contract: $t'_h = t'_l = p_h - c_D$ and $\alpha'_l = \alpha'_h = 0$. This alternative contract satisfies all the constraints $D_l - IR$, $D_h - IR$, $D_l - IC$ and $D_h - IC$, it does not lead to war and provides higher payoffs for S under peace, so $V(\alpha'_l, t'_l, \alpha'_h, t'_h) > V(\alpha_l, t_l, \alpha_h, t_h)$, which contradicts the optimality of $\{\alpha_l, t_l, \alpha_h, t_h\}$. So, $t_l \leq p_h - c_D$. ■

Then we have $\max\{t_l, p_h - c_D\} = p_h - c_D$ in $D_h - IC$. We also have $\max\{t_h, p_l - c_D\} = t_h$ in $D_l - IC$ since $t_h \geq p_h - c_D > p_l - c_D$. Then we can rewrite the problem (P2) as follows:

$$\max_{\alpha_l, t_l, \alpha_h, t_h} V(\alpha_l, t_l, \alpha_h, t_h) = \pi \left[\alpha_h(1 - p_h - c_S) + (1 - \alpha_h)(1 - t_h)\right] + (1 - \pi)\left[\alpha_l(1 - p_l - c_S) + (1 - \alpha_l)(1 - t_l)\right]$$

subject to

$$D_l - IR : t_l \geq p_l - c_D,$$

$$D_h - IR : t_h \geq p_h - c_D,$$

$$D_l - IC : \alpha_l(p_l - c_D) + (1 - \alpha_l)t_l \geq \alpha_h(p_l - c_D) + (1 - \alpha_h)t_h$$

$$D_h - IC : \alpha_h(p_h - c_D) + (1 - \alpha_h)t_h \geq p_h - c_D$$

and the usual feasibility constraints $\alpha_i \in [0, 1]$ and $t_i \in [0, 1]$.

Lemma 11 *$\alpha_h \geq \alpha_l$ for any contract $\{\alpha_l, t_l, \alpha_h, t_h\}$ that satisfies the incentive compatibility constraints.*

Proof. By $D_h - IC$ and Lemma 10,

$$\alpha_h(p_h - c_D) + (1 - \alpha_h)t_h \geq p_h - c_D \geq \alpha_l(p_h - c_D) + (1 - \alpha_h)t_h$$

so

$$\alpha_h(p_h - c_D) + (1 - \alpha_h)t_h \geq \alpha_l(p_h - c_D) + (1 - \alpha_h)t_h$$

Sum this inequality up with

$$D_l - IC : \alpha_l(p_l - c_D) + (1 - \alpha_l)t_l \geq \alpha_h(p_l - c_D) + (1 - \alpha_h)t_h$$

to obtain

$$(\alpha_h - \alpha_l)(p_h - p_l) \geq 0.$$

Since $p_h > p_l$, this implies $\alpha_h \geq \alpha_l$. ■

Lemma 12 $\alpha_l = 0$

Proof. To the contrary, suppose that $\alpha_l > 0$. Consider decreasing α_l by some small $\delta > 0$. Such a decrease does not affect the individual rationality constraints $D_h - IR$, $D_l - IR$ and $D_h - IC$. It does not change the right hand side of $D_l - IC$,and it does not decrease the left hand side of $D_l - IC$ because $t_l \geq p_l - c_D$ by $D_L - IR$. That is, the resulting contract is feasible for (P2) and provides a higher expected utility for S, which contradicts the optimality of $\{\alpha_l, t_l, \alpha_h, t_h\}$. ■

Lemma 13 *$D_h - IR$ implies $D_h - IC$, so we can ignore $D_h - IC$.*

Proof. Substitute $t_l \leq p_h - c_D$ and $\alpha_l = 0$ in $D_h - IC$, then $D_h - IC$ becomes

$$\alpha_h(p_h - c_D) + (1 - \alpha_h)t_h \geq p_h - c_D \Leftrightarrow$$
$$(1 - \alpha_h)t_h \geq (1 - \alpha_h)(p_h - c_D)$$

If $\alpha_h < 1$, then the last inequality is implied by $D_h - IR$. Otherwise the first inequality becomes $p_h - c_D \geq p_h - c_D$, which is true. So $D_h - IR$ implies $D_h - IC$ and we can ignore $D_h - IC$. ■

This proves the part of Result 4 that states that we can ignore the incentive compatibility constraint for the high-resolve type

Lemma 14 *$D_h - IR$ is binding, that is, $t_h = p_h - c_D$.*

Proof. To the contrary suppose that $t_h > p_h - c_D$. Consider an alternative contract by decreasing t_h by some small $\epsilon > 0$ without changing t_l, α_h and α_l. The alternative contract satisfies all the constraints because (i) $D_h - IR$ continues to hold if ϵ is small enough, (ii) $D_h - IR$ implies $D_h - IC$ (iii) if $\alpha_h > 0$ then $D_l - IC$ becomes slack, otherwise $D_l - IC$ is not affected, and (iv) $D_l - IR$ is not affected . The new contract provides a higher expected payoff for S, which is a contradiction. So, $t_h = p_h - c_D$. ■

This proves the part of Result 4 that states that the individual rationality constraint for the high-resolve type is binding.

Lemma 15 *If $\alpha_h < 1$, then $D_l - IR$ is slack so we can ignore it.*

Proof. Since $\alpha_l = 0$ by Lemma (12), $D_l - IC$ becomes

$$t_l \geq \alpha_h(p_l - c_D) + (1 - \alpha_h)t_h$$

Also $t_h \geq p_h - c_D$ by $D_h - IR$ and $p_H - c_D > p_l - c_D$, so that

$$\alpha_h(p_l - c_D) + (1 - \alpha_h)t_h > p_l - c_D$$

The two inequalities imply that $t_l > p_l - c_D$ so that we can ignore it when $\alpha_h < 1$. ■

This proves the part of Result 4 that states that we can ignore the individual rationality constraint for the low-resolve type.

Lemma 16 *If $\alpha_h < 1$, then $D_l - IC$ is binding.*

Proof. Suppose that $\alpha_h < 1$ and $D_l - IC$ is slack. Since $\alpha_l = 0$ by Lemma (12), $D_l - IC$ becomes

$$t_l > \alpha_h(p_l - c_D) + (1 - \alpha_h)t_h$$

Since $\alpha_h < 1$ implies that $D_l - IR$ is slack by Lemma (15), that is $t_l > p_l - c_D$, we can decrease t_l by some small $\epsilon > 0$ without violating $D_l - IR$ and $D_l - IC$. $D_h - IR$ is not affected so $D_h - IC$ also continues to hold by Lemma 13. This new contract provides a higher expected payoff for S, which is a contradiction. So $D_l - IC$ is binding when $\alpha_h < 1$. ■

This proves the part of Result 4 that states that the incentive compatibility constraint for the low-resolve type is binding.

These results imply that either $\alpha_h = 1$ so that $\alpha_l = 0$, $t_h = p_h - c_D$ and $D_l - IC$ and optimality of the contract imply that $t_l = p_l - c_D$; or $\alpha_h < 1$. When $\alpha_h \leq 1$, (P2) becomes

$$\max_{\alpha_h, t_l} \pi \left[\alpha_h(1 - p_h - c_S) + (1 - \alpha_h)(1 - p_h + c_D)\right] + (1 - \pi)(1 - t_l)$$
$$\text{subject to } t_l = \alpha_h(p_l - c_D) + (1 - \alpha_h)(p_h - c_D)$$

Substituting $t_l = \alpha_h(p_l - c_D) + (1 - \alpha_h)(p_h - c_D) = \alpha_h(p_l - p_h) + (p_h - c_D)$ in the objective, the problem becomes

$$\max_{\alpha_h \in [0,1]} \alpha_h \left[(1 - \pi)(p_h - p_l) - \pi(c_S + c_D)\right] + (1 - p_h + c_D)$$

Then

$$\alpha_h = \begin{cases} 0 \text{ if } \pi > \pi^{**} = \frac{p_h - p_l}{(p_h - p_l) + (c_S + c_D)} \\ 1 \text{ if } \pi \leq \pi^{**} \end{cases}$$

Thus, the optimal contract is either $C = \{\alpha_l = 0, t_l = p_h - c_D, \alpha_h = 0, t_h = p_h - c_D\}$ if $\pi > \pi^{**}$ or $C' = \{\alpha'_l = 0,\ t'_l = p_l - c_D,\ \alpha'_h = 1,\ t'_h = p_h - c_D\}$ if $\pi \leq \pi^{**}$.

In other words, if S is confident enough that D is likely to be a high-resolve type, $\pi > \pi^{**}$, then she avoids war by offering $t_h = t_l = p_h - c_D$. Otherwise, she solves her risk-return trade-off problem by offering $p_l - c_D$, which is accepted by a low resolve type D and which leads to war with a high-resolve type D.

We have so far ignored S's individual rationality constraints. Therefore our solution is a solution to a relaxed problem. If the solution satisfies S's individual rationality constraints, then it also solves the more constrained problem. Check that

$$S - IR_h : 1 - t_h \geq 1 - p_h - c_S \text{ and}$$
$$S - IR_l : 1 - t_l \geq 1 - p_l - c_S$$

are equivalent to $t_h \leq p_h + c_S$ and $t_l \leq p_l + c_S$. The first one is satisfied trivially. For the second one, either $t_l = p_l - c_D < p_l + c_S$ so that it holds, or $t_l = t_h = p_h - c_D \leq p_l + c_S$ which is satisfied if $p_h - p_l \leq c_S + c_D$ are satisfied. Denoting $\bar{c} = p_h - p_l$, $c_S + c_D \geq \bar{c}$ is Fey and Ramsay's (2011) sufficiency

condition for the existence of a crisis bargaining game with mediation that has voluntary agreements in which an always peaceful equilibrium exists. To the contrary, our result states that, in this case, if $\pi < \pi^*$, then war breaks out with probability π.

C The solution of S's problem with uncertainty about distribution of power when S holds private information

I only provide the solution of the optimal separating equilibrium here in the appendix. The solution of the optimal pooling equilibrium and the optimal equilibrium are in the main text.

Lemma 17 $\alpha_l^{sep} \geq \alpha_h^{sep}$

Proof. Summing up $S_l - IC$ and $S_h - IC$ yields $\alpha_l^{sep}(p_h - p_l) \geq \alpha_h^{sep}(p_h - p_l)$. Then $p_h > p_l$ implies $\alpha_l^{sep} \geq \alpha_h^{sep}$. ■

Lemma 18 $t_h^{sep} \geq t_l^{sep}$

Proof. By individual rationality for S, it must be the case that $1 - t_l^{sep} \geq 1 - p_l - c_S$. Then $\alpha_l^{sep} \geq \alpha_h^{sep}$ and $S_l - IC$ imply that $1 - t_l^{sep} \geq 1 - t_h^{sep}$, which is equivalent to $t_h^{sep} \geq t_l^{sep}$. ■

Lemma 19 $\alpha_h^{sep} = 0$

Proof. Suppose that $\alpha_h^{sep} > 0$. Then $\alpha_l^{sep} > 0$. Decrease α_h^{sep} by some small $\epsilon > 0$ and α_l^{sep} by some small $\delta > 0$ such that

$$\epsilon(p_h + c_S - t_h) = \delta(p_h + c_S - t_l)$$

Individual rationality for S implies $1-t_h^{sep} \geq 1-p_h-c_S$, equivalently $p_h+c_S \geq t_h^{sep}$. Also $t_h^{sep} \geq t_l^{sep}$ from the previous lemma so that the coefficients of ϵ and δ are both nonnegative. Then $S_h - IC$ and all other constraints continue to hold and V_h^{sep} and V_l^{sep} increase. This is a contradiction so $\alpha_h^{sep} = 0$. ■

Lemma 20 *$S_h - IC$ holds with equality.*

Proof. Suppose that $S_h - IC$ is slack. If $\alpha_l^{sep} > 0$ or $t_l^{sep} > p_l - c_D$, then slightly decreasing α_l^{sep} or t_l^{sep} increases V_l^{sep} without violating any of the constraints. So $\alpha_l^{sep} = 0$ and $t_l^{sep} = p_l - c_D$ must hold. Then $S_h - IC$ becomes $1 - t_h^{sep} \geq 1 - p_l + c_D$, equivalently $t_h^{sep} \leq p_l - c_D$. But this is a contradiction because $t_h^{sep} \geq p_h - c_D$ by $D_h - IR$ and $p_h - c_D > p_l - c_D$. Then $S_h - IC$ holds with equality. ■

Lemma 21 *$S_l - IC$ is slack.*

Proof. Substitute $\alpha_h^{sep} = 0$ in $S_h - IC$ and $S_l - IC$. Then

$$\begin{aligned} 1 - t_h^{sep} &= \alpha_l^{sep}(1 - p_h - c_S) + (1 - \alpha_l^{sep})(1 - t_l^{sep}) \\ &< \alpha_l^{sep}(1 - p_l - c_S) + (1 - \alpha_l^{sep})(1 - t_l^{sep}) \end{aligned}$$

where the equality is the $S_h - IC$ constraint and the inequality is implied by $p_l < p_h$. But the inequality is $S_l - IC$ so $S_l - IC$ is slack. ■

The last two Lemmata prove Result 7.

Lemma 22 $t_h^{sep} = p_h - c_D$

Proof. If $t_h^{sep} > p_h - c_D$ then slightly decrease t_h^{sep}. $D_h - IR$ and $S_l - IC$ continue to hold for a small enough decrease, $D_l - IR$ is not affected, $S_h - IC$ becomes slack and V_h^{sep} increases. This is a contradiction so $t_h^{sep} = p_h - c_D$.

■

So the solution to the problem of the low-resolve S is given by $\alpha_h^{sep} = 0$ and $t_h^{sep} = p_h - c_D$.

The problem of the high-resolve type *becomes*

$$(\alpha_l^{sep}, t_l^{sep}) \text{ solves } \max_{(\alpha_l, t_l)} \alpha_l(1 - p_l - c_S) + (1 - \alpha_l)(1 - t_l)$$

$$\text{subject to}$$

$$D_l - IR : t_l \geq p_l - c_D$$

$$S_h - IC : 1 - p_h - c_S = \alpha_l(1 - p_h - c_S) + (1 - \alpha_l)(1 - t_l)$$

Take the total differential of $S_h - IC$ with respect to α_l and t_l:

$$(p_h + c_S - t_l)d\alpha_l + (1 - \alpha_l)dt_l = 0$$

Take the total differential of the objective function with respect to α_l and t_l and replace the above equality:

$$\begin{aligned} d(\text{objective}) &= (t_l - (p_l + c_S))d\alpha_l - (1 - \alpha_l)dt_l \\ &= (p_h - p_l)d\alpha_l \end{aligned}$$

So increasing α_l and decreasing t_l increases the objective. Then the solution to the problem of the high-resolve type S is given by $t_l^{sep} = p_l - c_D$, and substitute that in $S_h - IC$ to obtain

$$\alpha_l^{sep} = \frac{p_h - p_l}{(p_h - p_l) + (c_D + c_S)}$$

D Take-it-or-leave-it offer

Suppose that S knows the true value of p and $c_S + c_D \geq \pi(p_h - p_l)$. Refer to S as S_h if $p = p_h$ and S_l if $p = p_l$.

In a bluffing equilibrium, S_h bluffs by imitating S_l. Here, I solve for the bluffing equilibria that arises with the take-it-or-leave-it bargaining protocol.

Let $\alpha \in [0, 1]$ be the probability that S_h bluffs in a bluffing equilibrium. Define ϕ as

$$\phi = \frac{\pi\alpha}{\pi\alpha + (1 - \pi)} \in [0, 1)$$

and t_l as

$$t_l = \phi(p_h - c_D) + (1-\phi)(p_l - c_D) = p_l - c_D + \phi(p_h - p_l) < p_h - c_D$$

and β as

$$\beta = \frac{c_D + c_S}{p_h + c_S - t_l} = \frac{c_D + c_S}{c_D + c_S + (1-\phi)(p_h - p_l)} \in (0,1)$$

Next I will provide a Bayesian Nash equilibrium with bluffing in which S_l always offers t_l, S_h bluffs by offering t_l with probability α and offering $t_h = p_h - c_D$ otherwise, and when D receives the low offer of t_l, D believes that $p = p_h$ with probability ϕ and fights with probability β.

For any offer $t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$, define

$$\hat{\phi}(t) = \frac{t - (p_l - c_D)}{p_h - p_l} \in [0,1]$$

Consider the following strategy and belief profile:

- S_l always offers t_l,
- S_h offers t_l with probability α and t_h with probability $1 - \alpha$,
- D's beliefs and strategy are given as follows:
  - If D receives an offer of t_h, he believes that $p = p_h$ with probability 1 and accepts the offer,

- If D receives an offer of t_l, he believes that $p = p_h$ with probability ϕ and he accepts the offer with probability β and rejects it with probability $1 - \beta$,
- If D receives any other offer $t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$, D believes that $p = p_h$ with probability $\hat{\phi}(t)$ or higher and he rejects the offer.

In a take-it-or-leave-it protocol, war breaks out when an offer is rejected.

First, I will show that D's strategy is optimal given his beliefs.

Suppose that D receives an offer of t_h. Then he believes that $p = p_h$ with probability 1. Now, accepting the offer is optimal for him because rejecting it provides him with the same payoff $p_h - c_D$.

If D receives an offer of t_l, then he believes that $p = p_h$ with probability ϕ. To accept this offer with probability $\beta \in (0, 1)$, he must be indifferent between accepting and rejecting the offer:

$$t_l = \phi(p_h - c_D) + (1 - \phi)(p_l - c_D)$$

where the left hand side is D's payoff from accepting the offer and the right hand side is his expected payoff from rejecting it given his beliefs. This equality holds by definition of t_l, and it is optimal for D to mix between accepting and rejecting the offer.

Suppose that D receives an offer of $t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$ and updates his beliefs so that he believes $p = p_h$ with probability $\tilde{\phi}$. For fighting to be optimal, it must be the case that

$$\tilde{\phi}(p_h - c_D) + (1 - \tilde{\phi})(p_l - c_D) \geq t$$

where the left hand side is D's expected payoff from rejecting the offer given his beliefs and the right hand side is his payoff from accepting it. This inequality is equivalent to $\tilde{\phi} \geq \hat{\phi}(t)$, so it is optimal for D to reject the offer when he receives an offer of $t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$.

Next I will show that S's strategy is optimal given D's strategy and beliefs.

Suppose $p = p_h$. If S_h offers $t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$ then D fights with probability 1 so S_h's payoff from offering t is $1 - p_h - c_S$. Her payoff from offering $t_h = p_h - c_D$ is $1 - p_h + c_D$, since D accepts it with probability 1. So offering $t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$ with positive probability cannot be optimal for S_h and she either offers t_l or t_h. For S_h to bluff by offering t_l with positive probability of α, it must be the case that she is indifferent between offering t_h and t_l:

$$1 - p_h + c_D = \beta(1 - t_l) + (1 - \beta)(1 - p_h - c_S)$$

where the left hand side is S's payoff from offering t_h, which D accepts with probability 1, and the right hand side is her expected payoff from offering t_l,

which D accepts with probability β. This equality holds by definition of β so it is optimal for S_h to bluff with positive probability.

Suppose that $p = p_l$. If S_l offers t_l, her payoff is

$$\beta(1 - t_l) + (1 - \beta)(1 - p_l - c_S)$$

If S offers t_h, her payoff is $1 - p_h + c_D$ since D accepts t_h with probability 1. By definition of β, this payoff is less than her payoff from offering t_l. So S_l does not offer t_h. If she offers $t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$ then her payoff is $1 - p_l - c_S$ because then D fights with probability 1. This payoff is less than her payoff from offering t_l if and only if $c_S + c_D \geq \pi(p_h - p_l)$, which I have assumed. So it is optimal for S_l to offer t_l with probability 1.

Finally, I confirm the consistency of D's beliefs on the equilibrium path. S offers t_h only when $p = p_h$, so it must be the case that $p = p_h$ after observing t_h. Since both types of S offer $t_{l,}$ D's belief after receiving t_l must follow Bayes' rule:

$$\Pr(p = p_h | t = t_l) = \phi = \frac{\pi\alpha}{\pi\alpha + (1 - \pi)} \in [0, \pi]$$

where, given the equilibrium strategies, the numerator is the ex ante probability that $t = t_l$ will be offered by S_h and the denominators is the ex ante probability that $t = t_l$ will be offered. Thus, D's beliefs are consistent with the Bayes' rule on the equilibrium path. Since no offer of $t \in (t_l, p_h - c_D)$ will be made on the equilibrium path, D's beliefs after receiving an offer of

$t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$ can be arbitrary.

$\alpha = 0$, $\phi = 0$, $t_l = p_l - c_D$ and $\beta = \frac{c_D + c_S}{(c_D + c_S) + (p_h - p_l)}$ describe a separating equilibrium with no bluffing.

S_h's equilibrium payoff is

$$\begin{aligned} u_h(\alpha) &= (1-\alpha)(1 - p_h + c_D) + \alpha\left[\beta(1 - t_l) + (1-\beta)(1 - p_h - c_S)\right] \\ &= 1 - p_h + c_D \end{aligned}$$

and S_l's equilibrium payoff is

$$\begin{aligned} u_l(\alpha) &= \beta(1 - t_l) + (1-\beta)(1 - p_l - c_S) \\ &= 1 - p_l - c_S + \beta(p_l + c_S - t_l) \end{aligned}$$

Applying the intuitive criterion (Cho and Kreps, 1987) selects the separating equilibrium with no bluffing: The first step is to check if there is a type of S that potentially benefits from deviating to an offer of $t \in [p_l - c_D, p_h - c_D] \backslash \{t_l, t_h\}$ off the equilibrium path. Consider the belief $\Pr(p = p_l | t) = 1$ for D after receiving such an offer. Given that belief, it is optimal for D to accept the offer. Then both S_h and S_l obtain a payoff of $1 - t$, which is greater than their equilibrium payoff if

$$\begin{aligned} &1 - t > 1 - p_h + c_D \text{ and} \\ &1 - t > 1 - p_l - c_S + \beta(p_l + c_S - t_l) \end{aligned}$$

The first inequality holds because $t < p_h - c_D$. The second inequality holds if

$$t < \hat{t}(\alpha) = p_l + c_S - \beta(p_l + c_S - t_l)$$

It is easily verified that $t_l \leq \hat{t}(\alpha) < p_h - c_D$. So, if $t \geq \hat{t}(\alpha)$, then D must believe that $p = p_h$ and since $p_h - c_D > t$, it is optimal for D to fight when $t \geq \hat{t}(\alpha)$. If $t < \hat{t}(\alpha)$ then both types of S can potentially benefit and so we need to check whether either type can do better after such a deviation. The war payoff is the minimum payoff for both types. D fights with probability 1 if

$$p - c_D > t$$

When D receives $t < \hat{t}(\alpha)$, he can assume any beliefs. If he assumes $p = p_h$ with probability 1, then $p_h - c_D > t$ so that he fights and both types of S obtain their minimum payoff off the equilibrium path. Since

$$1 - p_h - c_S < u_h(\alpha) = 1 - p_h + c_D$$

and

$$1 - p_l - c_S < u_l(\alpha) = 1 - p_l - c_S + \beta(p_l + c_S - t_l)$$

the α-bluffing equilibrium does not fail the intuitive criterion.