Verifiable Information Transmission in a Crisis Bargaining Model

Eva Bernauer
Center for Doctoral Studies in Social and Behavioral Sciences (CDSS)
Mannheim Center for European Social Research (MZES)
Mannheim University
eva.bernauer@lysium.de

Abstract

The present study develops a formal model on the rationale and effectiveness of verifiable information transmission - the core characteristic of confidence-building measures (CBMs). CBMs have received high attention in the case study literature and day-to-day political business, yet are largely neglected in formal literature. This is surprising given that the desirability and peace-inducing effects of CBMs are debated. The formal model shows that by unilaterally or bilaterally increasing the amount of verifiable information concerning military aspects, CBMs may increase the parties' expected payoffs and eliminate the risk of war. By offering a solution to the renowned information problem in crisis bargaining, without resorting to domestic audiences or third parties, this study also contributes to the crisis bargaining literature.

1 Introduction

Confidence-building measures (CBMs) have greatly proliferated since first being established by the United States and Soviet Union and the members of the Conference on Security and Cooperation in Europe (CSCE) during the Cold War. Despite sometimes falling short of their goal, CBMs are primarily intended to credibly increase military information between states in order to prevent wars occurring due to misunderstanding and miscalculation (Alford 1981; Desjardins 1996; Lachowski 2005). While the first Helsinki CSCE approach to CBMs proved unsuccessful owing to missing verification procedures (Alford 1981, 137; Desjardins 1996, 16), later attempts in the Stockholm and Vienna Documents provided for the extensive verification of military information and represented a success story (Lachowski 2005, 13).[1] Over the years, OSCE members have agreed both multilaterally and bilaterally to diverse modes of transmitting verifiable information, including the mutual observation of military manoeuvers, on-site inspections and aerial surveillance (Lachowski 2005). Outside the OSCE, Argentina and Brazil are among the most ardent proponents of CBMs. In 1990, both states agreed to establish the Brazilian-Argentine Agency for Accounting and Control of Nuclear Materials (ABACC), with the goal of verifying both states' use of all nuclear materials and facilities exclusively for peaceful purposes. In practice, the agreement has led to a joint inspection mechanism through which Argentinian nationals inspect Brazilian nuclear facilities, and vice versa, in order to mutually assure peaceful intentions (Redick 1999, 233). Other state dyads in Asia, the Middle East, and Africa have followed a similar path, albeit often with less ambitions and success (Krepon 1999).

Why do states agree to share verifiable information about military aspects? Do they benefit from this undertaking; moreover, does it reduce the risk of war?[2] This study's main idea is as follows: the concern leading to verifiable information transmission in the first place is that war might arise due to misunderstanding and miscalculation. Military actions are seen as inherently ambiguous, and thus can easily be misinterpreted: "A routine military training exercise, for instance, can be mistaken by another state as an offensive action and trigger an undesirable reaction, leading to unintended conflict" (Desjardins 1996, 21). Likewise, a harmless reconnaissance flight or naval deployments can be mistaken as an act of aggression. Similarly, military rhetoric that serves mere domestic purposes yet has little substance as a real danger can trigger a hostile response. In all such cases, uncertainty about the other's actual military capabilities and intentions can bring about military escalation. When allowing for CBMs in the form of external observers monitoring troop movements, inspecting weapon arsenals or observing nuclear facilities, the hope is that ambiguity is reduced. Observers receive informative signals about a potentially hostile state's true intentions, and despite this information not being perfect, it helps states to update their beliefs about each other's intentions and better assess the situation in place, which should subsequently lead to a risk reduction of accidental war.[3]

In order to capture this idea formally, the present study incorporates verifiable information transmission into a standard take-it-or-leave-it crisis bargaining model with uncertainty concerning war payoffs. This standard model depicts a bargaining situation between one state, the owner of a contested good, and another, a potential challenger who threatens to fight a war over the contested good. The owner makes an offer, which the challenger can accept or reject, with rejection resulting in ex-post inefficient war. Uncertainty concerning war payoffs leads to a risk-return–tradeoff with the owner sometimes making a low offer that a strong challenger rejects. In order to model the transmission of verifiable information, this study adds a pre-negotiation phase to the standard crisis bargaining model, in which the states, who do not yet know their own expected war payoffs, can commit themselves to the later revelation of a random variable, the value of which correlates with the uncertain war payoff. Upon learning the value of the random variable, the states update their belief about the other's type. For example, if the challenger commits itself to sharing verifiable information about its military budget, troop size or nuclear program, the owner of the contested good can better assert whether it faces a strong or weak challenger. This assertion subsequently helps the owner to make a more appropriate offer, which the challenger rejects with less frequency. Consequently, the prevention of war generates a surplus that can be potentially divided by the states.

While CBMs can be concluded unilaterally, bilaterally or multilaterally, this study focuses only on unilateral and bilateral CBMs.[4] Although multilateral CBMs clearly merit attention, unilateral and bilateral CBMs are more interesting theoretically, because they are deemed effective absent an external information generating or enforcement mechanism. The present study determines the conditions for no, unilateral, and bilateral verifiable information transmission to hold. It predicts that no information transmission occurs if the challenging state is likely to be strong irrespective of the other state's type; that unilateral information transmission occurs if the challenging state is likely to be weak and its opponent is likely to be strong; and that bilateral information transmission occurs if both states are likely to be weak. Under the model's assumptions, it also shows that both unilateral and bilateral information transmission reduce the risk of war to zero. That said, this study provides a formal backing to the often empirically assumed rationale and effectiveness of confidence-building measures. Simultaneously, the study meaningfully extends the standard take-it-or-leave-it crisis bargaining model in that it offers a solution to the information problem without relaxing the unitary actor assumption or introducing third parties.

The following section introduces confidence-building measures in general and verifiable information transmission in particular. Section 3 holds that verifiable information transmission is best analyzed as a solution to the information problem in a crisis bargaining model, thus complementing other possible solutions to the information problem. Section 4 presents the theoretical model. Section 5 discusses the model's implications and assumptions, and section 6 concludes.

2 Verifiable Information Transmission

The term confidence-building measures dates back to the early days of the Cold War, when the United States and Soviet Union agreed on a hotline and other measures of consultation and notification following the Cuban Missile Crisis, in order "to create more reliable communication channels for the exchange of information, particularly following unforeseen incidents or accidents" (Fisher 1999, 296; Krepon 1999, 1ff.). These initial steps were followed by more far-reaching measures taken in the Helsinki Final Act (1975) and the Stockholm and Vienna Agreement (1986, 1990), which included the obligatory observation of military exercises and on-site inspections. After the Cold War, the United States and Russia deepended the confidence-building process in the Vienna Documents (1992, 1994, 1999, 2011), the CFE Treaty, and the Open Skies Treaty, which provided inter alia the obervation of military exercises, the reduction of troop sizes, the renouncement of certain classes of weapons, and aerial surveillance.

The confidence-building process between the United States and Soviet Union exemplifies the stepwise nature of confidence-building with verifiable information transmission at its heart. Confidence-building typically proceeds in three steps.[5] First, states commonly convey strategically relevant, yet non-verifiable information about military aspects, such as via hotlines, visits, and reports.[6] Such measures of *non-verifiable* information transmission are often discussed as 'precursors' for rather than as 'real' CBMs (Krepon, Newbill, Khoja, and Drezin 1999, 4f.), unable to sincerely address insecurities between states (Desjardins 1996, 5, 16, 40, 61). Formally, they are analyzed as cheap talk (Fearon 1995). In a next step, confidence-building continues with measures that further increase the amount of militarily relevant information, but this time information is *verifiable*. Examples of verifiable information transmission include the observation of troop movements, and the inspection of weapons arsenals or nuclear facilities. These steps are at the core of CBMs, and are intended to prevent accidental war resulting from misunderstanding and misinterpretation (United Nations 1982, 6ff.). Finally, the confidence-building process may culminate in the conclusion of measures that *constrain* states militarily. Examples include constraints on troop exercises, the reduction of troop size, and the renouncement of certain classes of weapons. These measures are often discussed under the heading of arms control and arms reduction measures, and arguably go beyond confidence-building (United Nations 1982, 6ff.). In any case, verifiable information transmission is at the heart of a stepwise confidence-building process. However, despite its outstanding role, verifiable information transmission has been largely neglected in formal literature.

While CBMs can be concluded unilaterally, bilaterally, or multilaterally, the present study focuses only on unilateral and bilateral measures of verifiable information transmission. While multilateral CBMs clearly merit attention, one might argue that unilateral and bilateral verifiable information transmission is the more theoretically interesting phenomenon. For realists, it is surprising that states can unilaterally or bilaterally pacify their relations beyond coercion and deterrence in an anarchic international system (Mearsheimer 1995); liberalists wonder that any pair of states should be able to pacify relations, not just those with joint democracy, economic trade, or links via international organizations (Oneal and Russett 2001); and for bargaining theorists, it is surprising that states can manage the information problem without relaxing the unitary actor assumption or introducing third parties (Fearon 1997; Kydd 2003; Rauchhaus 2006). In contrast, multilateral measures are less puzzling because other states or organizations might possess an information advantage or enforcement capacities which facilitates the prevention of violent conflict (Boehmer, Gartzke, and Nordstrom 2004).

The lack of formal models on verifiable information transmission is surprising given that the desirability and effectiveness of these measures is debated. While most theorists would doubt the relevance of CBMs, case study scholars and practicioners tend to evaluate these measures positively (United Nations 1982; Krepon, Newbill, Khoja, and Drezin 1999). During the last twenty years, CBMs have been concluded between numerous dyads across the world, and if CBMs were indeed detrimental or inconsequential, it would be a puzzle why they are so frequently concluded. Nevertheless, case studies provide a mixed record on the effectiveness of CBMs. On the one hand is supposedly successfull cases such as the United States and Soviet Union, France and Germany, Argentina and Chile, or Argentina and Brazil, where CBMs were followed by the absence of militarized interstate disputes (MIDs).[7] However, on the other hand is examples like India and Pakistan, who took several bilateral steps of confidence-building, yet nevertheless experienced numerous MIDs (Devabhaktuni, Rudolph, and Newbill 1999; Kapur 2005). All in all, the ubiquity of CBMs and case study evidence suggest that CBMs *can* be beneficial and successful, but only under certain *conditions* (Desjardins 1996, 5, 38f., 61). However, such conditions have never been formally explored. The present study intends to fill this gap.

One notable exception to the scarcity of formal models on CBMs is Kydd (2005), who conceptualizes confidence-building as a two-round Security Dilemma Game, whereby the players can cooperate or defect in each round without knowing what the other party did in that particular round. However, in the second round, both players know what the other player did in the first round, and under certain conditions this first round behavior can serve as a signal concerning the other player's type. The shortcoming of Kydd's model is that it merely conceptualizes CBMs as a cooperative gesture, and does not model the most important aspect of CBMs, namely the reduction of uncertainty about each other's war payoffs via the transmission of verifiable information about military aspects. I argue that this aspect is best tackled in a model that explicitly identifies the lack of information as a key cause of war, as Fearon's (1995) crisis bargaining model does.

3 The Crisis Bargaining Framework

This section argues that unilateral and bilateral verifiable information transmission is best analyzed as a solution to the information problem in a crisis bargaining model, thus complementing other existing solutions to the information problem.

Fearon's (1995) take-it-or-leave-it crisis bargaining model is the logical starting point for modeling CBMs in general, and verifiable information transmission in particular. This crisis bargaining model provides an explanation for why certain classes of wars occur, namely accidental wars resulting from information problems, or preemptive and preventive wars resulting from commitment problems. As CBMs address the information problem via information transmission and the commitment problem via military constraints, CBMs address exactly the types of military conflicts identified by the crisis bargaining model as the two key causes for war (Krepon, Newbill, Khoja, and Drezin 1999). Evidently, there are other explanations for the occurrence of wars, including higher expected payoffs from war than peaceful settlement (Fearon 1995, 386). However, the bargaining model neither intends to explain these kinds of 'aggressive' conflicts, nor do CBMs have the capacity to tackle them (United Nations 1982, 8; Krepon, Newbill, Khoja, and Drezin 1999, 4; Desjardins 1996, 4, 22). Thus, the crisis bargaining perspective appears to lend itself perfectly for modeling CBMs.

Despite the closeness of the crisis bargaining perspective and CBMs, Kydd (2005) rejects the idea of using this perspective for modeling CBMs. Kydd focuses on the role of trust when conceptualizing CBMs, arguing that trust should be perceived as "the beliefs one side has about the likelihood that the other prefers to reciprocate cooperation rather than exploit it" (Kydd 2005, 29). He therefore uses a Prisoner's Dilemma and an Assurance Game framework to model these two different underlying motivations, rejecting the bargaining model because it focuses on "uncertainty about the bargaining leverage of the two sides" (Kydd 2005, 30) rather than on uncertainty about underlying motivations. Kydd adds, that in his view, "trust is more a matter of whether each side will honor whatever bargain has been reached, rather than on what bargain is reached in the first place", which represents his second reason for not using the bargaining model (Kydd 2005, 30). Although Kydd's choice for the classical conflict games is justifiable in light of his definition of trust, this study adopts a different perspective, holding that the main purpose of CBMs is to shed light on the uncertain *bargaining leverage* that two parties have in *future crisis bargaining situations*. That said, the crisis bargaining model is the logical starting point for modeling CBMs.

As CBMs are intended to prevent accidental and preemptive or preventive wars, let us recap why these wars occur in the crisis bargaining model (Fearon 1995). Accidental wars occur because one party has uncertainty about the other party's war payoffs. Facing a risk-return-tradeoff, the party might propose the division of a contested good, which the other party rejects because it expects higher payoffs from fighting. Thus, war occurs due to a lack of information (the so-called information problem). Fearon then shows that war can still happen with perfect information, namely due to changing military capabilities. If a country expects to lose relative power in the future, it might be better to launch a military strike today than to wait until tomorrow, as the rising country can hardly commit to not exploiting its advantage and challenging in the future. Thus, war results due to a lack of commitment (the so-called commitment problem). Given the ex-post inefficiency of war, the question of whether states can address the information and the commitment problem thus preventing inefficient war arises, and how this can be achieved through unilateral and bilateral verifiable information transmission.

The literature discusses at least three possible solutions to the information problem.[8] However, none of these solutions does justice to the present study's topic of interest - unilateral and bilateral verifiable information transmission. The first possible solution to the information problem discussed within the literature is non-verifiable information transmission or cheap talk (Fearon 1995, Kydd 1997). In an often-cited economic model, Crawford and Sobel (1982) show that two agents, a sender and a receiver, are better off if the better-informed sender provides a noisy, non-verifiable signal to the receiver, whose decision then affects both agents' payoffs. However, Crawford/Sobel reveal that this kind of non-verifiable information transmission only works if the parties' interests are not too far apart. In contrast, in a crisis bargaining model where the parties' interests are opposed, it is not possible to credibly communicate one's own degree of resolve: states have an incentive to say that they expect high war payoffs, despite in fact expecting low war payoffs. This implies that any talk is incredible and inconsequential (Fearon 1995, 412; Guisinger and Smith 2002, 176), unless conducted in the shadow of costly signals (Kurizaki 2007), occurring in multidimensional bargaining (Trager 2011), involving reputation effects (Sartori 2002; Guisinger and Smith 2002), or containing no pre-commitment to negotiation after diplomatic statements have been exchanged (Ramsay 2011). Cheap talk or non-verifiable information transmission differs from verifiable information transmission in the former being merely a verbal, non-verifiable announcement, whereas the latter actually *reveals*, though only partially, a military aspect of interest.

The second solution to the information problem that has received large coverage is costly signaling (Fearon 1994; Fearon 1997; Schultz 2001)[9], examples of which include the mobilization of troops or the announcement of one's resolve in front of domestic or foreign audiences. The idea is that costly signals separate resolved and unresolved types, namely that resolved types send the signal whereas unresolved types do not.[10] More concretely, only the resolved types mobilize troops and thereby risk a war, or only the resolved types threaten to fight a war because they know that a bluff will be punished by domestic or foreign audiences.[11] Formally, mobilization is modeled via sunk cost signals that impose costs on all outcomes thus having pure informational value, while audience costs are modeled via tying hands signals that impose costs only when a state backs down after threatening to fight.[12] Tying hands signals thus *commit* states to fighting, which is supposed to prevent the opponent from challenging in the first place (Fearon 1997, 412). Formally, costly signals differ from verifiable signals in that the former impose additional costs on the payoffs, whereas the latter do not. Verifiable signals directly reveal something about the uncertain aspect in the payoffs, whereas costly signals only do this indirectly via the costs.

The third solution to the information problem is information transmission with the help of a third party (Kydd 2003; Rauchhaus 2006). The idea is that a third party might be better informed about a state's opponent than the state itself is (though the third party is not perfectly informed). This information advantage could be owing to better intelligence gathering, communication, or past alliance ties. The third party then communicates the information about the opponent to the other state, and the state updates its belief. As information transmission is merely verbal (or cheap talk), the state does not know for certain whether the third party is telling the truth. However, one can show that the third party is credible if it has close ties with the state that it informs (Kydd 2003), or if it is impartial and has no strong interest in peace (Rauchhaus 2006). One shortcoming of this literature is that it neglects how the third party actually acquires the information (Kydd 2010, 117). Beyond that, it appears to only address rare empirical occasions, because the third party needs to have close ties with one party, but good information about another party with which it explicitly has no close ties.[13] Formally, cheap talk via third parties and verifiable information transmission are related: in both cases, information about the uncertain war payoff is conveyed via a variable imperfectly correlated with the variable of interest (third party knowledge vs. CBMs). Interestingly, however, the present study shows that it does not require a third party to tackle the information problem, as verifiable information transmission alone can achieve this.

The study most closely related to the present study is that of Black/Bulkley (1988).

Coming from an economic context, the authors model a situation whereby a seller and buyer bargain over the price of a commodity yet have uncertainty about its value, which can be high or low. Thus, the seller does not know whether to price high or low. Trade occurs if the bargain is struck, and otherwise it fails. Similarly to the present study's model, the parties can commit themselves to the revelation of a random variable, such as their income or wealth, which correlates with the value of the commodity. This transmission of verifiable information helps the seller to make a more appropriate offer, which the buyer subsequently rejects with less frequency. The present study borrows the formalization of verifiable information transmission via the revelation of a random variable from Black/Bulkley. However, it changes the interpretation of the model and the parties' expected payoff functions (Black and Bulkley 1988). It deepens and extends the analysis, presents all intermediate steps, and corrects for several mistakes.

4 The Model

This section develops a crisis bargaining model with uncertainty concerning war payoffs that allows for verifiable information transmission. In a first step, the study adds a pre-negotiation phase to the standard crisis bargaining model without verifiable information transmission (4.1). In a second and third step, it considers the model with unilateral (4.2) and bilateral (4.3) verifiable information transmission. It separately analyzes the countries' expected payoffs (4.1-4.3) and the probability of war (4.4). In the end, the study shows that unilateral and bilateral verifiable information transmission is beneficial to the countries in increasing their expected payoffs and that it is consequential in reducing the risk of war.

4.1 Extending the Standard Crisis Bargaining Model

According to the standard take-it-or-leave-it crisis bargaining model (Fearon 1995), country A owns a good that country B would also like to possess. Country B threatens to fight a war over the contested good. However, as war is ex-post inefficient, country A prefers to settle the dispute peacefully. With complete information, country A offers country B's reserve value from fighting, which country B accepts, and consequently no war occurs. With incomplete information about country B's expected war payoff, country A faces a risk-return-tradeoff: country A can make a high offer that country B always accepts, so that war is prevented. Alternatively, country A can make a low offer that only one type of country B accepts and the other type of country B rejects, so that war results. Depending on the belief about country B's type, country A makes a high or low offer.

The present study modifies the standard crisis bargaining model in adding a pre-negotiation phase at time point t_0, in which both countries have uncertainty about the other's *and their own* expected war payoff. At time point t_1, the model proceeds in the standard way. The modification influences country A's threshold for preferring one or the other offer, and affects the calculation of expected payoffs viewed from time point t_0.

The bargaining protocol of the extended model is depicted in Figure 1. At time t_0, Nature draws country A's and country B's type from a probability distribution over two types: country A has low expected war payoffs ($\underline{w_A}$)[14] with probability c_0 and high expected war payoffs ($\overline{w_A}$) with probability $1 - c_0$; and country B has low expected war payoffs ($\underline{w_B}$) with probability a_0 and high expected war payoffs ($\overline{w_B}$) with probability $1 - a_0$, with $1 - \underline{w_B} > 1 - \overline{w_B} > \overline{w_A} > \underline{w_A}$.[15] At time t_0, both countries only know the probabilities (correct prior beliefs), rather than the realization of the other's and their own expected war payoff, whereas at time t_1 they know the realization of their own expected war payoff. At time t_1, country A makes an offer x to country B about the division of a contested good, the value of which is normalized to 1, and country B accepts or rejects the offer. If country B accepts the offer, country B receives x and country A $1 - x$. If country B rejects the offer, both countries fight a war and receive their expected war payoff (which depends on Nature's draw).

If country A knew country B's type in advance, country A would offer country B's reserve value from fighting, $\overline{w_B}$ or $\underline{w_B}$ (plus ϵ), which country B would accept. Thus, with certainty about country B's expected war payoff, no war occurs. However, with uncertainty about country B's war payoff, country A offers $\overline{w_B}$ or $\underline{w_B}$ depending on its prior belief about country B's type. If country A finds country B likely to be strong, it offers $\overline{w_B}$, both types of country B accept the offer, and country A receives $1 - \overline{w_B}$. If country A finds country B likely to be weak, it offers $\underline{w_B}$, country B accepts the offer with probability a_0 (being $\underline{w_B}$), in which case country A receives $1 - \underline{w_B}$, and country B rejects the offer with probability $1 - a_0$ (being $\overline{w_B}$), in which case country A receives its expected war payoff, $\overline{w_A}$ or $\underline{w_A}$ depending on its type. Country A's threshold of beliefs for preferring one or the other offer is calculated separately for both types of country A:

Type $\overline{w_A}$ prefers to offer $\underline{w_B}$ over $\overline{w_B}$ and risk a war if

$$a_0 \cdot (1 - \underline{w_B}) + (1 - a_0) \cdot \overline{w_A} > 1 - \overline{w_B},$$

$$\Rightarrow a_0 > a_2 \equiv \frac{1 - \overline{w_B} - \overline{w_A}}{1 - \underline{w_B} - \overline{w_A}}.$$

Type $\underline{w_A}$ prefers to offer $\underline{w_B}$ over $\overline{w_B}$ and risk a war if

$$a_0 \cdot (1 - \underline{w_B}) + (1 - a_0) \cdot \underline{w_A} > 1 - \overline{w_B},$$

$$\Rightarrow a_0 > a_3 \equiv \frac{1 - \overline{w_B} - \underline{w_A}}{1 - \underline{w_B} - \underline{w_A}}.$$

Thus, there are two thresholds, a_2 and a_3, one for each type of country A, with $a_3 > a_2$. Let us now calculate the countries' expected payoffs in the case of no information transmission for the three situations, $a_0 \leq a_2$, $a_0 > a_3$, and $a_2 < a_0 \leq a_3$, and visualize these expected payoffs in Figure 2(a) for different values of a_0, with c_0 held constant at $c_0 = 0.3$. The other parameter values are $\underline{w_B} = 0.1$, $\overline{w_B} = 0.4$, $\underline{w_A} = 0.1$, and $\overline{w_A} = 0.4$. If $a_0 \leq a_2$, both types of country A offer $\overline{w_B}$, which both types of country B accept. Thus, country A's expected payoff is $1 - \overline{w_B}$, and country B's is $\overline{w_B}$, both constant in a_0. In comparison, if $a_0 > a_3$, both types of country A offer $\underline{w_B}$, which country B accepts if it expects low war payoffs (probability a_0), and rejects if it expects high war payoffs (probability $1 - a_0$). That said, without information transmission and $a_0 > a_3$, country A's expected payoff u_A at t_0 is

$$u_A = a_0 \cdot (1 - \underline{w_B}) + (1 - a_0) \cdot (c_0 \cdot \underline{w_A} + (1 - c_0) \cdot \overline{w_A}),$$

and country B's expected payoff u_B is

$$u_B = a_0 \cdot \underline{w_B} + (1 - a_0) \cdot \overline{w_B}.$$

Upon rearrangement, country A's expected payoff is increasing in a_0, while country B's expected payoff is decreasing in a_0 (see Appendix A and Figure 2(a)). This means that it is desirable for country A if country B likely has low expected war payoffs because country B accepts the low offer, while for country B it is more beneficial if it likely has high expected war payoffs.

Finally, in the third case of $a_2 < a_0 \leq a_3$, country A makes a high offer if it expects low war payoffs (threshold a_3; probability c_0), and country A makes a low offer if it expects high war payoffs (threshold a_2; probability $1 - c_0$). Country B always accepts the high offer, and it accepts the low offer with probability a_0, and rejects it with probability $1 - a_0$. Thus, without information transmission, if $a_2 < a_0 \leq a_3$, country A's expected payoff u_{A*} at t_0 is

$$u_{A*} = c_0 \cdot (1 - \overline{w_B}) + (1 - c_0) \cdot (a_0 \cdot (1 - \underline{w_B}) + (1 - a_0) \cdot \overline{w_A})$$

and country B's expected payoff u_{B*} at t_0 is

$$u_{B*} = c_0 \cdot \overline{w_B} + (1 - c_0) \cdot (a_0 \cdot \underline{w_B} + (1 - a_0) \cdot \overline{w_B}).$$

Thus, country A's expected payoff is still increasing in a_0, though less than in the previous situation of $a_0 > a_3$, and country B's expected payoff is still decreasing in a_0, though again less than previously (see Appendix A and Figure 2(a)).

If $a_0 \leq a_2$, fighting never occurs, and thus, there is no inefficiency. If $a_0 > a_2$, there is a loss of expected surplus due to fighting. Thus, the question of whether the transmission of verifiable information can reduce this loss of expected surplus and make the two countries better off arises. The following sections compare a situation of no verifiable information transmission with one of unilateral verifiable information transmission and of bilateral verifiable information transmission, in order to show that verifiable information transmission pays off under certain conditions.

4.2 Unilateral Verifiable Information Transmission

With unilateral verifiable information transmission, the bargaining protocol changes so that country B can announce a signaling procedure at time t_0, which commits country B to transmitting verifiable information to country A at time t_1. By using a signaling procedure if $a_0 > a_2$, country B intends to influence country A's beliefs about its type in order that country A more often makes a high offer.

Formally, information transmission occurs via a random variable, the value of which correlates with country B's expected war payoff. More precisely, information transmission leads to a randomization of country A's beliefs, sometimes producing one posterior (e.g. the posterior $P(\underline{w_B}) = a_2$) and sometimes producing another posterior (e.g. the posterior $P(\underline{w_B}) = 1$), yet always preserving the mean a_0. The only problem for country B is finding such a random variable that correlates with its war payoff, but as I argue below CBMs can do the trick.

In the following, I consider the two situations of $a_0 > a_3$ and $a_2 < a_0 \leq a_3$ separately.[16] In each situation, I compare two possible information transmission procedures that country B can take, showing that country B's decision for one or the other procedure depends on country B's belief about country A having high or low war payoffs.

Starting with $a_0 > a_3$, assume that country B can decide between two signals, signal z and signal y. Signal z takes the values $\underline{z}$ and $\overline{z}$ and is such that $P(\underline{w_B}|\underline{z}) = a_2$, and

$P(\underline{w_B}|\overline{z}) = 1$. It follows from Bayes rule that $P(\underline{z}) \cdot a_2 + P(\overline{z}) = a_0$[17], which means that country A remains indifferent between its post- and pre-signaling beliefs. Signal y takes the values $\underline{y}$ and $\overline{y}$ and is such that $P(\underline{w_B}|\underline{y}) = a_3$, and $P(\underline{w_B}|\overline{y}) = 1$. It also follows from Bayes rule that $P(\underline{y}) \cdot a_3 + P(\overline{y}) = a_0$, which means that country A again remains indifferent between its post- and pre-signaling beliefs. It is clearly visible that the z-signal induces a larger change in beliefs than the y-signal as $a_2 < a_3$. If country B knew country A's type in advance, it would want to choose the z-signal for type $\overline{w_A}$ (threshold a_2) and the y-signal for type $\underline{w_A}$ (threshold a_3). However, as country B is uncertain about country A's type, the choice depends on country B's belief about the likelihood for one or the other type. Country B decides between the two signals based on its expected payoffs, which are calculated separately for the z-signal and the y-signal as follows:

If $\underline{z}$ is sent, both types of country A believe that country B is $\underline{w_B}$ with probability a_2, and as $a_2 \leq a_2$, they offer $\overline{w_B}$. Both types of country B accept this offer. If $\overline{z}$ is sent, both types of country A believe that country B is $\underline{w_B}$ (for sure), and they offer $\underline{w_B}$ which country B always accepts.

Thus, given the z-signal, the expected payoff u_{Az} for country A at t_0 is

$$u_{Az} = P(\underline{z}) \cdot (1 - \overline{w_B}) + P(\overline{z}) \cdot (1 - \underline{w_B}),$$

and the expected payoff u_{Bz} for country B at t_0 is

$$u_{Bz} = P(\underline{z}) \cdot \overline{w_B} + P(\overline{z}) \cdot \underline{w_B}.$$

Inserting $P(\underline{z}) = \frac{1-a_0}{1-a_2}$ and rearranging the terms, one can show that country A's and country B's expected payoffs are always higher than the expected payoffs without information transmission (see Appendix B and Figure 2(b)). Thus, if the z-signal is sent, unilateral information transmission pays off for country A and for country B. In addition, with the z-signal, country A and country B always settle their dispute peacefully and no war occurs.

On the other hand, if $\underline{y}$ is sent, both types of country A believe that country B is $\underline{w_B}$ with probability a_3. In this case, country A offers $\overline{w_B}$ if its war payoff is $\underline{w_A}$ (threshold a_3, probability c_0), and country A offers $\underline{w_B}$ if its war payoff is $\overline{w_A}$ (threshold a_2, probability $1-c_0$). The offer $\overline{w_B}$ is accepted by both types of country B, while the offer $\underline{w_B}$ is accepted by only one type of country B ($\underline{w_B}$, probability a_3), and therefore war sometimes occurs. If $\overline{y}$ is sent, both types of country A believe that country B is $\underline{w_B}$ (for sure) and they offer $\underline{w_B}$ which country B always accepts.

Thus, given the y-signal, for country A the expected payoff u_{Ay} at t_0 is:

$$u_{Ay} = P(\underline{y}) \cdot \Big(c_0 \cdot (1 - \overline{w_B}) + (1 - c_0) \cdot (a_3 \cdot (1 - \underline{w_B}) + (1 - a_3) \cdot \overline{w_A})\Big) + P(\overline{y}) \cdot (1 - \underline{w_B}),$$

and for country B, the expected payoff u_{By} at t_0 is:

$$u_{By} = P(\underline{y}) \cdot \Big(c_0 \cdot \overline{w_B} + (1 - c_0) \cdot (a_3 \cdot \underline{w_B} + (1 - a_3) \cdot \overline{w_B})\Big) + P(\overline{y}) \cdot \underline{w_B}.$$

Inserting $P(\underline{y}) = \frac{1-a_0}{1-a_3}$ and rearranging the terms, one can show that country A's expected payoff is equal to the expected payoff without information transmission, and that country B's expected payoff is higher than the expected payoff without information transmission (see Appendix C and Figure 2(c)). Thus, unilateral information transmission via the y-signal has no effect on country A and is beneficial for country B. As country B sometimes rejects country A's offer, war still occurs.

When comparing country B's expected payoffs for the z- and the y-signal, one can note that country B's preference for one or the other signal depends on its belief about country A's type (c_0). Country B prefers the z-Signal if $c_0 \leq c_1$, and country B prefers the y-signal if $c_0 > c_1$ with $c_1 = \frac{1-\overline{w_B}-\overline{w_A}}{1-\overline{w_B}-\underline{w_A}}$ (see Appendix D).

Let us continue with the case of $a_2 < a_0 \leq a_3$. Now, country B decides between the z-signal and the w-signal (the y-signal cannot be used because mixing between $\underline{y}$ and

$\overline{y}$ never allows country B to keep country A indifferent between post- and pre-signaling beliefs). The z-signal works like before, and thus generates the same expected payoff function. The w-signal also takes two values, $\underline{w}$ and $\overline{w}$, and is such that $P(\underline{w_B}|\underline{w}) = a_2$, and $P(\underline{w_B}|\overline{w}) = a_3$. Country A again remains indifferent between its pre- and post-signaling beliefs so that $P(\underline{w}) \cdot a_2 + P(\overline{w}) \cdot a_3 = a_0$. Again, the z-signal induces a more far reaching change in beliefs than the w-signal. If $\underline{w}$ is sent, both types of country A believe that country B is type $\underline{w_B}$ with probability a_2, and as $a_2 \leq a_2$, they offer $\overline{w_B}$. Both types of country B accept this offer. If $\overline{w}$ is sent, both types of country A believe that country B is $\underline{w_B}$ with probability a_3, and country A offers $\overline{w_B}$ if it is type $\underline{w_A}$ (threshold a_3, probability c_0), and country A offers $\underline{w_B}$ if it is type $\overline{w_A}$ (threshold a_2, probability $1 - c_0$). The offer $\overline{w_B}$ is accepted by both types of country B, while the offer $\underline{w_B}$ is accepted by only one type of country B, so that war sometimes occurs.

Given the w-signal, country A's expected payoff u_{Aw} at t_0 is:

$$u_{Aw} = P(\underline{w}) \cdot (1 - \overline{w_B}) + P(\overline{w}) \cdot \Big(c_0 \cdot (1 - \overline{w_B})) + (1 - c_0) \cdot (a_3 \cdot (1 - \underline{w_B}) + (1 - a_3) \cdot \overline{w_A})\Big),$$

and country B's expected payoff u_{Bw} at t_0 is:

$$u_{Bw} = P(\underline{w}) \cdot (\overline{w_B}) + P(\overline{w}) \cdot \Big(c_0 \cdot (\overline{w_B})) + (1 - c_0) \cdot (a_3 \cdot (\underline{w_B}) + (1 - a_3) \cdot \overline{w_B})\Big).$$

Inserting $P(\underline{w}) = \frac{a_3 - a_0}{a_3 - a_2}$ and rearranging the terms, it is visible, such as in the case of the y-signal, that unilateral information transmission via the w-signal has no effect for country A and is beneficial for country B (see Appendix C and Figure 2(c)). When comparing country B's expected payoffs for the z- and the w-signal, one can show that the threshold for choosing one or the other signal is the same as previously, c_1. Country B prefers the z-Signal if $c_0 \leq c_1$, and country B prefers the w-signal if $c_0 > c_1$ with $c_1 = \frac{1 - \overline{w_B} - \overline{w_A}}{1 - \overline{w_B} - \underline{w_A}}$. Again, it holds that war is always prevented with the z-signal, while war may still occur with the w-signal.

To sum up, in all cases of unilateral information transmission, one can note that country B is better off with than without unilateral information transmission (Fig. 2(b), 2(c)). By contrast, for country A it depends on the signal sent: when the z-signal is sent (Fig. 2(b)), country A is better off with unilateral information transmission by country B compared to no information transmission, whereas, when the y- and the w-signal is sent (Fig. 2(c)), country A is indifferent between unilateral information transmission by country B and no information transmission.[18] Thus, in the latter case the question of whether country A can be better off if it also announces a signaling strategy arises.

4.3 Bilateral Verifiable Information Transmission

We next allow for bilateral verifiable information transmission by country B and country A. The bargaining protocol changes so that at t_0, after country B has decided to send a signal, country A can also decide to send a signal the value of which will be revealed at t_1. After Nature has determined the value of the signal at t_1, country A makes an offer that country B can accept or reject. As previously argued, if $a_0 > a_2$ and $c_0 \leq c_1$, so that the z-signal is sent, counter-signaling would not be beneficial to country A. However, if $a_0 > a_2$ and $c_0 > c_1$ and the y- or w-signal is sent, war still occurs and unilateral information transmission by country B has no effect for country A. In this situation, country A's motivation for information transmission is to influence country B's beliefs about its type so that country B more often plays the beneficial z-strategy.

Formally, information transmission by country A is also modeled via a random variable. If $a_0 > a_2$ and $c_0 > c_1$, country A sends a signal v that takes the values $\underline{v}$ and $\overline{v}$, and is such that $P(\underline{w_A}|\underline{v}) = 1$, and $P(\underline{w_A}|\overline{v}) = c_1$. This time, country B remains indifferent between its pre- and post-signaling beliefs so that $P(\underline{v}) \cdot 1 + P(\overline{v}) \cdot c_1 = c_0$ (following from Bayes rule). If $\underline{v}$ is sent, country B believes that country A is $\underline{w_A}$ (for sure), and sends the y- or w-signal ($c_0 > c_1$). If $\overline{v}$ is sent, country B believes that country A is $\underline{w_A}$ with probability c_1, and sends the z-signal ($c_1 \leq c_1$). Both countries get their expected payoffs from the respective signals.

Thus, if $c_0 > c_1$, for country A, the expected payoff at t_0 with the v-signal is:

$$u_{Av} = P(\underline{v}) \cdot \Big(P(\underline{y}) \cdot (1 - \overline{w_B}) + P(\overline{y}) \cdot (1 - \underline{w_B})\Big) + P(\overline{v}) \cdot \Big(P(\underline{z}) \cdot (1 - \overline{w_B}) + P(\overline{z}) \cdot (1 - \underline{w_B})\Big)$$

$$\text{if } a_0 > a_3, \text{ and}$$

$$u_{Av*} = P(\underline{v}) \cdot \Big(P(\underline{w}) \cdot (1 - \overline{w_B}) + P(\overline{w}) \cdot (1 - \overline{w_B})\Big) + P(\overline{v}) \cdot \Big(P(\underline{z}) \cdot (1 - \overline{w_B}) + P(\overline{z}) \cdot (1 - \underline{w_B})\Big)$$

$$\text{if } a_2 < a_0 \leq a_3.$$

For country B, the expected payoff at t_0 with the v-signal is:

$$u_{Bv} = P(\underline{v}) \cdot \Big(P(\underline{y}) \cdot \overline{w_B} + P(\overline{y}) \cdot \underline{w_B}\Big) + P(\overline{v}) \cdot \Big(P(\underline{z}) \cdot \overline{w_B} + P(\overline{z}) \cdot \underline{w_B}\Big)$$

$$\text{if } a_0 > a_3, \text{ and}$$

$$u_{Bv*} = P(\underline{v}) \cdot \Big(P(\underline{w}) \cdot \overline{w_B} + P(\overline{w}) \cdot \overline{w_B}\Big) + P(\overline{v}) \cdot \Big(P(\underline{z}) \cdot \overline{w_B} + P(\overline{z}) \cdot \underline{w_B}\Big)$$

$$\text{if } a_2 < a_0 \leq a_3.$$

By inserting the critical terms and rearranging the expressions, one can show that, for country A, bilateral information transmission is better than unilateral information transmission and no information transmission, while for country B, bilateral information transmission is as good as unilateral information transmission, yet better than no information transmission (see Appendix E and Figure 2(d)).

Figures 3(a)-(c) illustrate these findings in a three-dimensional setting when both a_0 and c_0 vary (with $\underline{w_B} = 0.1$, $\overline{w_B} = 0.4$, $\underline{w_A} = 0.1$, and $\overline{w_A} = 0.4$). For country A, unilateral information transmission by country B is beneficial below the critical threshold $c_1 = 0.4$, while bilateral information transmission is beneficial above this critical threshold. For country B, in contrast, unilateral information transmission is beneficial independent of c_0, while bilateral information transmission provides no extra gain compared to unilateral information transmission.[19]

4.4 The Probability of War

I have shown that both countries gain from unilateral information transmission if the z-signal is sent, and that they gain from bilateral information transmission in comparison with no information transmission. Next, I compare the risk of war in the case of no information transmission with the case of unilateral and bilateral information transmission. I show that unilateral information transmission eliminates the risk of war if the z-signal is sent, and reduces the risk of war if the y- or w-signal is sent. Bilateral information transmission always eliminates the risk of war. It generally holds that war only occurs, if country A makes a low offer that country B rejects because it expects high war payoffs. Thus, if $a_0 \leq a_2$, war never occurs because, even without information transmission, both types of country A always make a high offer which country B always accepts. The situation changes for $a_2 < a_0 \leq a_3$ and $a_0 > a_3$ as the following two paragraphs show.

Starting with the case of $a_2 < a_0 \leq a_3$ and no information transmission, country A makes a low offer if it expects high war payoffs, which country B rejects if it also expects high war payoffs. Thus, the probability of war at t_0 is $P(\overline{w_A}) \cdot P(\overline{w_B}) = (1 - c_0) \cdot (1 - a_0)$. If unilateral information transmission takes place, two cases have to be distinguished, $c_0 \leq c_1$ and $c_0 > c_1$. If $c_0 \leq c_1$, country B chooses the z-signal that eliminates the risk of war for the following reason: If the value of the z-signal is $\underline{z}$, both types of country A update the belief to $P(\underline{w_B}|\underline{z}) = a_2$ and make a high offer that country B always accepts, consequently no war occurs. If the value of the z-signal is $\overline{z}$, both types of country A update the belief to $P(\underline{w_B}|\overline{z}) = 1$ and make a low offer that country B again always accepts, as it is the low type for certain. Thus, the z-signal reduces the probability of war to zero.

In the second case, $c_0 > c_1$, country B chooses the w-signal. This signal reduces the risk of war, yet does not eliminate it, for the following reason: if the value of the w-signal is $\underline{w}$, both types of country A update the belief about country B to $P(\underline{w_B}|\underline{w}) = a_2$ and make a high offer that country B always accepts. However, if the value of the w-signal is $\overline{w}$, both types of country A update the belief to $P(\underline{w_B}|\overline{w}) = a_3$. In this situation, one type of country A ($\underline{w_A}$) makes a high offer as its threshold is a_3, while the other type of country A ($\overline{w_A}$) makes a low offer as its threshold is a_2. Country B rejects the low offer if it expects high war payoffs. Thus, the probability of war for $a_2 < a_0 \leq a_3$ and the w-signal at t_0 is $P(\overline{w}) \cdot P(\overline{w_A}) \cdot P(\overline{w_B}|\overline{w}) = \frac{a_0-a_2}{a_3-a_2} \cdot (1-c_0) \cdot (1-a_3)$, which is lower than the probability of war if no information transmission occurs, $(1-c_0) \cdot (1-a_0)$. The probability of war changes to zero, if we add information transmission by country A via the v-signal. If the value of the v-signal is $\overline{v}$, country B updates its belief about country A to $P(\underline{w_A}|\overline{v}) = c_1$ and sends the z-signal, which never results in war, as previously shown. If the value of the v-signal is $\underline{v}$, country B updates its belief about country A to $P(\underline{w_A}|\underline{v}) = 1$ and sends the w-signal randomizing between the beliefs $P(\underline{w_B}|\underline{w}) = a_2$ and $P(\underline{w_B}|\overline{w}) = a_3$. As $P(\underline{w_A}|\underline{v}) = 1$, country A expects low war payoffs for certain and has the threshold a_3. This implies that country A always makes a high offer, which country B always accepts. Thus, with bilateral information transmission, the risk of war changes to zero.

In the second case, if $a_0 > a_3$ and no information transmission occurs, country A always makes a low offer which country B rejects if it expects high war payoffs. Thus, the probability of war at t_0 is $P(\overline{w_B}) = (1-a_0)$. Once we allow for unilateral information transmission and distinguish $c_0 \leq c_1$ and $c_0 > c_1$, the probability of war is zero if $c_0 \leq c_1$, as the z-signal is sent (see above), and the probability of war is larger than zero if $c_0 > c_1$, as the y-signal is sent. More concretely, if the y-signal takes the value $\overline{y}$, both types of country A update the belief about country B to $P(\underline{w_B}|\overline{y}) = 1$ and make a low offer, which country B always accepts as it expects low war payoffs for certain. If the value of the y-signal is $\underline{y}$, both types of country A update the belief to $P(\underline{w_B}|\underline{y}) = a_3$. In this situation, one type of country A ($\underline{w_A}$) makes a high offer as its threshold is a_3, while the other type of country A ($\overline{w_A}$) makes a low offer as its threshold is a_2. Country B rejects this low offer if it expects high war payoffs. Thus, the probability of war with $a_0 > a_3$ and the y-signal at t_0 is $P(\underline{y}) \cdot P(\overline{w_A}) \cdot P(\overline{w_B}|\underline{y}) = \frac{1-a_0}{1-a_3} \cdot (1-c_0) \cdot (1-a_3) = (1-a_0) \cdot (1-c_0)$, which is lower than the probability of war if no information transmission occurs, $(1-a_0)$.

Similarly to the previous case, this risk of war changes to zero, if we add information transmission by country A via the v-signal, for the same reason as above.

To summarize, unilateral information transmission is beneficial to both countries and eliminates the risk of war if the signaling country is likely to be weak ($a_0 > a_2$) and its opponent is likely to be strong ($c_0 \leq c_1$) as this implies the z-signal. In contrast, unilateral information transmission is only beneficial to one country and only reduces the risk of war without eliminating it, if both countries are likely to be weak ($a_0 > a_2$ and $c_0 > c_1$), as this implies the w- or the y-signal. However, as this latter kind of unilateral information transmission has no effect for one country, yet can become beneficial through counter-signaling, it is likely that the country conditions its support for information transmission on its bilateral nature. This completely eliminates the risk of war.

5 Discussing Assumptions and Implications

One of the model's key assumptions is that verifiable information transmission occurs via a randomizing procedure, which allows the signaling state to have perfect influence on its opponent's posterior beliefs, provided that the prior belief is preserved on average. However, if the mean of the posteriors after signaling accords with the mean without signaling, sometimes the opponent finds the signaling state more likely to be strong, and at other times more likely to be weak. This happens due to states a priori committing themselves to the revelation of a certain piece of information, which later results in one of the two posteriors probabilistially. Let me further elaborate on this.

In order to gain a better understanding of what the theoretical mechanism means in practice, let us consider a realistic bargaining scenario and what the random variable might stand for in such a case. Let us assume that both country A and B know that they might be fighting about something in the future, but neither of them know the amount of troops that country B will be able to devote to the fight. The reasons for country B's uncertainty concerning its own troop allocation might be manifold: Country B might be threatened by other adversaries, internal or external, at the same time as the issue with country A emerges, in order that parts of the troops will be occupied and not all troops can be devoted to the issue of contention with country A; there might be an economic crisis or change to a new government with other priorities; alternatively, global politics might change prompting new suppporters to arise and others vanish. Let us keep matters simple and assume with the model that the random variable can only take two values. Accordingly, country A and B might be uncertain about whether country B's military capabilities are high or low. However, both states have the same correct prior belief about the probability, and if this belief is above a critical threshold in order that country B is likely to be weak, country B can only expect a low offer implying a positive risk of war.

In this situation of uncertainty, country B has an incentive to reveal military information to country A. For example, country B might let country A position observers on its ground so that country A might be better informed about the actual troop movements once a contentious issue arises. Subsequently, if troop movements are low, country A is assured that country B will not devote many troops, and the offer will remain low. However, this is inconsequential for country B, as the offer would have been low regardless, given the prior belief. Therefore, upon learning that it is weak, country B has no incentive to deviate from its signaling procedure as the result is the same - and even if it had, there would be practically no possibility of refraining from it, as country A's observers are already on the ground given the prior commitment. On the other hand, it might also be that troop movements are high, indicating that country B more likely devotes a high amount of military capabilities. In this case, country A is induced to make a high offer, which is beneficial to country B compared to the low offer according to the prior belief, and also beneficial to country A given the prevention of war. While one might think that a strong country B would even want to fully reveal information about its strength, given that the offer is already high, there is no additional gain. To the contrary, coming from the model's assumption that the correct prior belief is preserved on average, a change in belief to a posterior left of a_2 (resulting in a higher probability of country B being strong) implies a lower probability of this case occurring in the first place, which is not in country B's interest (see Appendix F).

It becomes apparent through this example that inspections should be tailored to the problem at hand. If states face uncertainty about their future level of troop deployments, like during the Cold War rivalry, on-site inspections of troop manoeuvers is probably the means of choice. However, if states face uncertainty concerning their maritime ambitions, like in the dispute over the Spratly Islands, cooperative monitoring via satellite imagery could represent one way to prevent incidents at sea. Alternatively, if states are unsure whether uranium enrichment facilities will be used for peaceful or aggressive purposes, they might allow observers to such facilities, like in the case of Argentina and Brazil. Despite such measures never providing full certainty about another state's intentions, they yield valuable information that helps to update the belief about the other's type.

The model also identifies the conditions under which we should see no information transmission, unilateral verifiable information transmission or bilateral verifiable information transmission. For each unilateral and bilateral information transmission, the model thus far considers one set of posteriors, calculates the expected payoffs for that case, and compares them to each other in order to determine the superiority of one or the other signaling procedure. However, as previously indicated, states can deliberately choose the posteriors in advance, and they could well have chosen others. Accordingly, the comparison of the signaling procedures only makes sense, if the induced choice of posterior beliefs assumed thus far is optimal from the signaling state's viewpoint. That this optimality is in fact given proves Appendix F. Once the optimality and comparability of the signaling procedures is established, one can illustrate the conditions for one or the other signaling procedure to hold in more detail. This is achieved in the following paragraphs with a brief outline of two cases.

First, the model predicts unilateral information transmission by a country if that country is likely to be weak ($a_0 > a_2$) and its opponent is likely to be strong ($c_0 \leq c_1$) as this implies the z-signal, which is beneficial to both parties. It is worth noting the following about the timing of unilateral information transmission: assuming that we start with a situation of no information transmission, which implies that the signaling party likely was strong ($a_0 \leq a_2$), it must be that the signaling party has lost military power shortly before it took a signaling stance.

This logic of unilateral information transmission is reflected in the Cold War rivalry between the United States and Soviet Union. Despite this rivalry not being characterized purely by unilateral information transmission, nevertheless it is clearly evident that the weaker side, the Soviet Union, made much larger concessions with respect to confidence-building in general and verifiable information transmission in particular than the stronger side, the United States (Leffler 2007, 392). For example, when the Soviet Union agreed on a moratorium on nuclear testing, the United States doubted the verifiability of the nuclear halt. In response, the Soviet Union unilaterally allowed the United States to set up monitoring stations close to the testing range on Soviet soil, in order to allow for better verifiability of the halt (Kydd 2005, 225f.). Another example for the Soviet Union's willingness to signal is the Intermediate-Range Nuclear Forces Treaty, where Gorbachev wanted to proceed by destroying far more missiles than the United States, and by allowing unilaterally for intrusive on-site inspections in order to assure the United States of a radical change in Soviet policy (Kydd 2005, 227f.; Hanes, Hanes, and Baker 2003, 354; and Brown 2010, 263). Other unilateral confidence-building measures by the Soviet Union included the withdrawal from Afghanistan, unilateral troop reductions, and the permission of Eastern regimes to fall, in order to indicate a lack of territorial ambitions (Leffler 2007, 403, 421, 437). All such initiatives were taken at a point in time when the Soviet Union had arguably lost power: Starting in the 1970s, the Soviet Union had suffered from severe economic decline and the burdens of empire (Kydd 2005, 216; Leffler 2007, 350).

Second, the model predicts bilateral information transmission if both countries are likely to be weak ($c_0 > c_1$ and $a_0 > a_2$) as this implies the y- and w-signaling strategy by country B, which country A counters with the v-signaling strategy in order to induce country B to more often play the z-signaling strategy. The following claim can be made as regarding the timing of bilateral information transmission: assuming that we start with a situation of no information transmission or unilateral information transmission, bilateral information transmission is a likely outcome if at least one country looses military power, i.e. if it loses in military capabilities, devalues the contested good, or has higher costs of fighting a war.

The Argentina-Brazil rivalry exemplifies this logic. In the 1970s, Argentina and Brazil competed for regional influence (Child 1985, 98). Both countries started to develop nuclear programs aimed at self-sufficiency, and both were embroiled in a border struggle over hydropower resources (Child 1985, 100f.). Given their opposed ambitions and uncertainty about their military power, militarized conflict due to misperception or miscalculation could not be excluded (Lipson 2003, 144). At the end of the 1970s and the beginning of the 1980s, however, the political situation changed, with Argentina and Brazil facing severe economic downturns and civil unrests, Argentina emerged as a weak party from the lost Falkland war, Brazil had strained relations to its former strong ally, the United States, and the military dictatorships were coming to an end (Lipson 2003, 144). In short, both countries, particularly Argentina, expected less from fighting a war with each other than before, which opened the possibility of confidence-building steps.

In this situation, Argentina advanced with confidence-building gestures, which Brazil reciprocated. In 1983, Argentina briefed Brazil about the development of a gaseous diffusion enrichment facility before this development became public, and Brazil made a similar gesture in 1987. In 1987, Argentinian president Alfonsin invited Brazilian president Sarney to visit the unsafeguarded Pilcaniyeu gaseous diffusion facility as a first verification step, and this measure was taken despite Brazil having knownly not informed Argentina about its nuclear test site at Cachimbo, which was denied until 1990. However, Brazil reciprocated Argentina's benevolent gesture by counter-inviting Alfonsin to the official inauguration of the Aramar gas centrifuge facility (Redick 1999, 228f.). A third significant step in verifiable information transmission followed in 1990, with

Brazil's economy further deteriorating: Argentina and Brazil signed the Foz de Iguacu Declaration on the Common Nuclear Policy, formally renouncing nuclear weapons and establishing a framework for the implementation of a bilateral nuclear accounting and inspection arrangement and full-scope IAEA safeguards. This arrangement led to the mutual inspection of nuclear facilities (Lipson 2003, 144).

Third, the model predicts that the risk of war disappears with unilateral and bilateral information transmission, which is supported by the two previous cases. There has been no direct military confrontation between the United States and Soviet Union or Russia since the mid-1980s, and regarding the Argentina-Brazil case, one author assessed that there was "not even a remote possibility of an armed clash involving these two countries" (Child 1985, 104).

6 Conclusion

I presented a game-theoretic take-it-or-leave-it crisis bargaining model that allows for unilateral and bilateral verifiable information transmission. My contribution is two-fold. From the perspective of verifiable information transmission as the key aspect of confidence-building measures, I contend that we regularly observe this phenomenon empirically, yet do not know much about the conditions for verifiable information transmission occurring in the first place and being successful in reducing the risk of war in the second place. Formal models can inform us about these two conditions, and thus help us to address an empirical puzzle. In contrast, when considering the study's contribution from the perspective of the take-it-or-leave-it bargaining model, this model poses the question of how states can prevent the ex-post inefficient outcome of war. Existing solutions relax the unitary actor assumption or introduce third parties. By contrast, this study highlights that unilateral or bilateral verifiable information transmission succeed in preventing the ex-post inefficient outcome of war, and thus offers a solution to the bargaining problem without relaxing the unitary actor assumption or introducing third parties.

The analysis predicts unilateral verifiable information transmission if the signaling state expects low payoffs and the other expects high payoffs from fighting. It predicts bilateral verifiable information transmission if both states expect low payoffs from fighting. Finally, verifiable information transmission successfully reduces the risk of war when it induces a sufficiently large change in beliefs. Interestingly, these predictions are very similar to the predictions of Kydd (2005). Kydd holds that confidence-building takes place if both states have a sufficient level of trust, i.e. they believe each other likely to be security seekers rather than expansionists, and that confidence-building successfully prevents the occurrence of war if the confidence-building steps are large enough (Kydd 2005, 193, 197). However, unlike the present study, Kydd models confidence-building as a cooperative gesture in the first stage of a 2x2 conflict game with uncertainty about motivations. In contrast, the present study explicitly formalizes the most prevalent aspect of confidence-building, the transmission of verifiable information about military aspects, and thus focuses on a different aspect of CBMs. However, a systematic comparison of the two models and an assessment of which holds superiority stands out for future research.

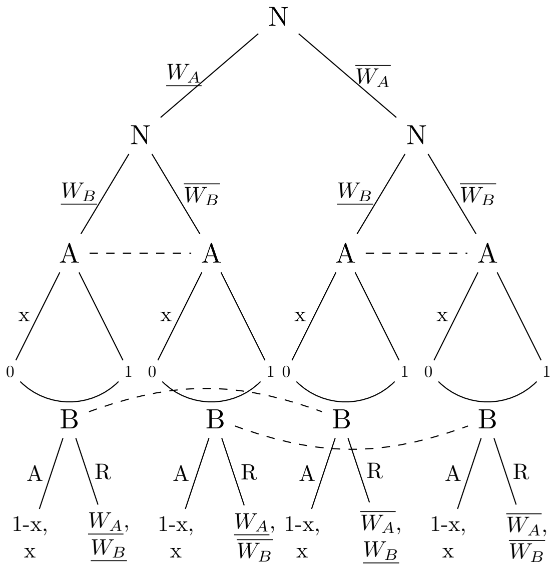

Figure 1: Extended Version of the Standard Crisis Bargaining Model

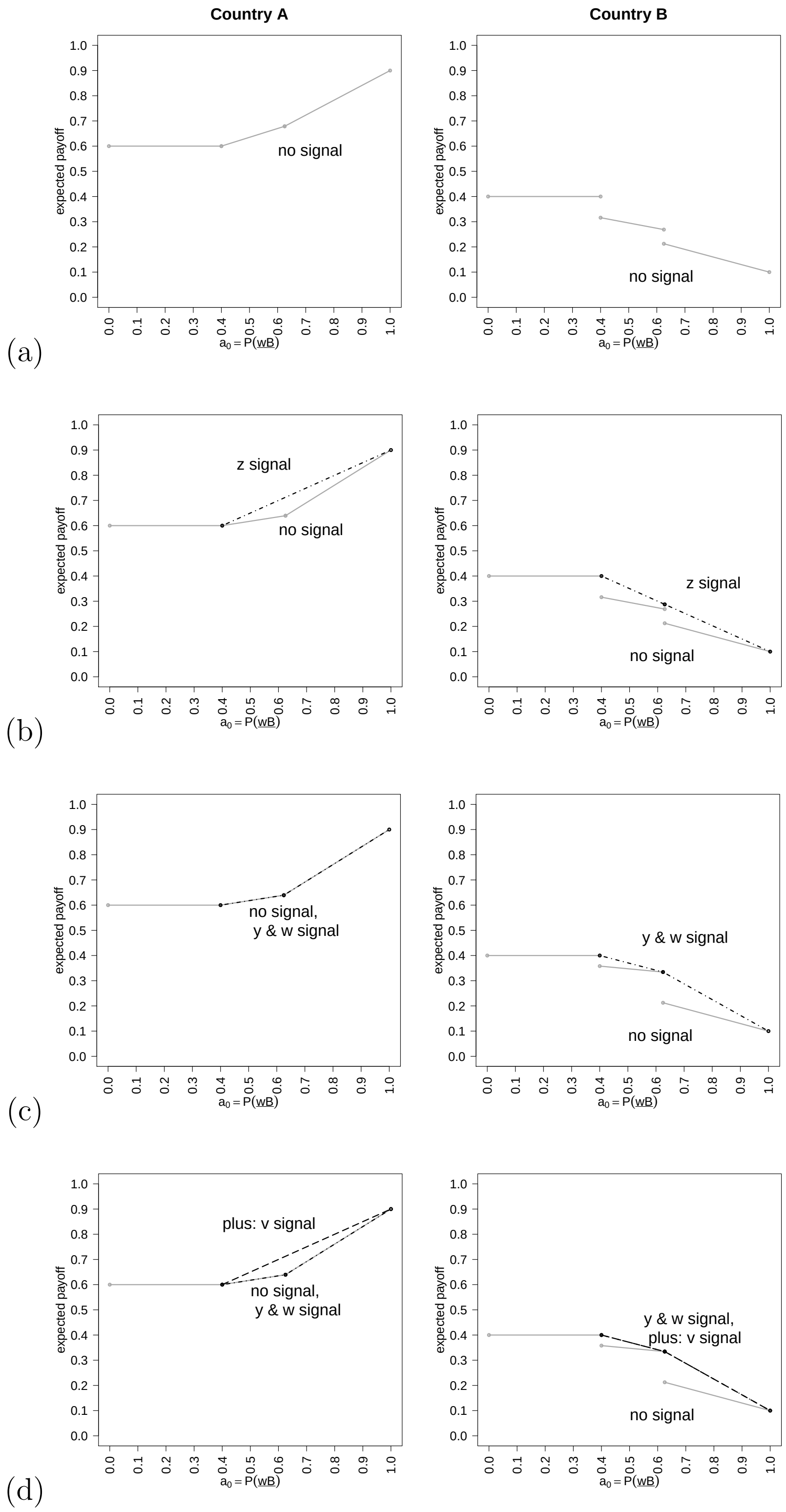

Figure 2: Expected payoffs for country A and country B with a) no signal ($c_0 = 0.3$), b) z signal ($c_0 = 0.3$), c) y & w signal ($c_0 = 0.65$), d) plus: v signal ($c_0 = 0.65$)

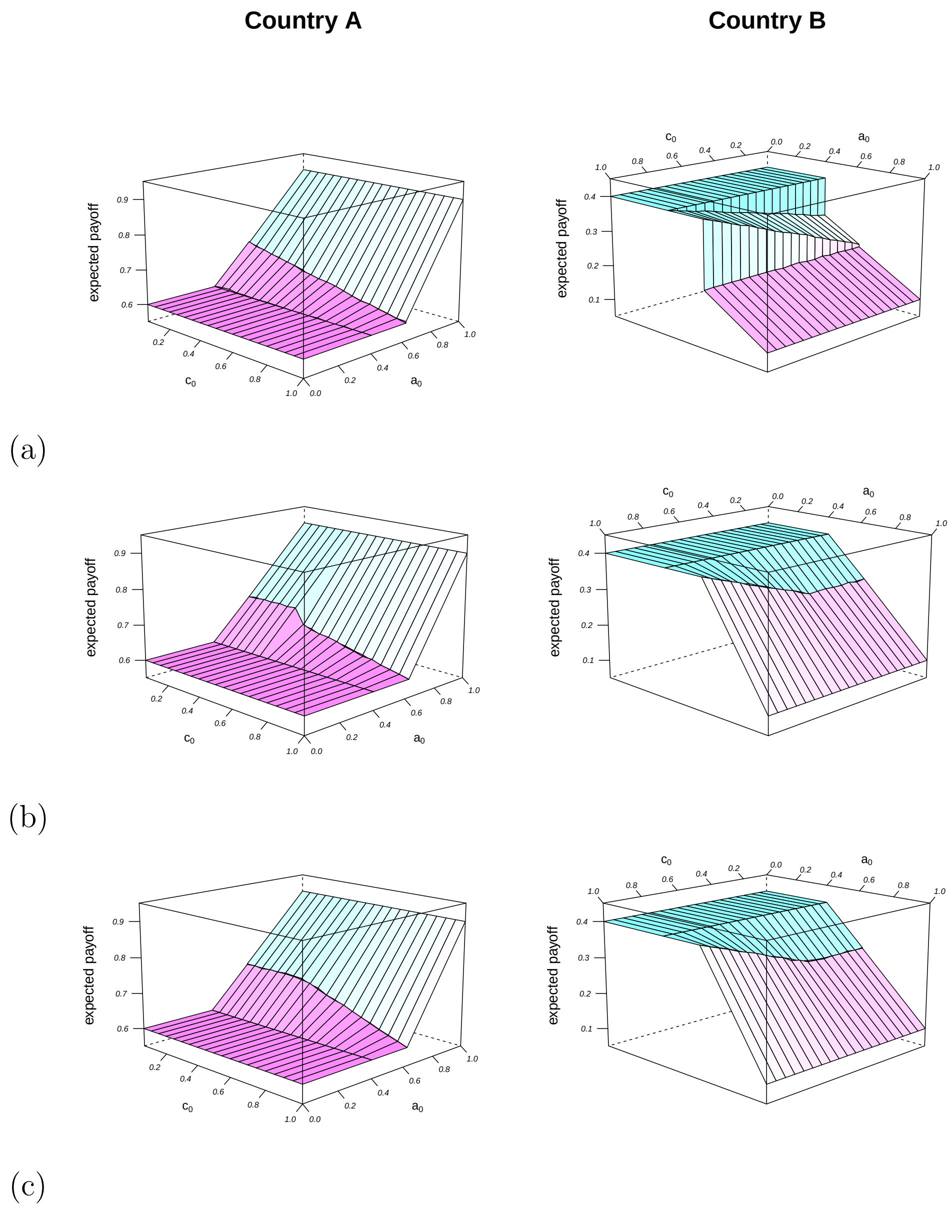

Figure 3: Expected payoffs for country A with a) no information transmission, b) unilateral information transmission, and c) bilateral information transmission in a three-dimensional setting

Notes

[1]These later measures are also known under the term 'Confidence- and Security-Building Measures' (CSBMs).

[2]Some recent studies examine the incentives of state to create, instead of eliminate, strategic ambiguity; see Meirowitz and Sartori (2008) and Baliga and Sjöström (2008).

[3]In the words of Alford (1981, 135f.):"Indeed, signalling is what CBMs are mainly concerned with. They are intended to help separate unambiguous signals of hostile intent from the random noise of continuous military activity. They are devices to filter out all extraneous events, leaving behind certain indicators that, either singly or in combination, provide hard information that a state is preparing for war."

[4]Unilateral CBMs are measures taken by one state alone vis a vis another state, a group of states, or the international community. Bilateral CBMs commit two states to reciprocally sharing military information, whereas multilateral measures commit either two states with the help of a third party, or more than two states. Of the three types of CBMs, unilateral CBMs are the least known. They have been discussed in the context of the Cold War (Krepon 1999, 1), and were recently mentioned with respect to the China-Taiwan dispute (Glosserman 2005, v), the Iran crisis (http://www.presstv.ir/detail/234770.html), and in relation to space security (Robinson, Schaefer, Schrogel, and Dunk von der 2011, 7).

[5]Cp. http://www.un.org/disarmament/convarms/infoCBM/.

[6]Some authors have offered a broad conceptualization of CBMs, including non-military cultural and economic CBMs, such as people-to-people contact or the lifting of travel and trade barriers. In this study, CBMs include all measures that inform about the countries' expected war payoff. At first sight, this seems to only include military CBMs, however, as the expected war payoff also consists of the costs of fighting, such as opportunity costs of foregone trade, this study does not constrain CBMs to the military dimension.

[7]Arguably, the absence of MIDs in these dyads is also due to realist or liberalist influences, such as the balance of power, or joint democracy, trade, and international organizations. In a case study or quantitative analysis, one would carefully have to control for these influences and examine whether CBMs have an extra pacifying effect.

[8]Solutions to the commitment problem are not discussed here because this is not the present study's focus, however for an example see Chadefaux (2011).

[9]Note that these studies model costly signaling in a defender-challenger-framework, and not in the present study's crisis bargaining framework.

[10]This argument was later qualified by Slantchev (2010) showing that it sometimes pays off for the resolved type not to use the signal and instead pretend to be weak. Although this does not necessarily discourage the other party from challenging in the first place, pretending to be weak can increase the chance of victory if a war is fought, as it prompts the other party to mobilize less resources.

[11]It is not quite clear why the audience would punish a government for backing down, if such an action is rational for the country. In order to make sense of punishment by an audience, some authors introduce a reputation effect, like Sartori (2002), and combine it with audience costs, like Guisinger and Smith (2002).

[12]Divergent is Slantchev (2005), who argues that militarization has a dual role.

[13]Relatedly, Meirowitz and Sartori (2008) examine whether perfect monitoring, presumably via a third party, can eliminate the risk of war. Though the authors show that there is a beneficial effect of perfect monitoring, they concede that such a scenario is unrealistic and that the beneficial effect vanishes once only a little bit of noise is introduced.

[14]Expected war payoffs (w) are usually calculated as the probability of winning (p) times the value of the contested good (here normalized to 1), minus the cost of fighting (c) which typically varies for the two players. Thus: $w_A = p - c_A$ for one player and $w_B = 1 - p - c_B$ for the other player. States can be uncertain about the cost of fighting, the value of the good, or the probability of winning. Uncertainty about the probability of winning makes calculations more complicated, thus, for simplification, most authors assume uncertainty about costs or about the value of the contested good (see Fearon (1995), div. Wittman (2009)).

[15]Assuming that $\overline{w_A} = p - \underline{c_A}$, $\underline{w_A} = p - \overline{c_A}$, $\overline{w_B} = 1 - p - \underline{c_B}$, and $\underline{w_B} = 1 - p - \overline{c_B}$, it follows that $1 - \underline{w_B} = p + \overline{c_B} > 1 - \overline{w_B} = p + \underline{c_B} > \overline{w_A} = p - \underline{c_A} > \underline{w_A} = p - \overline{c_A}$.

[16]For $a_0 \leq a_2$, there is no inefficiency due to fighting, which is why such a case is not considered here. In this situation, country A would nevertheless like to know more about country B's type, but as country B can expect a high offer from country A, it has no incentive to reveal information about its type. Thus, no information transmission occurs.

[17]

$$P(\underline{z}) \cdot a_2 + P(\overline{z}) \cdot 1 =$$

$$P(\underline{z}) \cdot P(\underline{w_B}|\underline{z}) + P(\overline{z}) \cdot P(\underline{w_B}|\overline{z}) =$$

$$P(\underline{w_B}) \cdot P(\underline{z}|\underline{w_B}) + P(\underline{w_B}) \cdot P(\overline{z}|\underline{w_B}) =$$

$$P(\underline{w_B}) = a_0$$

[18]Although the z-signal induces country A to sometimes make a high offer, which provides lower payoffs to country A than a low offer, country A nevertheless gains from the z-signal. The reason is that country A is better informed and can thus make a more appropriate offer which prevents the inefficient outcome of war. However, with the y- or w-signal, country A has to make a high offer too often and thus the beneficial effect of information transmission disappears.

[19]Figure 3(a) shows that country A's expected payoff is independent of c_0 if $a_0 \leq a_2$ or decreases with c_0 if $a_0 > a_2$, and that country B's expected payoff is independent of c_0 if $a_0 \leq a_2$ and $a_0 > a_3$ or increases with c_0 if $a_2 < a_0 \leq a_3$. The interpretation is that with c_0 increasing (and $a_2 < a_0 \leq a_3$), country A can expect less from fighting, which is bad for country A yet good for country B, as country B more often receives a high offer.

References

Alford, J. (1981). The Usefulness and the Limitations of CBMs. *New Directions in Disarmament*, 134.

Baliga, S. and T. Sjöström (2008). Strategic Ambiguity and Arms Proliferation. *Journal of Political Economy 116*(6), 1023–1057.

Black, J. and G. Bulkley (1988). The Role of Strategic Information Transmission in a Bargaining Model. *The Economic Journal 98*(390), 50–57.

Boehmer, C., E. Gartzke, and T. Nordstrom (2004). Do Intergovernmental Organizations Promote Peace? *World Politics 57*(1), 1–38.

Brown, A. (2010). *The Gorbachev Revolution and the End of the Cold War*, Volume 3, Chapter 12, pp. 244–266. Cambridge University Press.

Chadefaux, T. (2011). Bargaining over Power: When do Shifts in Power Lead to War? *International Theory 3*(2), 228–253.

Child, J. (1985). *Geopolitics and Conflict in South America: Quarrels among Neighbors.* Praeger New York.

Crawford, V. P. and J. Sobel (1982). Strategic Information Transmission. *Econometrica: Journal of the Econometric Society 50*(6), 1431–1451.

Desjardins, M.-F. (1996). *Rethinking Confidence-Building Measures: Obstacles to Agreement and the Risks of Overselling the Process.* Oxford University Press for the International Institute for Strategic Studies.

Devabhaktuni, S., M. C. Rudolph, and M. Newbill (1999). *Key Developments in CBMs between India and Pakistan*, Chapter 9, pp. 189–196. In Krepon et al. Krepon et al. (1999).

Fearon, J. D. (1994). Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review 88*(3), 577–592.

Fearon, J. D. (1995). Rationalist Explanations for War. *International Organization 49*(3), 379–379.

Fearon, J. D. (1997). Signaling Foreign Policy Interests. *Journal of Conflict Resolution 41*(1), 68–90.

Fisher, C. S. (1999). *The Preconditions of Confidence Building: Lessons from the European Experience*, Chapter 18, pp. 291–312. In Krepon et al. Krepon et al. (1999).

Glosserman, B. (2005). Cross-Strait Confidence Building Measures. *Issues and Insights 5*(2).

Guisinger, A. and A. Smith (2002). Honest Threats. *Journal of Conflict Resolution 46*(2), 175–200.

Hanes, S. M., R. C. Hanes, and L. W. Baker (2003). *Cold War: Almanac*, Volume 2. UXL.

Kapur, S. P. (2005). India and Pakistan's Unstable Peace: Why Nuclear South Asia is not like Cold War Europe. *International Security 30*(2), 127–152.

Krepon, M. (1999). *Conflict Avoidance, Confidence Building, and Peacemaking.* In Krepon et al. Krepon et al. (1999).

Krepon, M., M. Newbill, K. Khoja, and J. S. Drezin (1999). *Global Confidence Building: New Tools for Troubled Regions.* Palgrave Macmillan.

Kurizaki, S. (2007). Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy. *American Political Science Review 101*(03), 543–558.

Kydd, A. (2003). Which Side are You on? Bias, Credibility, and Mediation. *American Journal of Political Science 47*(4), 597–611.

Kydd, A. H. (1997). Game Theory and the Spiral Model. *World Politics 49*, 371–400.

Kydd, A. H. (2005). *Trust and Mistrust in International Relations.* Princeton University Press.

Kydd, A. H. (2010). Rationalist Approaches to Conflict Prevention and Resolution. *Annual Review of Political Science 13*, 101–121.

Lachowski, Z. (2005). *Confidence-and Security-Building Measures in the New Europe*, Volume 18. SIPRI.

Leffler, M. P. (2007). *For the Soul of Mankind: the United States, the Soviet Union, and the Cold War.* Hill and Wang.

Lipson, C. (2003). *Reliable Partners: How Democracies Have Made a Separate Peace.* Princeton University Press.

Mearsheimer, J. J. (1994/1995). The False Promise of International Institutions. *International Security 19*(3), 5–49.

Meirowitz, A. and A. E. Sartori (2008). Strategic Uncertainty as a Cause of War. *Quarterly Journal of Political Science 3*(4), 327–352.

Oneal, J. and B. Russett (2001). *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: Norton.

Ramsay, K. W. (2011). Cheap Talk Diplomacy, Voluntary Negotiations, and Variable Bargaining Power. *International Studies Quarterly 55*, 1003–1023.

Rauchhaus, R. W. (2006). Asymmetric Information, Mediation, and Conflict Management. *World Politics 58*(2), 207–241.

Redick, J. R. (1999). *Nuclear Confidence Building in the Southern Cone*, Chapter 12, pp. 227–244. In Krepon et al. Krepon et al. (1999).

Robinson, J., M. P. Schaefer, K.-U. Schrogel, and F. Dunk von der (2011). Transparency and Confidence-Building Measures for Space Security. *Space Policy 27*(1), 27–37.

Sartori, A. E. (2002). The Might of the Pen: A Reputational Theory of Communication in International Disputes. *International Organization 56*(1), 121–149.

Schultz, K. A. (2001). Looking for Audience Costs. *Journal of Conflict Resolution 45*(1), 32–60.

Slantchev, B. L. (2005). Military Coercion in Interstate Crises. *American Political Science Review 99*(4), 533–547.

Slantchev, B. L. (2010). Feigning Weakness. *International Organization 64*(3), 357–388.

Trager, R. F. (2011). Multidimensional Diplomacy. *International Organization 65*(3), 469–506.

United Nations (1982). Comprehensive Study on Confidence-Building Measures. Technical report, United Nations Publication.

Wittman, D. (2009). Bargaining in the Shadow of War: When is a Peaceful Resolution Most Likely? *American Journal of Political Science 53*(3), 588–602.

Appendices

A Expected Payoffs Without Verifiable Information Transmission

Without information transmission and $a_0 > a_3$, country A's expected payoff u_A can be rearranged to read

$$u_A = ((1 - c_0) \cdot \overline{w_A} + c_0 \cdot \underline{w_A}) + (1 - \underline{w_B} - \overline{w_A} + c_0 \cdot (\overline{w_A} - \underline{w_A})) \cdot a_0.$$

It holds that $u_A(1) = 1 - \underline{w_B}$, and $\frac{\partial u_A}{\partial a_0} > 0$.

Without information transmission and $a_2 < a_0 \leq a_3$, country A's expected payoff u_{A*} can be rearranged to read

$$u_{A*} = ((1 - c_0) \cdot \overline{w_A} + c_0 \cdot (1 - \overline{w_B})) + (1 - \underline{w_B} - \overline{w_A} - c_0 \cdot (1 - \underline{w_B} - \overline{w_A})) \cdot a_0.$$

It holds that $u_{A*}(a_2) = 1 - \overline{w_B}$, and $\frac{\partial u_{A*}}{\partial a_0} < \frac{\partial u_A}{\partial a_0}$. Moreover, $u_A(a_3) = u_{A*}(a_3)$ as the following calculation shows:

$$u_A(a_3) = u_{A*}(a_3)$$

$$\Rightarrow ((1 - c_0) \cdot \overline{w_A} + c_0 \cdot \underline{w_A}) + (1 - \underline{w_B} - \overline{w_A} + c_0 \cdot (\overline{w_A} - \underline{w_A})) \cdot a_3$$

$$= ((1 - c_0) \cdot \overline{w_A} + c_0 \cdot (1 - \overline{w_B})) + (1 - \underline{w_B} - \overline{w_A} - c_0 \cdot (1 - \underline{w_B} - \overline{w_A})) \cdot a_3$$

$$\Rightarrow a_3 = \frac{1 - \overline{w_B} - \underline{w_A}}{1 - \underline{w_B} - \underline{w_A}},$$

which is true.

Without information transmission and $a_0 > a_3$, country B's expected payoff u_B can be rearranged to read

$$u_B = \overline{w_B} - (\overline{w_B} - \underline{w_B}) \cdot a_0.$$

It holds that $u_B(1) = \underline{w_B}$, and $\frac{\partial u_B}{\partial a_0} < 0$.

Without information transmission and $a_2 < a_0 \leq a_3$, country B's expected payoff u_{B*} can be rearranged to read

$$u_{B*} = \overline{w_B} - (1 - c_0) \cdot (\overline{w_B} - \underline{w_B}) \cdot a_0.$$

It holds that $u_{B*}(a_2) < \overline{w_B}$, and $\frac{\partial u_{B*}}{\partial a_0} > \frac{\partial u_B}{\partial a_0}$. Moreover, $u_{B*}(a_3) > u_B(a_3)$ as the following calculation shows:

$$u_{B*}(a_3) > u_B(a_3)$$

$$\Rightarrow \overline{w_B} - (1 - c_0) \cdot (\overline{w_B} - \underline{w_B}) \cdot a_3 > \overline{w_B} - (\overline{w_B} - \underline{w_B}) \cdot a_3$$

$$\Rightarrow -c_0 < 0,$$

which is true.

B Expected Payoffs with Unilateral Information Transmission, z-Signal

Show that country A's expected payoff is higher with the z-signal than without signaling. With the z-signal, country A's expected payoff is

$$u_{Az} = \frac{1 - a_0}{1 - a_2} \cdot (1 - \overline{w_B}) + \frac{a_0 - a_2}{1 - a_2} \cdot (1 - \underline{w_B}).$$

It holds that $u_{Az}(a_2) = 1 - \overline{w_B}$, and $u_{Az}(1) = 1 - \underline{w_B}$. Country A's expected payoff can be rearranged to read

$$u_{Az} = \frac{1}{1 - a_2} \cdot (1 - \overline{w_B} - a_2 \cdot (1 - \underline{w_B})) + \frac{1}{1 - a_2} \cdot (\overline{w_B} - \underline{w_B}) \cdot a_0.$$

It holds that $\frac{\partial u_{Az}}{\partial a_0} < \frac{\partial u_A}{\partial a_0}$ as the following calculation shows:

$$\frac{\partial u_{Az}}{\partial a_0} < \frac{\partial u_A}{\partial a_0}$$

$$\Rightarrow \frac{1}{1 - a_2} \cdot (\overline{w_B} - \underline{w_B}) < 1 - \underline{w_B} - \overline{w_A} + c_0 \cdot (\overline{w_A} - \underline{w_A})$$

$$\Rightarrow 1 - \underline{w_B} - \overline{w_A} < 1 - \underline{w_B} - \overline{w_A} + c_0 \cdot (\overline{w_A} - \underline{w_A}),$$

which is true.

Given the same end point, it follows that the z-signal provides a higher expected payoff to country A than no signaling.

Show that country B's expected payoff is higher with the z-signal than without signaling.

With the z-signal, country B's expected payoff is

$$u_{Bz} = \frac{1-a_0}{1-a_2} \cdot \overline{w_B} + \frac{a_0 - a_2}{1-a_2} \cdot \underline{w_B}.$$

It holds that $u_{Bz}(a_2) = \overline{w_B} > u_{B*}(a_2) = \overline{w_B} - (1-c_0) \cdot (\overline{w_B} - \underline{w_B}) \cdot a_2$, and $u_{Bz}(1) = \underline{w_B}$. Country B's expected payoff can be rearranged to read

$$u_{Bz} = \frac{1}{1-a_2} \cdot (\overline{w_B} - a_2 \cdot \underline{w_B}) - \frac{1}{1-a_2} \cdot (\overline{w_B} - \underline{w_B}) \cdot a_0.$$

It holds that $u_{Bz}(a_3) = \frac{1-a_3}{1-a_2} \cdot \overline{w_B} + \frac{a_3-a_2}{1-a_2} \cdot \underline{w_B}$ which is higher than the expected payoff without information transmission as the following calculation shows:

$$u_{Bz}(a_3) = \frac{1-a_3}{1-a_2} \cdot \overline{w_B} + \frac{a_3 - a_2}{1-a_2} \cdot \underline{w_B} > u_{B*}(a_3) = \overline{w_B} - (1-c_0) \cdot (\overline{w_B} - \underline{w_B}) \cdot a_3$$

$$\Rightarrow a_3 + a_2 < (1-c_0) \cdot a_3 \cdot (1-a_2)$$

$$\Rightarrow (1 - \overline{w_B} - \overline{w_A}) \cdot (\underline{w_B} - \overline{w_B}) < 0,$$

which is true.

In summary, it follows that the z-signal provides a higher expected payoff to country B than no signaling.

C Expected Payoffs with Unilateral Information Transmission, y-&w-Signal

Show that the y- and the w-signal provide the same expected payoff to country A as no signaling. With the y-signal ($a_0 > a_3$), country A's expected payoff is

$$u_{Ay} = \frac{1-a_0}{1-a_3} \cdot \Big(c_0 \cdot (1 - \overline{w_B}) + (1-c_0) \cdot (a_3 \cdot (1 - \underline{w_B}) + (1-a_3) \cdot \overline{w_A})\Big) + \frac{a_0 - a_3}{1-a_3} \cdot (1 - \underline{w_B}).$$

It holds that $u_{Ay}(1) = 1 - \underline{w_B}$. Country A's expected payoff can be rearranged to read

$$u_{Ay} = \frac{1}{1-a_3} \cdot \Big(c_0 \cdot (1 - \overline{w_B}) + (1-c_0) \cdot (a_3 \cdot (1 - \underline{w_B}) + (1-a_3) \cdot \overline{w_A}) - a_3 \cdot (1 - \underline{w_B})\Big)$$

$$- \frac{1}{1-a_3} \cdot \Big(c_0 \cdot (1 - \overline{w_B}) + (1-c_0) \cdot (a_3 \cdot (1 - \underline{w_B}) + (1-a_3) \cdot \overline{w_A}) - (1 - \underline{w_B})\Big) \cdot a_0.$$

The slope can be simplified to read

$$\frac{\partial u_{Ay}}{\partial a_0} = -\Big((1-c_0)\cdot\overline{w_A} - 1 + \underline{w_B} - c_0\cdot(1-\underline{w_A}) + c_0\Big),$$

which is equal to the slope of no signaling

$$\frac{\partial u_A}{\partial a_0} = 1 - \underline{w_B} - \overline{w_A} + c_0\cdot(\overline{w_A} - \underline{w_A}).$$

Given the same end point and slope, country A's expected payoff for the y-signal and no signaling are the same.

With the w-signal ($a_2 < a_0 \leq a_3$), country A's expected payoff is

$$u_{Aw} = \frac{a_0 - a_3}{a_2 - a_3}\cdot(1-\overline{w_B}) + \frac{a_2 - a_0}{a_2 - a_3}\cdot\Big(c_0\cdot(1-\overline{w_B}) + (1-c_0)\cdot(a_3\cdot(1-\underline{w_B}) + (1-a_3)\cdot\overline{w_A})\Big).$$

It holds that $u_{Aw}(a_2) = 1 - \overline{w_B}$. Country A's expected payoff can be rearranged to read

$$u_{Aw} = \frac{a_2}{a_2 - a_3}\cdot\Big(c_0\cdot(1-\overline{w_B}) + (1-c_0)\cdot(a_3\cdot(1-\underline{w_B}) + (1-a_3)\cdot\overline{w_A})\Big) - \frac{a_3}{a_2 - a_3}\cdot(1-\overline{w_B})$$

$$-\frac{1}{a_2 - a_3}\cdot\Big(c_0\cdot(1-\overline{w_B}) + (1-c_0)\cdot(a_3\cdot(1-\underline{w_B}) + (1-a_3)\cdot\overline{w_A}) - (1-\overline{w_B})\Big)\cdot a_0.$$

The slope can be simplified to read

$$\frac{\partial u_{Aw}}{\partial a_0} = (1-c_0)\cdot(1-\overline{w_A} - \underline{w_B}),$$

which is equal to the slope of no signaling

$$\frac{\partial u_{A*}}{\partial a_0} = 1 - \underline{w_B} - \overline{w_A} - c_0\cdot(1 - \underline{w_B} - \overline{w_A}).$$

Given the same starting point and slope, country A's expected payoffs for the w-signal and no signaling are the same.

Show that the y- and the w-signal provide higher expected payoffs for country B than no signaling. With the y-signal ($a_0 > a_3$), country B's expected payoff is

$$u_{By} = \frac{1-a_0}{1-a_3}\cdot\Big(c_0\cdot(\overline{w_B}) + (1-c_0)\cdot(a_3\cdot\underline{w_B} + (1-a_3)\cdot\overline{w_B})\Big) + \frac{a_0 - a_3}{1-a_3}\cdot\underline{w_B}.$$

It holds that $u_{By}(1) = \underline{w_B}$. Country B's expected payoff can be rearranged to read

$$u_{By} = \frac{1}{1-a_3} \cdot \Big(c_0 \cdot \overline{w_B} + (1-c_0) \cdot (a_3 \cdot \underline{w_B} + (1-a_3) \cdot \overline{w_B}) - a_3 \cdot \underline{w_B}\Big)$$

$$-\frac{1}{1-a_3} \cdot \Big(c_0 \cdot \overline{w_B} + (1-c_0) \cdot (a_3 \cdot \underline{w_B} + (1-a_3) \cdot \overline{w_B}) - \underline{w_B}\Big) \cdot a_0.$$

The slope can be simplified to read

$$\frac{\partial u_{By}}{\partial a_0} = -\overline{w_B} + \underline{w_B} - (1 - \overline{w_B} - \underline{w_A}) \cdot c_0,$$

which is smaller (and thus steeper, due to being negative) than the slope of no signaling

$$\frac{\partial u_B}{\partial a_0} = -\overline{w_B} + \underline{w_B}.$$

With the w-signal ($a_2 < a_0 \leq a_3$), country B's expected payoff is

$$u_{Bw} = \frac{a_0 - a_3}{a_2 - a_3} \cdot \overline{w_B} + \frac{a_2 - a_0}{a_2 - a_3} \cdot \Big(c_0 \cdot \overline{w_B} + (1-c_0) \cdot (a_3 \cdot \underline{w_B} + (1-a_3) \cdot \overline{w_B})\Big).$$

It holds that $u_{Bw}(a_2) = \overline{w_B}$ which is higher than the expected payoff without signaling

$$u_{B*}(a_2) = \overline{w_B} - (1-c_0) \cdot (\overline{w_B} - \underline{w_B}) \cdot a_2.$$

It also holds that

$$u_{Bw}(a_3) = c_0 \cdot \overline{w_B} + (1-c_0) \cdot \Big(a_3 \cdot \underline{w_B} + (1-a_3) \cdot \overline{w_B}\Big),$$

which is the same as the expected payoff without signaling

$$u_{B*}(a_3) = \overline{w_B} - (1-c_0) \cdot a_3 \cdot (\overline{w_B} - \underline{w_B}).$$

D Country B's Preference for Signals

Country B prefers to send the z-signal over the y-signal, if the former's expected payoff exceeds the latter:

$$P(\underline{z}) \cdot \overline{w_B} + P(\overline{z}) \cdot \underline{w_B} \geq P(\underline{y}) \cdot \Big(c_0 \cdot \overline{w_B} + (1-c_0) \cdot (a_3 \cdot \underline{w_B} + (1-a_3) \cdot \overline{w_B})\Big) + P(\overline{y}) \cdot \underline{w_B}$$

$$\Rightarrow \frac{a_0 - 1}{a_2 - 1} \cdot \overline{w_B} + \frac{a_2 - a_0}{a_2 - 1} \cdot \underline{w_B} \geq \frac{a_0 - 1}{a_3 - 1} \cdot \Big(c_0 \cdot \overline{w_B} + (1-c_0) \cdot (a_3 \cdot \underline{w_B} + (1-a_3) \cdot \overline{w_B})\Big) + \frac{a_3 - a_0}{a_3 - 1} \cdot \underline{w_B}$$

$$\Rightarrow \frac{a_2 \cdot (a_3 - 1)}{a_2 - 1} \cdot \overline{w_B} + \frac{(1 - a_0) \cdot (a_3 - a_2)}{(a_2 - 1) \cdot (a_0 - 1)} \cdot \underline{w_B} - a_3 \cdot \underline{w_B} \geq c_0 \cdot a_3 \cdot (\overline{w_B} - \underline{w_B})$$

$$\Rightarrow \frac{a_2 \cdot (a_3 - 1)}{a_2 - 1} \cdot (\overline{w_B} - \underline{w_B}) \geq c_0 \cdot a_3 \cdot (\overline{w_B} - \underline{w_B})$$

$$\Rightarrow c_0 \leq \frac{a_2 \cdot (a_3 - 1)}{a_3 \cdot (a_2 - 1)}$$

$$\Rightarrow c_0 \leq \frac{1 - \overline{w_B} - \overline{w_A}}{1 - \overline{w_B} - \underline{w_A}}.$$

E Expected Payoffs with Bilateral Information Transmission

Show that the v-signal provides higher expected payoffs to country A than no (and unilateral) signaling. With the v-signal and $a_0 > a_3$, country A's expected payoff is

$$u_{Av} = \frac{c_0 - c_1}{1 - c_1} \cdot \Big(\frac{1 - a_0}{1 - a_3} \cdot (1 - \overline{w_B}) + \frac{a_0 - a_3}{1 - a_3} \cdot (1 - \underline{w_B})\Big) +$$

$$\frac{1 - c_0}{1 - c_1} \cdot \Big(\frac{1 - a_0}{1 - a_2} \cdot (1 - \overline{w_B}) + \frac{a_0 - a_2}{1 - a_2} \cdot (1 - \underline{w_B})\Big)$$

It holds that $u_{Av}(1) = 1 - \underline{w_B}$. The slope is

$$\frac{\partial u_{Av}}{\partial a_0} = -\frac{1 - c_0}{1 - c_1} \cdot \frac{1}{1 - a_2} \cdot (\underline{w_B} - \overline{w_B}) - \frac{c_0 - c_1}{1 - c_1} \cdot \frac{1}{1 - a_3} \cdot (\underline{w_B} - \overline{w_B}) > 0$$

$$\Rightarrow \frac{1 - c_0}{1 - c_1} \cdot (1 - \underline{w_B} - \overline{w_A}) + \frac{c_0 - c_1}{1 - c_1} \cdot (1 - \underline{w_B} - \underline{w_A})$$

$$\Rightarrow 1 - \underline{w_B} - \frac{1 - c_0}{1 - c_1} \cdot \overline{w_A} - \frac{c_0 - c_1}{1 - c_1} \cdot (\underline{w_A}),$$

which is smaller than the slope without signaling, as the following calculation shows:

$$\frac{\partial u_{Av}}{\partial a_0} = 1 - \underline{w_B} - \frac{1 - c_0}{1 - c_1} \cdot \overline{w_A} - \frac{c_0 - c_1}{1 - c_1} \cdot (\underline{w_A}) < \frac{\partial u_A}{\partial a_0} = 1 - \underline{w_B} - \overline{w_A} + c_0 \cdot (\overline{w_A} - \underline{w_A}).$$

$$\Rightarrow c_0 - c_1 < c_0 \cdot (1 - c_1)$$

$$\Rightarrow c_0 < 1,$$

which is true.

Given the same end point, it follows that bilateral information transmission provides higher expected payoffs than no information transmission.

With the v-signal and $a_2 < a_0 \leq a_3$, country A's expected payoff is

$$u_{Av*} = \frac{c_0 - c_1}{1 - c_1} \cdot \Big(\frac{1 - a_0}{1 - a_2} \cdot (1 - \overline{w_B}) + \frac{a_0 - a_2}{1 - a_2} \cdot (1 - \overline{w_B})\Big)$$

$$+\frac{1 - c_0}{1 - c_1} \cdot \Big(\frac{a_3 - a_0}{a_3 - a_2} \cdot (1 - \overline{w_B}) + \frac{a_0 - a_2}{a_3 - a_2} \cdot (1 - \underline{w_B})\Big).$$

It holds that $u_{Av*}(a_2) = 1 - \overline{w_B}$. The slope is

$$\frac{\partial u_{Av*}}{\partial a_0} = -\frac{1 - c_0}{1 - c_1} \cdot \frac{1}{1 - a_2} \cdot (\underline{w_B} - \overline{w_B}),$$

$$\Rightarrow \frac{1 - c_0}{1 - c_1} \cdot (1 - \underline{w_B} - \overline{w_A}),$$

which is larger than the slope without signaling

$$\frac{\partial u_{A*}}{\partial a_0} = (1 - c_0) \cdot (1 - \underline{w_B} - \overline{w_A}).$$

Given the same starting point, it follows that bilateral signaling provides higher expected payoffs to country A than no signaling.

Show that the v-signal provides higher expected payoffs to country B than no signaling and the same expected payoffs as unilateral signaling. With the v-signal and $a_0 > a_3$, country B's expected payoff is

$$u_{Bv} = \frac{c_0 - c_1}{1 - c_1} \cdot \Big(\frac{1 - a_0}{1 - a_3} \cdot \overline{w_B} + \frac{a_0 - a_3}{1 - a_3} \cdot \underline{w_B}\Big) + \frac{1 - c_0}{1 - c_1} \cdot \Big(\frac{1 - a_0}{1 - a_2} \cdot \overline{w_B} + \frac{a_0 - a_2}{1 - a_2} \cdot \underline{w_B}\Big)$$

It holds that $u_{Bv}(1) = \underline{w_B}$. The slope is

$$\frac{\partial u_{Bv}}{\partial a_0} = -\frac{1 - c_0}{1 - c_1} \cdot \frac{1}{1 - a_2} \cdot (\overline{w_B} - \underline{w_B}) - \frac{c_0 - c_1}{1 - c_1} \cdot \frac{1}{1 - a_3} \cdot (\overline{w_B} - \underline{w_B}) < 0$$

$$\Rightarrow \frac{\partial u_{Bv}}{\partial a_0} = -(1 - \underline{w_B}) + \frac{1 - c_0}{1 - c_1} \cdot \overline{w_A} + \frac{c_0 - c_1}{1 - c_1} \cdot \underline{w_A},$$

which is smaller than the slope without signaling, as the following calculation shows:

$$\frac{\partial u_{Bv}}{\partial a_0} = -(1 - \underline{w_B}) + \frac{1 - c_0}{1 - c_1} \cdot \overline{w_A} + \frac{c_0 - c_1}{1 - c_1} \cdot \underline{w_A} < \frac{\partial u_B}{\partial a_0} = -(\overline{w_B} - \underline{w_B})$$

$$\Rightarrow \frac{1 - c_0}{1 - c_1} \cdot \overline{w_A} + \frac{c_0 - c_1}{1 - c_1} \cdot \underline{w_A} < 1 - \underline{w_B},$$

which is true.

Given the same end point, it follows that bilateral information transmission provides a higher expected payoff to country B than no signaling. At the same time, the slope for bilateral signaling is equal to the slope of unilateral signaling, as the following calculation shows:

$$\frac{\partial u_{Bv}}{\partial a_0} = -(1-\underline{w_B}) + \frac{1-c_0}{1-c_1}\cdot\overline{w_A} + \frac{c_0-c_1}{1-c_1}\cdot\underline{w_A} = \frac{\partial u_{By}}{\partial a_0} = -(1-\overline{w_B}-\underline{w_A})\cdot c_0 - \overline{w_B} + \underline{w_B}$$

$$\Rightarrow c_1 = \frac{1-\overline{w_B}-\overline{w_A}}{1-\overline{w_B}-\underline{w_A}},$$

which is true.

Thus, bilateral and unilateral information transmission provide the same expected payoffs to country B.

With the v-signal and $a_2 < a_0 \leq a_3$, country B's expected payoff is

$$u_{Bv*} = \frac{c_0-c_1}{1-c_1}\cdot\Big(\frac{a_3-a_0}{a_3-a_2}\cdot\overline{w_B} + \frac{a_0-a_2}{a_3-a_2}\cdot\overline{w_B}\Big) + \frac{1-c_0}{1-c_1}\cdot\Big(\frac{1-a_0}{1-a_2}\cdot\overline{w_B} + \frac{a_0-a_2}{1-a_2}\cdot\underline{w_B}\Big)$$

It holds that $u_{Bv*}(a_2) = \overline{w_B} = u_{Bw}(a_2) > u_{B*}(a_2)$. Country B's expected payoff can be rearranged to read

$$u_{Bv*} = \frac{1-c_0}{1-c_1}\cdot(1-\overline{w_A}) + \frac{c_0-c_1}{1-c_1}\cdot\overline{w_B} - \frac{1-c_0}{1-c_1}\cdot(1-\underline{w_B}-\overline{w_A})\cdot a_0.$$

It holds that $u_{Bv*}(a_3) = u_{B*}(a_3)(= u_{Bw}(a_3))$, as the following calculation shows:

$$u_{Bv*}(a_3) = \frac{1-c_0}{1-c_1}\cdot(1-\overline{w_A}) + \frac{c_0-c_1}{1-c_1}\cdot\overline{w_B} - \frac{1-c_0}{1-c_1}\cdot(1-\underline{w_B}-\overline{w_A})\cdot a_3 >$$

$$u_{B*}(a_3) = \overline{w_B} - (1-c_0)\cdot(\overline{w_B}-\underline{w_B})\cdot a_3$$

$$\Rightarrow (1-a_3)\cdot(1-\overline{w_A}-\overline{w_B}) > a_3\cdot c_1\cdot(\overline{w_B}-\underline{w_B})$$

$$\Rightarrow c_1 = \frac{1-\overline{w_B}-\overline{w_A}}{1-\overline{w_B}-\underline{w_A}},$$

which is true.

Given the higher starting point and the same end point, it follows that bilateral information transmission provides higher expected payoffs to country B than no information transmission. Given the same starting and end point, the expected payoffs of bilateral and one-sided information transmission are the same.

F Optimal Posterior Beliefs

In the following, I show that the posterior beliefs for the z-, y-, w-, and v-signal, which were so far assumed in the text, are optimal from the signaling state's viewpoint.

The z-signal mixes between the posterior beliefs $P(\underline{w_B}|\underline{z}) = a_2$ and $P(\underline{w_B}|\overline{z}) = 1$. If, a), the change in belief through $\underline{z}$ is larger than previously assumed (left to a_2), it formally holds that $P(\underline{w_B}|\underline{z}) = a_2 - x$ with $0 < x \leq a_2$. In this case, both types of country A make a high offer, but this occurs with less frequency than with no deviation from the originally assumed posterior belief as $P(\underline{z}) = \frac{1-a_0}{1-a_2+x}$ is decreasing in x. Consequently, country B's expected payoff decreases with x, so that country B would like to maintain the deviation as small as possible, preferably at 0. If, b), the change in belief through $\underline{z}$ is smaller than previously assumed (right to a_2), it formally holds that $P(\underline{w_B}|\underline{z}) = a_2 + x$ with $0 < x \leq a_3 - a_2$, and $P(\underline{z}) = \frac{1-a_0}{1-a_2-x}$. In this case, only one type of country A ($\underline{w_A}$, threshold a_3) makes a large offer, while the other type of country A ($\overline{w_A}$, threshold a_2) makes a small offer. Country B rejects the small offer and starts to fight a war if it expects high war payoffs, which is the case with updated probability $1 - (a_2 + x)$. Thus, country B expects high payoffs with probability $P(\underline{z}) \cdot (c_0 + (1 - c_0) \cdot (1 - (a_2 + x))) = \frac{1-a_0}{1-a_2-x} \cdot \Big(c_0 + (1 - c_0) \cdot (1 - a_2 - x)\Big) = \frac{1-a_0}{1-a_2-x} \cdot c_0 + (1 - a_0) \cdot (1 - c_0)$. This probability of high payoffs is increasing in x up to the point that $x = a_3 - a_2$ so that $P(\underline{w_B}|\underline{z}) = a_3$, whereby the probability of high payoffs is $\frac{1-a_0}{1-a_3} \cdot c_0 + (1 - a_0) \cdot (1 - c_0)$. Nonetheless, this probability of high war payoffs is smaller than the probability of high payoffs without a deviation, as the following calculation shows:

$$\frac{1-a_0}{1-a_3} \cdot c_0 + (1-a_0) \cdot (1-c_0) < \frac{1-a_0}{1-a_2}$$

$$\Rightarrow c_0 < \frac{a_2}{1-a_2} \cdot \frac{1-a_3}{a_3}$$

$$\Rightarrow c_0 < \frac{1 - \overline{w_B} - \overline{w_A}}{1 - \overline{w_B} - \underline{w_A}},$$

which is true if we start out with the z-signal, as we did.

It follows that country B prefers a posterior belief of a_2 over a posterior belief right to a_2. If, c), the change in belief through $\overline{z}$ is smaller than previously assumed (left to 1), it formally holds that $P(\underline{w_B}|\overline{z}) = 1 - x$ with $0 < x \leq 1 - a_3$, and $P(\underline{z}) = \frac{1-a_0-x}{1-a_2-x}$. A posterior belief $P(\underline{w_B}|\overline{z}) = 1 - x$ implies a positive probability of high war payoffs if $\overline{z}$ is sent. Thus, the overall probability of high payoffs is $P(\underline{z}) + P(\overline{z}) \cdot x = \frac{1-a_0-x\cdot(1-a_0+a_2)}{1-a_2-x}$.

This probability of high payoffs is decreasing in x up to the point that $x = 0(+\epsilon)$. Thus, country B prefers a posterior belief of 1 over a posterior belief left to 1.

The y-signal mixes between the posterior beliefs $P(\underline{w_B}|\underline{y}) = a_3$ and $P(\underline{w_B}|\overline{y}) = 1$. a) A belief to the left of a_3, $P(\underline{w_B}|\underline{y}) = a_3 - x$ with $0 < x \leq a_3 - a_2$, corrsponds to a situation of a belief to the right of a_2 in case of the z-signal, which implies that the probability of high payoffs is decreasing in x up to the point that $P(\underline{w_B}|\underline{z}) = a_3$. Thus, the deviation should be kept as small as possible in order to maximize country B's expected payoffs. b) A belief to the right of a_3, $P(\underline{w_B}|\underline{y}) = a_3 + x$ with $0 < x \leq a_0 - a_3$, which implies $P(\underline{y}) = \frac{1-a_0}{1-a_3-x}$, induces none of the two types of A to make a high offer. Thus, the probability of high payoffs is $P(\underline{y}) \cdot (1 - a_3 - x) = \frac{1-a_0}{1-a_3-x} \cdot (1 - a_3 - x) = 1 - a_0$, which is equal to the probability of high payoffs without information transmission. c) A belief to the left of 1, $P(\underline{w_B}|\overline{y}) = 1 - x$ with $0 < x \leq 1 - a_0$, implies $P(\underline{y}) = \frac{1-a_0-x}{1-a_3-x}$ and a probability of high payoffs that is $P(\underline{y}) \cdot (c_0 + (1-c_0) \cdot a_3) + P(\overline{y}) \cdot x = \frac{1-a_0-x}{1-a_3-x} \cdot (c_0 + (1-c_0) \cdot (1-a_3)) + \frac{a_0-a_3}{1-a_3-x} \cdot x$, which is decreasing in x. Thus, from country B's point of view, the deviation should be kept as small as possible, preferably at zero.

The w-signal mixes between the posterior beliefs $P(\underline{w_B}|\underline{w}) = a_2$ and $P(\underline{w_B}|\overline{w}) = a_3$. a) A belief to the left of a_2, $P(\underline{w_B}|\underline{w}) = a_2 - x$ with $0 < x \leq a_2$, implies that $P(\underline{w}) = \frac{a_3-a_0}{a_3-a_2+x}$, which is decreasing in x. Given that $\underline{w}$ yields higher expected payoffs for country B than $\overline{w}$, country B wants to keep a deviation from a_2 as small as possible, preferably at zero. b) A belief to the right of a_2, $P(\underline{w_B}|\underline{w}) = a_2 + x$ with $0 < x \leq a_0 - a_2$, implies that the probability of high payoffs is $P(\underline{w}) \cdot (c_0 + (1-c_0) \cdot (1-a_2-x)) + P(\overline{w}) \cdot (c_0 + (1-c_0) \cdot (1-a_3)) = \frac{a_3-a_0}{a_3-a_2-x} \cdot (c_0 + (1-c_0) \cdot (1-a_2-x)) + (1 - \frac{a_3-a_0}{a_3-a_2-x}) \cdot (c_0 + (1-c_0) \cdot (1-a_3)) = c_0 + (1-c_0) \cdot (1-a_0)$, which is the same probability of high payoffs as without information transmission, and thus, the signaling surplus has vanished. c) A belief to the left of a_3, $P(\underline{w_B}|\overline{w}) = a_3 - x$ with $0 < x \leq a_3 - a_0$, implies that $P(\underline{w}) = \frac{a_3-a_0-x}{a_3-a_2-x}$, which is decreasing in x. Given that $\underline{w}$ yields higher expected payoffs than $\overline{w}$, a lower probability of the former is not desirable for country B. d) A belief to the right of a_3, $P(\underline{w_B}|\overline{w}) = a_3 + x$ with $0 < x \leq 1 - a_3$, yields a probability of high payoffs that is $P(\underline{w}) + P(\overline{w}) \cdot (1 - (a_3 + x)) = 1 - \frac{a_0-a_2}{a_3-a_2+x} \cdot (a_3 + x)$. This probability is increasing in x, but is lower than the probability of high payoffs with the original version of the w-signal, as the following calculation shows (when assuming a maximal deviation of x, which is $1 - a_3$):

$$1 - \frac{a_0 - a_2}{a_3 - a_2 + (1 - a_3)} \cdot (a_3 + (1 - a_3)) < P(\underline{w}) + P(\overline{w}) \cdot (c_0 + (1 - c_0) \cdot (1 - a_3))$$

$$\Rightarrow 1 - \frac{a_0 - a_2}{1 - a_2} \cdot (1) < 1 - \frac{a_0 - a_2}{a_3 - a_2} \cdot (1 - c_0) \cdot a_3$$

$$\Rightarrow \frac{1}{1 - a_2} > \frac{a_3}{a_3 - a_2} \cdot (1 - c_0)$$

$$\Rightarrow c_0 > \frac{a_2 \cdot (1 - a_3)}{a_3 \cdot (1 - a_2)}$$

$$\Rightarrow c_0 > \frac{1 - \overline{w_B} - \overline{w_A}}{1 - \overline{w_B} - \underline{w_A}}$$

which is true if we start out with the w-signal, as we did.

The v-signal mixes between the posterior beliefs $P(\underline{w_A}|\overline{v}) = c_1$ and $P(\underline{w_A}|\underline{v}) = 1$. a) A belief to the left of c_1 implies that the z-signal is still played by country B, but less often, as $P(\overline{v}) = \frac{1-c_0}{1-c_1+x}$ is decreasing in x. As country A prefers $\overline{v}$ over $\underline{v}$, this is disadvantageous for country A. b) A belief to the right of c_1 implies that the y- and w-signal is played instead of the z-signal. The probability $P(\overline{v}) = \frac{1-c_0}{1-c_1-x}$ is increasing in x, but at the maximum value of x, which is $c_0 - c_1$ and implies a posterior belief of c_0 and $P(\overline{v}) = 1$, the expected payoff is the same as without information transmission. Thus, the additional gains of information transmission have vanished. c) Finally, a belief to the left of 1, $P(\underline{w_A}|\underline{v}) = 1 - x$ with $0 < x \leq 1 - c_0$, implies that $P(\underline{v}) = \frac{c_0 - c_1}{1-c_1-x}$. This probability is increasing in x, but as country A prefers $\overline{v}$ over $\underline{v}$, country A would like to keep the deviation as small as possible.