Mediation and Peace*

Johannes Hörner† Massimo Morelli‡ Francesco Squintani§

First version, May 2010;
this version, November 2013

Abstract

This paper applies mechanism design to the study of international conflict resolution. Standard mechanisms in which an arbitrator can enforce her decisions are usually not feasible because disputants are sovereign entities. Nevertheless, we find that this limitation is inconsequential. Despite only being capable of making non-binding recommendations, we find that mediators are equally effective as arbitrators. We determine optimal mediation techniques. They consist in not precisely reporting information to conflicting parties, and specifically, in not revealing to a player with probability one that the opponent is weak. These obfuscation techniques strictly improve the chance of peace when the intensity of conflict is high, or when asymmetric information is significant.

*We thank Sandeep Baliga, Alberto Bisin, Yeon-Koo Che, Matt Jackson, Philippe Jehiel, Philipp Kircher, Adam Meirowitz, Clara Ponsati, Kris Ramsay and Tomas Sjöström for their insightful comments, and Yingni Guo and Michael Rubin for excellent research assistance. The audiences of the ESSET 2009, SAET 2009, Wallis Conference 2010, conflict and mechanism design conferences at Princeton, Warwick, Madrid and UCL, and regular seminars at Cambridge, Columbia, Essex, EUI, Bristol, Melbourne, Oxford, Sydney, UNSW, Siena, Toulouse School of Economics, University of Pennsylvania, UCL, UCLA and Warwick are gratefully acknowledged. This paper also benefited from various conversations with Mr. Gianluca Esposito, Principal Administrator at the Council of Europe, and currently Consulting Counsel at the International Monetary Fund (IMF) (the views expressed by Gianluca Esposito do not necessarily reflect the views of the Council of Europe, of the IMF or of any of their respective bodies).

†Yale University, Department of Economics.

‡Columbia University, Department of Political Science and Department of Economics.

§University of Warwick, Department of Economics.

1 Introduction

Over the years, the formal theory of international relations has developed the positive analysis of conflict by making use of advanced game theoretical techniques.[1] On the contrary, the powerful tools of mechanism design have not yet been extensively used to explore which institutions may be more effective for conflict resolution and prevention.[2] The revelation principle, a fundamental result in mechanism design due to Myerson (1979, 1982), identifies mediation as an efficient institution to deal with conflicts that arise because of asymmetric information, one of the main rationalist explanations for wars.[3] Indeed, mediation has played an increasingly important role in the organization of peace talks to resolve recent international crises. According to the International Crisis Behavior (ICB) project, 30% of international crises for the entire period 1918–2001 were mediated, and the fraction rises to 46% for the period 1990–2001 (see Wilkenfeld et al., 2005).

Outside international relations, the fact that private information may cause bargaining failures has been invoked as an explanation for costly trials in the case of litigation, and strikes in the case of wage bargaining (see Kennan and Wilson, 1993, for an early review). In these types of 'conflicts', it is possible to use the standard mechanisms identified by the revelation principle by Myerson (1979): An arbitrator with enforcement power (the power of law or of the State), collects information from the disputants privately and then makes binding recommendations. However, the key distinguishing feature of international crises or disputes is that the players involved are sovereign entities, and hence there is no legitimate or recognized third party to which they can credibly delegate decision and enforcement power

[1]See Jackson and Morelli (2011) for an updated survey of such positive analysis.

[2]A few papers do exist that study mechanism design in international relations. See e.g. Fey and Ramsay (2009) and all the other papers in that special issue. A discussion of the importance of institutional design for conflict resolution or international cooperation is in Koremenos et al (2001).

[3]Blainey (1988) famously argued that wars begin when states disagree about their relative power and end when they agree again (see also, Brito and Intriligator, 1985, and Fearon, 1995). Wars may arise because of asymmetric information about military strength, but also about the value of outside options or about the contestants' political resolve, i.e., about the capability of the leaders and the peoples to sustain war. For example, it is known that Saddam Hussein grossly under-estimated the US administration political resolve, when invading Kuwait in 1990.

(see e.g. Waltz, 1959). For this reason, any mechanism design model aimed to deal with international relations has to dispense with the assumption that the third party can enforce her decisions and, instead, focus on self enforcing mechanisms. This paper aims to compare the effectiveness of forms of third party intervention that are used in international relations, in the absence of enforcement power and under the assumption that the sole objective of an unbiased mediator should be the maximization of the probability of peace.[4]

We consider mediation, as well as also a less potent form of third party intervention for dispute resolution, which we call 'communication facilitation,' following the terminology by Touval and Zartman (1985). By the version of the revelation principle by Myerson (1982), optimal mediation requires that the mediator collects information privately in closed door meetings. This practice, often called 'shuttle diplomacy' requires the mediator to meet the disputants privately, and to keep the communications private, before organizing a summit in which proposals will be discussed.[5] A communication facilitator, instead, does not hold bilateral closed door meetings: the content of any communication is open to all the disputants.[6]

The specific simple model of conflict we consider comprises two players who contest a fixed amount of resources.[7] A player cannot observe the opponent's strength, political resolve, or willingness to fight. Specifically, each player is strong (hawk) with some probability and weak (dove) otherwise. If the two players are of the same type, war is a fair lottery; else, the stronger wins with higher probability. For simplicity, we assume that all wars are equally costly.[8] We

[4]Bester and Warneryd (2006) study the optimal arbitration mechanism that maximizes the probability of a peaceful solution in a stylized conflict game. We adopt their model of conflict, but study also mediators who cannot enforce their decisions. Among the few papers studying self-enforcing mechanisms in contexts different from international relations, see Matthews and Postlewaite (1989), Banks and Calvert (1992), Cramton and Palfrey (1995), Forges (1999), Compte and Jehiel (2008), and Goltsman et al. (2009).

[5]See, for example, Kydd (2006) and Fey and Ramsay (2010) for detailed discussions of shuttle diplomacy. Even when the press describes a mediation attempt as if the parties were communicating at a high frequency, like for example in the famous Camp David peace conference, the actual negotiations often follow the pattern of shuttle diplomacy. See e.g. Stein (1999).

[6]For example, the WTO settlement procedure requires explicitly full transparency of all communications among disputants. See in particular items 5,6,10 in the Annex 2 of the WTO procedural rules.

[7]This is a standard metaphor for many types of wars, for example those related to territorial disputes or to the present and future sharing of the rents from the extraction of natural resources. Indeed, Bercovitch et al (1991) show that mediation is useful mostly when the disputes are about resources, territory, or in any case divisible issues.

[8]It might be interesting to allow for different costs for symmetric and asymmetric wars, but the additional

consider revelation mechanisms in which the two players report their types to the third party, who then renders decisions. In the most potent form of third party intervention, arbitration, participation is voluntary, reports are private and decisions are binding. Under mediation, reports are private but decisions are only non binding recommendations. Hence, participation is voluntary after the mediator presents her proposals and not only at the beginning of the mediation. When the third party only facilitates communication, reports are open to both disputants and, again, decisions are not binding. With each form of third party intervention, we calculate the equilibrium that maximizes the *ex ante* chances of peace.

Our application of mechanism design to international conflict allows us to answer three different questions, respectively, on whether the limitations of mediation relative to arbitration matter, on what the optimal mediation strategies should be, and on when shuttle diplomacy strictly improves over open facilitated communication.

We find that optimal mediation can achieve *the same* probability of peace as optimal arbitration, in spite of the absence of enforcement power. Because the mediator is incapable of enforcing settlements that yield the hawk a lower payoff than war when meeting a dove, she confuses the hawk by not always revealing that it is meeting a dove when this is the case. By doing so, the mediator manages to give the hawk facing a dove, in expectation, exactly the same payoff as the arbitrator.

Second, we derive optimal mediation strategies: the mediator should use shuttle diplomacy to *obfuscate* the information received by disputants. Effectively, she should avoid revealing that one player is weak to the other with probability one. When the intensity of conflict is low and asymmetric information is significant, the mediator should not always reveal to hawks when they are facing a dove. Whereas when the intensity of conflict is low, the mediator should obfuscate self-reported doves.

Our final result is a full characterization of the conditions under which optimal mediation with shuttle diplomacy strictly dominates communication facilitation. We find that this is

notational and computational costs appear a heavy price to pay.

the case when the intensity of conflict is high, or asymmetric information is significant.[9,10] When this is not the case, obfuscation strategies are not needed to achieve the same welfare as arbitration.

We conclude this introduction by briefly motivating our general modelling choices. In line with the mechanism design literature, we consider unbiased mediators who have no private information.[11] Further, the mediator's objective is the minimization of the *ex ante* probability of war. Hence, our mediator must be able to commit to quit in some circumstances, instead of seeking a peaceful agreement in all contingencies (see Watkins, 1998). Such commitments, in fact, facilitate information disclosure by the contestants, and ultimately improve the *ex ante* chances of peaceful conflict resolution. In section 5 we provide evidence that mediators as well as disputants do recognize the value of commitment to quit in case the disputants' demands cannot be reconciled. Finally, we study mediators who have no independent budget for transfers or subsidies, and cannot impose peace to the contestants. To be sure, third-party states that mediate conflict, such as the United States, are neither unbiased nor powerless; However, single states account for less than a third of the mediators in mediated conflicts (Wilkenfeld, 2005), so that we view our assumption not only as a useful theoretical benchmark, but also as a reasonable approximation for numerous instances of mediated crises.[12]

[9] Interestingly, the intensity of conflict and asymmetric information are considered among the most important variables explaining when mediation is most successful (see e.g. Bercovitch and Houston, 2000, and Bercovitch et al., 1991). Our findings resonate with well-documented stylized facts in the empirical literature on negotiation (Bercovich and Jackson 2001, Wall and Lynn, 1993), that show that parties are less likely to reach an agreement without a mediator when the intensity of conflict is high than when it is low. Rauchhaus (2006) provides quantitative analysis showing that mediation is especially effective when it targets asymmetric information.

[10] Unlike us, Fey and Ramsay (2010) do not find any advantage of shuttle diplomacy. Private information in their model is about costs of war (private values), whereas in our model it is about the probability of winning (interdependent values). Further, while we only consider third party intervention revelation mechanisms, they also consider explicit bargaining protocols that shape the constraints of third party intervention.

[11] As some scholars claim, "mediator impartiality is crucial for disputants' confidence in the mediator, which, in turn, is a necessary condition for his gaining acceptability, which, in turn, is essential for mediation success to come about" (see e.g., Young, 1967, and the scholars mentioned in Kleiboer, 1996). On the other hand, when a mediator possesses independent information that needs to be credibly transmitted, some degree of bias may be optimal (see Kydd, 2003, Rauchhaus, 2006).

[12] In the terminology of Fisher (1995), our mediators perform "pure mediation," i.e., they gather information and propose settlements, rather than "power mediation," which also involves mediator's power to reward, punish or enforce.

To achieve their objectives, our mediators can facilitate communication, formulate proposals, and manipulate the information transmitted (see Touval and Zartman, 1985, for a discussion of these three roles; and Wall and Lynn, 1993, for an exhaustive discussion of all observed mediation techniques). Shuttle diplomacy requires private and separate caucuses, and has become popular since Henry Kissinger's efforts in the Middle East in the early 1970s and the Camp David negotiations mediated by Jimmy Carter, in which a third party conveys information back and forth between parties, providing suggestions for moving the conflict toward resolution. In the real world, mediators also often prevent conflict by facilitating communication or coordinating discussions among parties unwilling to communicate without a mediator. Such instances of mediation correspond to what we call communication facilitation.

The paper is organized as follows. Section 2 provides the benchmark characterization of the optimal arbitration mechanism, section 3 characterizes the optimal shuttle diplomacy mediation and shows that it achieves the same peace probability as the optimal arbitration benchmark; section 4 explains the role of the obfuscation strategy used in shuttle diplomacy. Before concluding, we describe in section 5 some evidence that mediators are able to commit to quit when disputants make demands that cannot be reconciled, as prescribed by our model. Section 6 concludes. All the proofs are in the appendix.

2 Arbitration

This section presents a simple model of conflict with asymmetric information, and lays out the arbitration program to minimize the chances that the conflict leads to war.

Two players contest a pie of size normalized to one.[13] War shrinks the value of the pie to $\theta < 1$. The expected payoffs in case of war depends on both players' private *types.* Each player

[13]Depending on the context, of course, the interpretation of the pie ranges from territory or exploitation of natural resources to any measure of social surplus in a country or partnership.

can be of type H or L with probability q and $(1-q)$, respectively.[14] This private characteristic can be thought of as being related to resolve, military strength, leaders' stubbornness, etc. We will often refer to type H as a "hawk" and to a L type as a "dove" (with no reference to the hawk-dove game). When the two players are of the same type, the expected share of the pie in case of war is 1/2 for both. When a type H player fights against an L type, its expected share of the pie is $p > 1/2$, and hence its expected payoff is $p\theta$. If $p\theta < 1/2$, the problem is trivial, as war can always be averted with the anonymous split $(1/2, 1/2)$; we shall assume henceforth that $p\theta > 1/2$.

The model has three parameters: θ, p, and q. Yet, it turns out that a more parsimonious description of all results can be given in terms of only two statistics:

$$\lambda \equiv \frac{q}{1-q} \quad and \quad \gamma \equiv \frac{p\theta - 1/2}{1/2 - \theta/2}.$$

The parameter λ is the hawk/dove odds ratio, and $\gamma \geq 0$ represents the ratio of benefits over cost of war for a hawk: the numerator is the gain for waging war against a dove instead of accepting the anonymous split $(1/2, 1/2)$, and the denominator is the loss for waging war against a hawk rather than accepting $(1/2, 1/2)$. Given that γ is increasing in θ, we will also interpret situations with low γ as situations of high intensity or cost of conflict. Note that when $\lambda \geq \gamma$, war can always be averted with the split $(1/2, 1/2)$ because the expected payoff of war for hawks, $(1-q)\,p\theta + q\theta/2$ is smaller than 1/2. We shall henceforth assume that $\lambda < \gamma$.

Among the institutions that one can consider to avert the risk of war, arbitration is the simplest institution to model and the best understood one. Under arbitration, a third party collects information privately from the disputing parties and makes binding decisions on how to resolve the dispute. Instead, a mediator can only make non-binding recommendations after collecting information privately. The implications of arbitration are well understood in economic theory, because in economic applications there typically exist external agencies

[14] To simplify the analysis, and keep the problem's dimensionality in check, we adopt a fully symmetric model. We believe that our results will hold approximately, for models that are close to symmetric.

such as the court to enforce the arbitrator's recommendations. In the case of international relations, instead, arbitration is not usually feasible, because disputants are sovereign entities, and enforcement of contracts or agreements is often impossible. However, it is useful to present arbitration in full here, as it provides a useful benchmark against which mediation can be compared.

Invoking the version of the revelation principle proved by Myerson (1979), and proceeding as in Bester and Wï¿½rneryd (2006), we can set up the arbitration game as a revelation mechanism, without loss of generality.

- After being informed of its type, each player i chooses whether to participate to the arbitration or not.

- If both players agree, each player i privately sends a report $m_i \in \{l, h\}$ to the arbitrator.

- Given reports $m = (m_1, m_2)$, the arbitrator prescribes a peaceful split $(x, 1 - x)$ with probability $p(m)$ and with probability $1 - p(m)$ the arbitration fails, so that the players escalate the conflict and fight a war.[15] Unlike the reports, the arbitrator's recommendation is public.

Again by the revelation principle, without loss of generality, we restrict attention to equilibria in which the players agree to participate in the arbitration and adopt truthful report strategies. Further, it can be shown that restricting attention to symmetric recommendations is without loss of generality, because the optimal arbitration program is linear.[16] Symmetry entails that the settlement is $(1/2, 1/2)$ if the players report the same type, that the split is

[15] Framed in the context of revelation mechanism, the model appears to assume that the arbitrator recommends war to disputants. This feature of the model should not be taken literally. In the real world, often mediators quit or freeze the dispute resolution process; this usually results in conflict escalation by the contestants. We discuss this feature of the model in detail in the section 5.

[16] We will see that each player's constraints are linear in the maximization arguments. Thus, the constraint set is convex. Hence, suppose that an asymmetric mechanism maximizes the probability of peace. Because the set-up is symmetric across players, the anti-symmetric mechanism, obtained by interchanging the players' identities, is also optimal. But then, the constraint set being convex, it contains also the symmetric mechanism obtained by mixing the above optimal mechanisms. As the objective function is linear in the maximization argument, such a symmetric mixed mechanism is also optimal.

$(b, 1-b)$ if the reports are (h, l) —and $(1-b, b)$ if they are (l, h), for some $b \in [1/2, 1]$. Let $p_L \equiv p(l, l)$, $p_M \equiv p(l, h) = p(h, l)$ and $p_H \equiv p(h, h)$. The optimal arbitration program determines b, p_L, p_M and p_H so as to minimize the war probability:

$$\min_{b, p_L, p_M, p_H} (1-q)^2 (1-p_L) + 2q(1-q)(1-p_M) + q^2 (1-p_H)$$

subject to the constraints that players are willing to reveal their type truthfully (the so-called *ex interim* incentive compatibility constraints) and to the constraint that the players are willing to participate to the arbitration regardless of their types (*ex interim* participation constraints). Specifically, the incentive compatibility constraint for either player of dove type is as follows:

$$(1-q)((1-p_L)\theta/2 + p_L/2) + q((1-p_M)(1-p)\theta + p_M(1-b)) \geq$$
$$(1-q)((1-p_M)\theta/2 + p_M b) + q((1-p_H)(1-p)\theta + p_H/2).$$

The left-hand side is the dove's payoff when truthfully revealing its type. With probability $1-q$, the opponent is also a dove, in which case the equal split $1/2$ is achieved with probability p_L and the payoff from war, $\theta/2$, is collected with probability $1-p_L$. With probability q, the opponent is hawk. With probability p_M, this leads to the split $1-b$, and with probability $1-p_M$ to the payoff from war $(1-p)\theta$. The right-hand side is the expected payoff from 'exaggerating strength.' When the opponent is a dove, the split b is recommended with probability p_M; whereas war takes place with probability $1-p_M$, leading to the payoff of $\theta/2$. When the opponent is a hawk, the split $1/2$ is recommended with probability p_H, and war occurs with probability $1-p_H$ yielding payoff of $(1-p)\theta$.

Similarly, for the hawk, the constraint

$$(1-q)((1-p_M)p\theta + p_M b) + q((1-p_H)\theta/2 + p_H/2) \geq$$
$$(1-q)((1-p_L)p\theta + p_L/2) + q((1-p_M)\theta/2 + p_M(1-b))$$

must hold, where the left-hand side is the payoff for truthfully revealing the type, and the right-hand side is the expected payoff from 'hiding strength.'

The *ex interim* participation constraints (for the dove and the hawk, respectively) are:

$$(1-q)(p_L/2+(1-p_L)\theta/2)+q(p_M(1-b)+(1-p_M)(1-p)\theta) \geq (1-q)\theta/2+q(1-p)\theta,$$
$$(1-q)(p_M b+(1-p_M)p\theta)+q(p_H/2+(1-p_H)\theta/2) \geq (1-q)p\theta+q\theta/2.$$

Again the left-hand sides are the dove and the hawk's payoffs when accepting the arbitration and truthfully revealing their types. Evidently, they coincide with the left-hand sides of the incentive compatibility constraints. The right hand sides are now the payoffs for refusing arbitration and triggering war. The player meets a dove with probability $1-q$, obtaining payoff $\theta/2$ if it is a dove and $p\theta$ if it is a hawk, whereas it meets a hawk with probability q, in which case its payoff is $(1-p)\theta$ if the player is a dove and $\theta/2$ if it is a hawk.[17]

We now report the main features of the arbitration solution.

Proposition 1 *The solution of the optimal arbitration program is such that, for $\lambda < \gamma$,*

1. *When delivered the arbitrator's decisions, the players' types become common knowledge;*

2. *The incentive compatibility constraint of the low type and the participation constraint of the high type always bind, the other constraints do not;*

3. *For $\lambda < \gamma/2$, so that the chance of a hawk opponent is low, dove dyads do not fight ($p_L = 1$), asymmetric dyads fight with positive probability ($0 < p_M < 1$), hawk dyads always fight ($p_H = 0$), and the arbitrators recommendations are self-enforcing ($b = p\theta$);*

4. *For $\gamma > \lambda > \gamma/2$, so that the chance of a hawk opponent is intermediate, dove dyads and asymmetric dyads do not fight ($p_L = 1$ and $p_M = 1$), hawk dyads fight with posi-*

[17] Although these constraints are not linear because of the products $p_M b$, they can be turned into linear constraints by changing the variable b with $p_B = p_M b$ and the constraint $1/2 \leq b \leq 1$ with $p_B \leq p_M \leq 2p_B$.

tive probability $p_H \in (0, 1)$ *, and the arbitrators recommendations are not self enforcing* $(1/2 < b < p\theta)$.

We now elaborate on the characterization described above.

First, the decisions delivered by the arbitrator reveal to each player the opponent's type. While the players do not know each other's types when choosing whether to participate to the arbitration or not, their types will be common knowledge after the arbitration is concluded. In fact, if the settlement is $(1/2, 1/2)$ each player knows that the opponent has the same type as its own. When the settlement is $(b, 1-b)$ or $(1-b, b)$, each player knows that the opponent has the opposite type, because b is different from $1/2$.

Second, we see that the arbitrator never quits (leading to a conflict escalation) when both players report low strength: $p_L = 1$; intuitively, there is no need to punish self-reported doves by quitting and triggering a conflict escalation, as they receive lower splits on average than if reporting to be hawks.

Third, the incentive compatibility constraint for the low type is always binding, because on average hawks receive higher peaceful splits than doves. Given that the incentive to exaggerate strength must be discouraged, there needs to be positive probability of war following a high report. The most potent channel through which the low type's incentive to exaggerate strength can be kept in check is by quitting and instigating a conflict escalation whenever there are two self-proclaimed high types (i.e., equivalently, the players' demands cannot be reconciled). When the odds ratio λ is low (few high types) it is indeed optimal to set $p_H = 0$ and $p_M > 0$, whereas for higher values of λ, $p_H < 1$ and $p_M = 1$. When λ is sufficiently high, the likelihood of a hawk is sufficiently high that instigating war against a dove is not needed to deter a dove to exaggerate strength. But when λ is low, deterring misreporting by a dove requires having self-reported hawks fight both against hawks and doves, with positive probability.

Fourth, when the odds ratio λ is low, the binding recommendation by the arbitrator is also self-enforcing, because $b = p\theta$. In fact, a high type is willing to accept the split b even *ex post*, when knowing that the opponent is a dove. (If the opponent had been a hawk, the arbitrator's settlement would have been 1/2). If it were to fight, its payoff against the dove would be exactly $p\theta$, so in fact the hawk is indifferent between fighting or not. But when the odds ratio is high, the binding recommendation by the arbitrator is not self-enforcing, because $b < p\theta$. Upon realizing that the opponent is a dove, the hawk would like to not accept the arbitrator's decision and go to war against the dove.

3 Mediation

In the previous section, we have characterized the optimal solution for the case of third party intervention by an arbitrator endowed with enforcement power. In this section, we consider mediation: The third party's prescriptions are not binding anymore.

The version of the revelation principle proved in Myerson (1982) guarantees that the following game form entails no loss of generality, when representing mediation:

- After being informed of its type, each player i privately sends a report $m_i \in \{l, h\}$ to the mediator.

- Given reports $m = (m_1, m_2)$, the mediator recommends a split $(b, 1-b)$ according to some cumulative distribution function $F(b|m)$, where the only recommendation leading to war in the support of $F(\cdot|m)$ is $b = 0$.[18] Unlike the reports, the mediator's recommendation is public.

- Each one of the contestants separately decide whether to accept the mediator's recommended split. Unless they both agree, war takes place.

Again by the revelation principle, we may restrict attention to distributions F such that the players reveal their type truthfully, and accept the mediator's recommendation (unless they are meant to lead to war) in equilibrium. This restriction imposes both *ex interim*

[18]Clearly all recommendations leading to war induce the same payoffs, and hence can be subsumed by the recommendation $b = 0$.

incentive compatibility constraints and *ex post* participation constraints, that are at least as demanding as the incentive compatibility constraints and participation constraints required by arbitration.

The reason for this is two fold. First, requiring each type of each player to participate in an arbitration before knowing which settlement will be proposed is less demanding than requiring them to agree to participate in each settlement that can be proposed on path by a mediator. In principle, they may be willing to agree to some and not to others, but still be better off in expectation by participating in the arbitration. Second, when entertaining the possibility of not telling the truth in the mediation game, each type of player knows that it can also deviate and trigger war after the mediator has proposed a peaceful settlement, whereas this is impossible under arbitration. Also note that, because the constraints associated with mediation are weakly more demanding than the arbitration's constraints, the welfare achieved under optimal mediation cannot be larger than the welfare achieved under optimal arbitration.

Turning to describing in details the constraints associated with mediation, to simplify notation, we restrict attention to mechanisms that are symmetric across players, where $F(\cdot|m_1, m_2) = 1 - F(\cdot|m_2, m_1)$ for all (m_1, m_2), and to discrete distributions F. We shall later see that this entails no loss of generality. Let $\Pr[m_{-i}, b, m_i]$ denote the equilibrium joint probability that the players send messages (m_i, m_{-i}) and that the mediator offers $(b, 1-b)$, and set $\Pr[b, m_i] \equiv \Pr[h, b, m_i] + \Pr[l, b, m_i]$. When player i is a hawk, it reports $m_i = h$ in equilibrium, and *ex post* individual rationality requires that

$$b \Pr[b, h] \geq \Pr[l, b, h] p\theta + \Pr[h, b, h]\theta/2, \text{ for all } b \in (0, 1), \tag{1}$$

which ensures that, if recommended the peaceful split b, i.e., for all $b \in (0, 1)$ such that $\Pr[b, h] > 0$, the hawk prefers accepting the split b to starting a war and receiving payoff $p\theta$ when the opponent is a dove —an event occurring with probability $\Pr[l, b, h]$— and $\theta/2$ when the opponent is a hawk —which occurs with probability $\Pr[h, b, h]$.

Similarly, when i is a dove, the *ex post* participation constraint dictates that

$$b\Pr[b,l] \geq \Pr[h,b,l](1-p)\theta + \Pr[l,b,l]\theta/2, \text{ for all } b \in (0,1). \tag{2}$$

Ex interim incentive compatibility requires that, when player i is a hawk, it truthfully reports $m_i = h$. The associated constraint dictates that

$$\begin{aligned} &qF(0|h,h)\theta/2 + (1-q)F(0|h,l)p\theta + \int_0^1 bdF(b|h) \geq qF(0|l,h)\theta/2 + \\ &(1-q)F(0|l,l)p\theta + \int_0^1 \max\{b, \Pr[l|b,l]p\theta + \Pr[h|b,l]\theta/2\}dF(b|l), \end{aligned} \tag{3}$$

where $\Pr[m_{-i}|b,m_i] = \Pr[m_{-i},b,m_i]/\Pr[b,m_i]$ whenever $\Pr[b,m_i] > 0$, and $F(\cdot|m_i) \equiv qF(\cdot|m_i,h)+$ $(1-q)F(\cdot|m_i,l)$, for m_i and m_{-i} taking values l and h. The left-hand side is the hawk's payoff if reporting truthfully to the mediator. The hawk obtains payoff $\theta/2$ when meeting a hawk (which occurs with probability q), if war takes place —an event of probability $F(0|h,h)$. It obtains the payoff $p\theta$ when meeting a dove and war takes place, and event of joint probability $(1-q)\,F\,(0|h,l)$. The expected value of a peaceful b depends on the distribution $F(b|h)$ and is thus the third term on the left hand side. The right hand side similarly displays the expected payoff when deviating and pretending to be a dove. The first two terms display the expected payoff in case of war. With joint probability $qF(0|l,h)$ the opponent is a hawk and war erupts, and the payoff is $\theta/2$; whereas with probability $(1-q)F(0|l,l)$ war erupts with a dove, leading to a payoff of $p\theta$. The third term represents the expected payoff when the mediator prescribes a peaceful split b. Note that when a player i contemplates deviating at the report stage, it also anticipates and takes into account that it might prefer to declare war *ex post*, even when the mediator recommends a peaceful solution. This fact is reflected by the maximum operator in the third term on the right hand side of the above inequality: the player may choose whether to accept a peaceful proposal b or go to war, again with expected payoff $\Pr[l|b,l]p\theta + \Pr[h|b,l]\theta/2$. If instead the mediator quits and players fight, the player is left with no choice but to fight, as it anticipates that this will be the equilibrium strategy of the opponent.

To ensure truth-telling by player i when a dove, the following constraint must be satisfied:

$$qF(0|lh)(1-p)\theta + (1-q)F(0|ll)\theta/2 + \int_0^1 (1-b)dF(b|l) \geq qF(0|hh)(1-p)\theta +$$
$$(1-q)F(0|lh)\theta/2 + \int_0^1 \max\{1-b, \Pr[l|b,h]\theta/2 + \Pr[h|b,h](1-p)\theta\}dF(b|h). \tag{4}$$

In the best equilibrium, the mediator seeks to minimize the probability of war, i.e.,

$$(1-q)^2 F(0|hh) + 2q(1-q)F(0|lh) + q^2 F(0|ll).$$

Because recommendations must be self-enforcing, there is *a priori* no reason to restrict the mediator in the number of splits to which he assigns positive probability. In fact, recommendations convey information about the most likely opponents' revealed types, and it might be in the mediator's best interest to scramble such information by means of multiple recommendations. Nevertheless, Proposition 2 below shows that relatively simple mechanisms reach the maximum probability of peace among all possible mechanisms, including asymmetric ones. These simple mechanisms can be described as follows. Given reports (h, h), the mediator recommends the peaceful split $(1/2, 1/2)$ with probability q_H, and war with probability $1 - q_H$. Given reports (h, l), the mediator recommends the peaceful split $(1/2, 1/2)$ with probability q_M, the split $(b, 1-b)$ with probability p_M, and war with probability $1 - p_M - q_M$, for some $b \geq 1/2$. Given reports (l, l), the mediator recommends the peaceful split $(1/2, 1/2)$ with probability q_L, the splits $(b, 1-b)$ and $(1-b, b)$ with probability p_L each, and war with probability $1 - 2p_L - q_L$.

Again, we relegate the explicit formulas of the solution to the Appendix, and restrict ourselves here to the description of its main features.

Proposition 2 *A solution to the mediator's problem is such that, for all $\lambda < \gamma$:*

1. *Both the low-type incentive compatibility constraint and the high-type participation constraint bind* ($b = p\theta$)*;*

2. *For* $\lambda > \gamma/2$, *doves do not fight* ($q_L + 2p_L = 1$); *hawk dyads fight with positive probability,* $q_H \in (0,1)$, *asymmetric dyads do not fight* ($p_M + 2q_M = 1$) *and the mediator recommends both the splits* $(b, 1-b)$ *and* $(1/2, 1/2)$ *with positive probability* ($q_M > 0$ *and* $p_M > 0$);

3. *For* $\lambda \leq \gamma/2$, *doves do not fight* —$q_L + 2p_L = 1$, *asymmetric dyads fight with positive probability* —$q_M = 0$, $p_M \in (0,1)$, *whereas hawk dyads always fight,* $q_H = 0$*;*

4. *For* $\gamma < 1$, *the mediator assigns the unequal splits* $(p\theta, 1-p\theta)$ *and* $(1-p\theta, p\theta)$ *to dove dyads with strictly positive probability* ($p_L > 0$).

We now comment on the solution, discussing informally the optimal mediator's techniques uncovered by the above Proposition.

As in the arbitration solution, doves never fight among each other. Again, this is because the mediator finds no need for quitting when the players reports being doves. Further, the low-type incentive compatibility constraint binds, because on average hawks receive higher peaceful splits than doves. Again, this leads the mediator to need to punish self-reported hawks by quitting and triggering war. Again, when the chance of facing a hawk is high, $\lambda > \gamma/2$, this is done by making hawk dyads fight, without the need to make asymmetric dyads fight.

However, there is a major difference with respect to optimal arbitration. When not quitting and triggering war, the mediator does not assign the unequal split $(b, 1-b)$ with probability one to (H, L) asymmetric dyads in which the first player is a hawk. With probability $q_M > 0$, she assigns the equal split $(1/2, 1/2)$ instead. This is equivalent to not reporting to the hawk that the opponent is a dove, because the split $(1/2, 1/2)$ is the one assigned to dyads in which both players are hawks.

This *obfuscation* strategy obtains to lower the reward for a dove from mimicking a hawk, hence it lowers the incentive to exaggerate strength. When achieving a peaceful split against a dove, a self reported hawk need not achieve the payoff $p\theta$ and may get $1/2$ instead. This obfuscation is possible, when $\lambda > \gamma/2$, because when two hawks meet and are assigned the equal split $(1/2, 1/2)$, they have a strict incentive not to fight. Hence, they will still have a incentive not to fight, even after the mediator mixes to obfuscate them.

The obfuscation is needed when $\lambda > \gamma/2$ because unlike the arbitrator, the mediator cannot force the hawk to accept an unequal split $b < p\theta$ when facing a dove. In fact, the *ex post* participation constraint of the hawk $b \geq p\theta$ cannot be violated. Hence, the mediator needs to *circumvent* this constraint by obfuscating the hawk meeting a dove. By doing so, she lowers the expected payoff of the hawk when meeting a dove below $p\theta$.

When the chance of facing a hawk is low, $\lambda \leq \gamma/2$, the mediator (like the arbitrator) ensures that doves do not self report that they are hawk by making both hawk dyads and asymmetric dyads fight with positive probability. Noting that, for this parameter range, it is the case that $b = p\theta$ for both the optimal mediation and the optimal arbitration solution, leads to the interesting result that optimal mediation and optimal arbitration coincide, as long as $\lambda \leq \gamma/2$ and $\gamma \geq 1$, i.e., when the benefit cost ratio for war is sufficiently large. Because $b = p\theta$ for both the optimal mediation and the optimal arbitration solution, it turns out that in this parameter region there is no room for obfuscation.

However, even when $\lambda \leq \gamma/2$ as well as when $\lambda > \gamma/2$, there is a noticeable difference between optimal mediation and the optimal arbitration when $\gamma < 1$. In this case, the mediator needs to obfuscate the players who report to be dove, by choosing one at random and not reporting that the opponent is a dove. The mediator does this by sending either one of the unequal reports $(b, 1-b)$ to dove dyads. This obfuscation strategy is used to keep in check the incentive for a hawk to pretend that it is a dove and hide strength. In fact, such a hawk can improve its payoff by further deviating from the mediator's prescription and waging war, in the case where the mediator's settlement reveals that the opponent is a dove. By obfuscating the self-reported dove and not revealing that it is facing a dove, the mediator keeps in check the benefit for this 'double-deviation' strategy.

Evidently, this double deviation is impossible with arbitration. There, players cannot deviate after being delivered the arbitrator's decision. Because the value of this double deviation is particularly high when $\gamma < 1$, i.e., the benefit of war is small relative to the cost, this case is when the mediator needs to adopt this obfuscation strategy.

Having explained in detail the main qualitative features of optimal mediation, we can now precisely answer the first set of questions presented in the introduction:

- What are the optimal mediation strategies?

    1. When the odds ratio of hawks is sufficiently large, the mediator does not always give the lion's share to a declared hawk facing a dove (or, equivalently does not always reveal to a self-reported hawk that it is facing a dove). This lowers the reward for a dove from mimicking a hawk, hence it lowers the incentive to exaggerate strength and achieve a favorable peace settlement with a dove.

    2. When the benefit/cost ratio of declaring war is small, the mediator offers with some probability unequal split to two parties reporting low type (or, equivalently the mediator does not always reveal to a dove that it is facing a dove). This lowers the incentive to doubly deviate: hide strength and fight if finding out that the opponent is a dove.

Turning to comparing the peace probability under optimal mediation and under optimal arbitration, surprisingly, we find that they are the same.[19] Specifically, for $\lambda \leq \gamma/2$, the mechanisms with and without enforcement coincide. When $\lambda > \gamma/2$, the simplest optimal mechanism with enforcement is such that $b < p\theta$, which is not self-enforcing. But the optimal mechanism without enforcement obfuscates the players' reports, and this obfuscation succeeds in fully circumventing the enforcement problem.

[19]This result facilitates the proof of Proposition 2. It is enough to establish that the simple mechanism characterized there, and described in closed form in the Appendix, satisfies the more stringent constraints of the mediator's program. Because this mechanism achieves the same welfare as the solution to the arbitration problem, it must be optimal, *a fortiori*, in the mediator's program.

Proposition 3 *An arbitrator who can enforce recommendations is exactly as effective in promoting peace as a mediator who can only propose self-enforcing agreements.*

The intuition is as follows. First, note that the dove's *ex interim* incentive compatibility constraint and hawk's *ex interim* participation constraint are the only ones binding in the solution of the arbitrator's program. Conversely, the only binding constraints in the mediator's program with self-enforcing recommendations are the dove's incentive compatibility constraint and the *ex post* hawk's participation constraints $b \geq p\theta$. Recall that, in our solution, the hawk is always indifferent between war and peace if recommended a peaceful settlement. Further, the dove's incentive compatibility constraint in the mediator's problem with self-enforcing recommendations is identical to the dove's incentive compatibility constraint in the arbitrator's program, because a dove never wages war after exaggerating strength in the solution of mediator's problem with self-enforcing recommendations.

Further, the hawk's *ex interim* participation constraint integrates the two binding hawk's *ex post* participation constraints in the arbitrator's problem. While requiring a constraint to hold in expectation is generally a weaker requirement than having the two constraints, it turns out that the induced welfare is the same. This is easiest to see when $\lambda \leq \gamma/2$, as in this case the only settlement ever granted to a hawk is b, when the opponent is dove. For any mechanism with this property, the *ex interim* participation and the *ex post* participation constraints trivially coincide. Let us now consider the case $\lambda > \gamma/2$. In this case, the optimal truthful arbitration mechanism prescribes a settlement $b < p\theta$ that does not satisfy the *ex post* participation constraint of a hawk meeting a dove, as well as prescribing a settlement with slack, equal to $1/2$, to same type dyads. The mediator cannot reproduce this mechanism. But she circumvents the problem with the obfuscation strategy whereby the hawk is made exactly indifferent between war and peace when recommended either the split $1/2$ or the split $b = p\theta$. Hence, she optimally rebalances the *ex post* participation constraints so as to achieve the same welfare as the arbitrator.

Specifically, we find that the mediator chooses p_M and q_M optimally so that the expected payoff of a hawk meeting a dove under mediation, $p_M \cdot p\theta + q_M \cdot 1/2$, is exactly equal to the split imposed by a mediator to a hawk meeting a dove, $b \in (1/2, p\theta)$. The fact that this can be done follows exactly from the fact that b lies between $1/2$ and $p\theta$ and that p_M and q_M can be chosen so as to sum up to one when $\lambda > \gamma/2$. Evidently, the hawk's payoff when meeting a hawk can be made to coincide under mediation and arbitration, as there is no issue of enforcement there. For the same reason, the dove interim payoff can be made the same under mediation and under arbitration. These arguments conclude that the mediator's obfuscation strategy achieves the same equilibrium interim payoff for both types. Now, we note that mediation is not more restrictive than arbitration at the interim stage: all the differences are induced by the additional *ex post* stage. Hence, a consequence of the interim payoff equality between mediation and arbitration is the fact that these two institutions achieve the same welfare.

We can now answer the second question that we posed in the introduction.

- *How do mediation and arbitration differ in terms of conflict resolution?*
  - In our game, there is no difference in terms of optimal *ex ante* probability of peace between the two institutions.
  - Unless the chance of meeting a hawk is intermediate, the two optimal mechanisms coincide. When they do not coincide, the mediator's optimal obfuscation strategy fully circumvents the mediator's lack of enforcement power.

4 The importance of obfuscation

To assess the importance of obfuscation and shuttle diplomacy, we now solve the mediation problem by making obfuscation impossible, and forcing the mediator to report all the information she gathered by the disputants. We call this form of mediation *communication facilitation,* because the only role of the mediator is to facilitate communication among the two disputants, without hiding any of the information that she is given by the disputants.[20]

For comparability with optimal mediation, we adopt the same game form that we considered in the previous section. So, again, after privately learning its type, each player i sends a message $m_i \in \{l, h\}$ to the facilitator, who reveals them to the players and makes a recommendation $(x, 1-x)$ on the basis of the messages $m = (m_1, m_2)$ and possibly randomizing. With probability $1 - p(m)$, the facilitator instigates war; with probability $p(m)$, she chooses a peaceful split $x(m)$. To make facilitated communication comparable with the full fledged model of mediation in the previous section, which includes the possibility of shuttle diplomacy, we restrict attention to pure-strategy equilibria of this game in which players report truthfully their type, i.e., to separating equilibria. Further we focus on equilibria with peaceful splits $x(m)$ and probabilities $p(m)$ that are symmetric across players. Such symmetry restriction entails that $x(h, h) = x(l, l) = 1/2$, and that we only need to find another split value, i.e., $b \equiv x(h, l) = 1 - x(l, h)$, given that the message space contains only two elements. We here let $p_L \equiv p(l, l)$, $p_H \equiv p(h, h)$ and $p_M \equiv p(h, l) = p(l, h)$.

Before proceeding with the analysis, we briefly comment on the characteristics of our model of communication facilitation. In fact, because this type of mediator reports all information gathered by disputants, we observe that this mechanism can also be understood as a form of unmediated communication. Specifically, the communication to the facilitator can also be understood as direct cheap talk among the players. Further, the possibly randomized recommendation by the facilitator can be understood just as a public correlation random device. Following Aumann and Hart (2003), such a public device can be replicated by an additional round of communication (using so-called jointly controlled lotteries). Hence our game can be reformulated as a two-round communication game without any third party involvement.

For the sake of comparability with the mediation program in the previous section, we do not consider the possibility of further rounds of cheap talk.[21] The restriction to binary

[20]See Annex 2 of the WTO Legal Texts as an example of communication facilitation procedures with full transparency.

[21]This might help, however. Aumann and Hart (2003) provide examples of games in which longer, indeed unbounded, communication protocols improve upon finite round communication.

messages is natural given the binary type space. When focusing on pure-strategy equilibria, these two restrictions are without loss of generality. But it is possible that more rounds or larger message sets might help in mixed-strategy equilibria (although numerical optimization shows that expanding the message space by allowing one more message, or allowing one more round of cheap talk does not help in our game).[22] On the other hand, it can be shown that the restriction to a single peaceful split $x(m)$, for every m, rather than the consideration of a lottery over peaceful splits, is without loss of generality.[23]

Evidently, by the version of the revelation principle of Myerson (1982), we know that *any* unmediated communication protocol cannot outperform the mediation mechanism we studied in the previous section. In general, communication facilitation will yield a lower welfare than mediation. We will nevertheless show that the communication facilitation protocol we study here can match mediation's welfare when the cost of war is low relative to the potential benefit and when the chances of meeting a hawk is low (specifically, this occurs when $\gamma < \lambda/2$ and $\gamma \geq 1$).

The optimal separating equilibrium is characterized by the following program. Minimize the war probability

$$\min_{b, p_L, p_M, p_H} (1-q)^2(1-p_L) + 2q(1-q)(1-p_M) + q^2(1-p_H)$$

subject to the following *ex post* participation constraints and *ex interim* incentive compatibility constraints. First, reporting truthfully must be optimal. For the dove, this incentive

[22]Indeed, there exist games in which restricting the cardinality of the message space could play a role when considering mixed strategies. For example, Bester and Strausz (2000) study a mechanism design problem with a principal and two agents. The principal cannot commit to implement the mechanism recommendations. Because of this, they show that the principal can improve the optimal mixed strategy equilibrium, by asking agents to send a message out of a space with a larger cardinality than their type space. See also Bester and Strausz (2001) and (2007) on mechanism design when the principal cannot commit to the mechanism.

[23]Note that we can replace without loss any lottery over peaceful recommendations with its certainty equivalent. In fact, at the agreement stage, the requirement that the players accept such a deterministic average split is less stringent than the requirement that they accept all splits in the support of the lottery. Further, lotteries over peaceful splits affect each player's equilibrium payoff at the communication stage only through their expectations. Finally, the payoff of a player who deviates at the communication stage is convex in the recommended split, as we shall later see. Hence, the deviation payoff is lower when replacing a lottery with its certainty equivalent, thus making the equilibrium requirement less stringent.

compatibility constraint states that

$$(1-q)\left((1-p_L)\theta/2+p_L/2\right)+q\left((1-p_M)(1-p)\theta+p_M(1-b)\right)\geq$$
$$(1-q)\left((1-p_M)\theta/2+p_M\max\{b,\theta/2\}\right)+q\left((1-p_H)(1-p)\theta+p_H\max\{1/2,(1-p)\theta\}\right).$$

The left-hand side is the dove's equilibrium payoff. With probability $1-q$, the opponent is also a dove, in which case the equal split 1/2 occurs with probability p_L and the payoff from war, $\theta/2$, is collected with probability $(1-p_L)$. With probability q, the opponent is hawk. With probability p_M, this leads to the split $1-b$, and with probability $1-p_M$ to the payoff from war $(1-p)\,\theta$. The right-hand side is the expected payoff from exaggerating strength. When the opponent is a dove, the split b is recommended with probability p_M. In principle, the player may deviate from the recommendation, and collect the war payoff $\theta/2$, hence the payoff is $\max\{b,\theta/2\}$.[24] Further, war takes place with probability $1-p_M$. When the opponent is a hawk, the split 1/2 is recommended with probability p_H, and war occurs with probability $1-p_H$. Similarly, for the hawk, the incentive compatibility constraint

$$(1-q)\left((1-p_M)p\theta+p_Mb\right)+q\left((1-p_H)\theta/2+p_H/2\right)\geq$$
$$(1-q)\left((1-p_L)p\theta+p_L\max\{1/2,p\theta\}\right)+q\left((1-p_M)\theta/2+p_M\max\{1-b,\theta/2\}\right),$$

must hold, where the left-hand side is the equilibrium payoff and the right-hand side is the expected payoff from hiding strength.

Second, players must find it optimal to accept all peaceful splits. Given that, in a separating equilibrium, messages reveal types, this requires that

$$b\geq p\theta,\quad 1-b\geq(1-p)\theta.$$

[24] Because of the maxima on the right-hand side of the incentive compatibility constraints, as anticipated in footnote 23, the payoff of a player who deviates at the communication stage is strictly convex in the recommended split b, instead of linear in b.

That is, a hawk facing a self-proclaimed dove must get a share b that makes war unprofitable against a dove. Similarly, the dove's share against a hawk cannot be so low that it is better to go to war. The constraint that a player would accept an equal split when the opponent's type is the same as its own, $1/2 > \theta/2$, is always satisfied.

Solving this program yields the following characterization. The complete equilibrium specification is tedious and relegated to the Appendix.

Proposition 4 *A solution of the facilitated communication program is such that, for $\lambda < \gamma$:*

1. *The dove's incentive compatibility constraint binds;*

2. *If $\gamma \geq 1$ and/or $\lambda \geq (1+\gamma)^{-1}$, then the hawk's incentive compatibility constraint does not bind and $b = p\theta$; and further:*

   - *if $\lambda < \gamma/2$, then hawk dyads always fight, $p_H = 0$, asymmetric dyads fight with positive probability, $p_M \in (0,1)$ and doves do not fight, $p_L = 1$; the mediation and facilitation mechanisms coincide;*

   - *if $\lambda \geq \gamma/2$ (which covers also the case $\lambda \geq (1+\gamma)^{-1}$), then hawk dyads fight with positive probability, $p_H \in (0,1)$, dove and asymmetric dyads do not fight, $p_M = p_L = 1$; further, shuttle diplomacy strictly improves welfare over facilitated communication;*

3. *If $\gamma < 1$ and $\lambda < (1+\gamma)^{-1}$, then the hawk's incentive compatibility constraint binds and $b > p\theta$; and further $p_H = 0$, $p_M \in (0,1)$ and $p_L = 1$ for $\lambda < \gamma/(1+\gamma)$, whereas $p_H \in (0,1)$ and $p_M = p_L = 1$ otherwise as above; further, shuttle diplomacy strictly improves welfare over facilitated communication.*

Elaborating on this characterization, we first find, again, that war is never optimal among doves ($p_L = 1$) and that, again, the truth-telling constraint for the low type is always binding. As in the case of mediation, the facilitator achieves this by setting set $p_H = 0$ and $p_M > 0$ when λ is low (few high types), whereas for higher values of λ, she sets $p_H < 1$ and $p_M = 1$.

Now, suppose that $\gamma \geq 1$, so that the cost of war is low. Then the *ex post* participation constraint always binds, and hence $b = p\theta$; and the *ex interim* high-type truth-telling constraint never binds. This is because, when the cost of war is low, the hawk hiding strength always prefers to wage war (both against hawks and doves). Whether shuttle diplomacy improves upon facilitated communication or not depends on whether $\lambda > \gamma/2$ or $\lambda \leq \gamma/2$. Recall that when $\gamma > 1$ and $\lambda > \gamma/2$, then the mediator sometimes recommends the equal split $(1/2, 1/2)$ when one player reports to be a hawk, and the other claims to be a dove. Evidently, this strategy is unfeasible under facilitated communication, and shuttle diplomacy improves the chances of peace. When $\lambda \leq \gamma/2$ instead, shuttle diplomacy does not improve upon facilitated communication because there is no obfuscation in optimal mediation.

Finally, suppose that $\gamma < 1$, so that the cost of war is high. For $\lambda \leq 1/(1+\gamma)$, in the best equilibrium of the facilitated communication game, the high-type truth-telling constraint binds, and $b > p\theta$, unlike in the case of mediation. To see why, suppose by contradiction that $b = p\theta$. For $\gamma < 1$, this would imply that $1 - b > \theta/2$. Consider a hawk pretending to be a dove. If it meets a dove, its payoff is the same as if revealing to be a hawk, because it can secure the payoff $p\theta$ by waging war. If it meets a hawk, it gets $1 - b$ with probability p_M and $\theta/2$ with probability $1 - p_M$. By revealing to be a hawk, it gets $1/2$ with probability p_H and $\theta/2$ with probability $1 - p_H$ against a hawk. But we know that p_M is larger than p_H, and because $1 - b > \theta/2$, the hawk would prefer to pretend to be a dove, violating the incentive compatibility constraint. To make sure that all constraints are satisfied, we must have $b > p\theta$, so as to reduce the payoff from pretending to be a dove, as the split b is proposed for self-reported hawks. As a result, when $\lambda \leq 1/(1+\gamma)$, the high-type incentive compatibility constraint binds, whereas the high type participation constraint $b \geq p\theta$ does not.[25]

In contrast with the case of $\gamma \geq 1$, shuttle diplomacy always yields a strict welfare im-

[25] To see why $b = p\theta$ when $\lambda \in [1/(1+\gamma), \gamma]$, even if $\gamma < 1$, note that p_H increases in λ, as in the case where $\gamma \geq 1$. Because the incentive to hide strength decreases as p_H increases relative to p_M, we can reduce b as λ increases. When λ reaches the threshold $1/(1+\gamma)$, the offer b required for the hawk truth-telling constraint to bind is exactly $p\theta$. Further increasing λ cannot induce a further decrease in b, because the *ex post* participation constraint $b \geq p\theta$ becomes binding. So in the region where $\lambda \in [1/(1+\gamma), \gamma]$, the incentive compatibility constraint does not bind and $b = p\theta$.

provement when $\gamma < 1$. When $\lambda > 1/(1+\gamma)$, so that $b = p\theta$ in the perfectly separating equilibrium of the facilitated communication game, it is also the case that $\lambda > \gamma/2$ (note that $1/(1+\gamma) > \gamma/2$), and hence the mediator improves over the facilitator for the same reasons as when $\gamma \geq 1$. When $\gamma < 1$ and $\lambda < 1/(1+\gamma)$, unlike with facilitated communication, the mediator makes sure that the high-type incentive compatibility constraint is satisfied with $b = p\theta$. In fact, recall that the mediator offers $(b, 1-b)$ with positive probability when both players report to be doves. A hawk who is hiding strength, and who is offered $1-b$ believes that the opponent is most likely a hawk, and does not wage war. This reduces the above described incentive to hide strength in order to wage war if revealed that the opponent is weak.

We can now answer the final set of questions asked in the introduction.

- When does obfuscation matter, and how?
  - When the conflict is expected to be very costly or intense and the proportion of hawks is intermediate, shuttle diplomacy strictly improves upon facilitated communication by obfuscating communication to hawks when they are paired with doves.
  - When the conflict is expected to be very costly or intense and the the proportion of hawks is either low, or high, the separating equilibrium of the facilitated communication game exactly reproduces mediation and arbitration. They all yield the same welfare.
  - When the intensity and/or cost of conflict is high, shuttle diplomacy strictly improves welfare compared to facilitated communication by obfuscating communication to self-reported doves when paired with doves.

5 Evidence on Commitment

One of the main assumptions of our model is the mediator's commitment to lead to a conflict escalation if this is prescribed by the optimal solution. In the real world, this translates into the mediator's commitment to quit and terminate the mediation, at least with some probability, when the disputants' demands cannot be reconciled. In this section we provide two types of evidence that mediators in the real world are capable of keeping this commitment, and even that disputants prefer having a mediator with a credible reputation of being able to keep this commitment. The first type of evidence is from a data set that has settlement attempts as units of observation, namely the Issue Correlates of War dataset (ICOW), by Hensel et al (2008). The second type of evidence will be from an illustrative case study.

In fact, mediators often make clear to the disputants under which circumstances they will quit. Such contingent plans of action often include deadlines. According to Avi Gil, one of the key architects of the Oslo peace process, "A deadline is a great but risky tool. Great because without a deadline it's difficult to end negotiations. [The parties] tend to play more and more, because they have time. Risky because if you do not meet the deadline, either the process breaks down, or deadlines lose their meaning" (Watkins, 1998). Among the many cases in which this technique was used, see for instance Curran and Sebenius (2003)'s account of how a deadline was employed by former Senator George Mitchell in the Northern Ireland negotiations. Committing to such deadlines might be somewhat easier for professional mediators whose reputation is at stake, but they have been also used both by unofficial and official individuals, including Pope John Paul II and former U.S. President Jimmy Carter.[26] Meanwhile, institutions like the United Nations increasingly set time limits to their involvement upfront (see, for instance, the U.N. General Assembly report, 2000).

[26]Bebchik (2002) describes how Clinton and Ross attempted to impress upon Arafat the urgency of accepting the proposal being offered for a final settlement, calling it a "damn good deal" that would not be within his grasp indefinitely.

5.1 Evidence from settlement attempts data

The ICOW data set considers each settlement attempt as a distinct observation.[27] The key source of variation is whether mediators resume the mediation process through a new settlement attempt following failure or whether they withdraw for good. The theory assumes the mediator's commitment to withdraw from mediation, expecting that this condition is essential to successful mediation. Mediation attempts in the presence of commitment to quit should be expected to be more successful, and mediators with such a reputation more sought after to serve as a mediator compared to mediators that cannot commit to quit. We probe the ICOW data for information that is consistent with these two hypotheses.

To ascertain whether these patterns exist, we construct rough indicators of mediation termination and differentiate cases according to the circumstances under which they terminated. The settlement attempt may end with 1) an agreement fully implemented; 2) an agreement reached but not ratified by at least one party; 3) an agreement reached and ratified but not complied with by at least one party; 4) no agreement reached and/or escalation of violent hostilities.[28] We say a mediator has "quit" if the mediator does not initiate a new mediation attempt involving the same two primary disputants within at least 10 years of the end of the mediation failure. We consider this a "responsive quit" if the settlement attempt ended under any conditions 2–4. A non trivial proportion of failed mediation attempts satisfy this criteria and a significant portion of the mediators in the sample will have quit at least once. Mediators who have established a reputation for actually following through on commitments to withdraw from the process in response to disputant transgressions should be considered more effective mediators and therefore should be more likely to be invited as a mediator in future settlement attempts.

[27]We use ICOW instead of other datasets because it's level of analysis is appropriate for our purposes. The International Crisis Bargaining dataset (ICB), Brecher and Wilkenfeld (1997 and 2000) for example, would not be useful because it collapses all settlement attempts for each crisis into one observation.

[28]We use two alternative definitions of agreement, leading to quit1 and quit2 variables, but they lead to very similar results.

Variable	frequency	Rel. Frequency (total)	Rel. Frequency (failed)
quit1	38	28.4% (134 total)	41.8% (91 total)
quit2	39	29.1% (134 total)	41.5% (94 total)

Table 1 describes how frequently mediators quit in the sample of 134 mediation attempts present in the ICOW dataset. The relative frequency of quitting in the panel varies slightly based on the definition of quitting used, which vary by how strict the criteria are for coding the result as a responsive quit, but overall the data suggest that approximately 30% of the mediation attempts resulted in the mediator quitting by our definition. Of failed mediation attempts, mediators quit responsively in approximately 42% of cases. It is clear that mediators do quit in a significant portion of cases in which quitting is called for.

Variable	frequency	Rel. Frequency (total)
Quit1	116	86.6% (134 total)
Quit2	116	86.6% (134 total)

Table 2 demonstrates that an overwhelming majority of mediators in the sample quit the process (by our definition) at least once in their career. 116 of the 134 mediation attempts in the data were mediated by a third party that had quit at least once previously in the panel. This suggests mediators that demonstrate they will withdraw in response to belligerent transgressions are those asked again to mediate crises. The imprecision of the measurement and the limited variation does not allow us to claim that this is airtight evidence that the mediator's capability to credibly threaten to quit when appropriate is a key factor in determining the mediator selection, but the evidence is at least consistent with the general perception that the ability to commit we assume in our model is an important characteristic of real world mediation.

5.2 Illustrative Case Study: Kofi Annan in Kenya and Syria

When the contested December 2007 Kenyan Presidential election resulted in widespread ethnic violence, threatening to escalate to full-scale civil war, the international community rushed to establish peace-making efforts. After weeks of multiple uncoordinated mediation attempts that failed to bring the disputants closer to a negotiated settlement, the African Union appointed former UN Secretary-General Kofi Annan to head the Panel of Eminent African Personalities, a coalition of political actors to bring about a solution to the escalating violence. The most important pillar of Annan's mediation strategy was the consolidation of the disparate mediation efforts and interested third parties behind a single process. He demanded, and received, assurances that the international community back the Panel's efforts and refrain from pedaling alternative or competing mediation attempts. Throughout the negotiations, Annan constantly devised signals to communicate the lack of alternative negotiating channels to the disputants in order to ensure their engagement with the Panel's efforts and proposals.

While working hard to overcome the challenges associated with bridging the bargaining gap separating Kibaki and Odinga, Annan also made a point early on to communicate to the disputants that the Panel would not remain involved indefinitely. As the pace of progress slowed, Annan reiterated that he 'would not be available forever and that an alternative had to be found' (Lindenmayer and Kaye, 2009). When the two sides began stalling for time, refusing to make meaningful concessions on the political issue, Annan announced a suspension of the talks. In an interview with the Centre for Humanitarian Dialogue, Griffiths (2008), Annan described the motivation for his decision:

"...Look, this is getting nowhere, so I've decided I'm suspending talks." ..."I had done my duty and believed the leaders should do theirs."

Annan clearly used the suspension of the mediation process, and the threat of its termination, to coerce the parties into giving up more than they hoped at the bargaining table. He even went so far to strengthen his signal by suggesting his successor, Cyril Ramaphosa of South Africa. Kibaki and the PNU vehemently rejected this proposed replacement, claiming

Ramaphosa to be partial to Odinga's claim. But nevertheless, the signal seems to have been perceived as credible, as the disputants increased the urgency with which they engaged with the Panel's peace-making process (Lindenmayer and Kaye 2009).

Annan's commitment to withdraw from mediation in the event the parties refused to abide by his negotiating plan was a calculated complement to his demand to consolidate the international community behind one mediation process. These two tactics served the strategy to remove the disputants' ability to shop around for alternative mediators and conflict resolution avenues, leaving them with the option of following Annan's lead or continuing a fight neither side wanted. This strategy fit the crisis well because it was clear that both parties wanted to end the conflict, but that the possibility of alternative means to attain their political goals tempted them away from conceding too much at the negotiating table.

Despite a mixed record of success (failure in Rwanda and Bosnia; success in Kenya), Annan was asked again to step into the mediator's role as Special Envoy for the U.N and the Arab League in Syria's escalating civil war. One of the prominent reasons cited for his appointment was his recent success in Kenya. But, unlike in the Kenya case, in Syria Annan faced a deeply divided international community. Though nominally unanimous in support of Annan's role and his six-point peace plan, the competing world powers almost immediately took actions to undermine the mediation process. The U.S. and Western allies began arming rebel forces and continued to call for Assad to step down while Russia and China continued to prop up Assad's regime. Without the international community united behind the process, the situation deteriorated. Annan quit in response to continued escalation of violence and both parties' violation of the six-point plan. As such, Annan maintained his credibility to commit to withdraw from his role in the event disputants become uncooperative.

6 Concluding Remarks

By applying mechanism design techniques to the study of international conflict resolution, this paper derives a number of lessons on mediation in international relations. Given the sovereignty of States, it is often not feasible in international relations to rely on arbitration, i.e., to give authority to a third party that would have enforcement power over the disputants. Nonetheless, our first result is that an unbiased and completely uninformed mediator without any enforcement power can achieve the exact same probability of peace that would be guaranteed by an optimal arbitration process with enforcement. The inability to enforce is effectively replaced by an obfuscation strategy, which is however feasible only if the mediator operates in a manner similar to what is called "shuttle diplomacy,"i.e., only when the mediator talks to the disputants privately, reserving the right not to reveal all the information (s)he receives from them.

Other forms of communication facilitation where the disputants communicate directly or in any case there is full transparency in the mediation process, cannot achieve the same outcome, because the obfuscation strategy is obviously ruled out by public communication. The full transparency rules used in WTO settlement procedures (see WTO Legal Texts) probably respond to a different objective function of the designer, but our paper shows that such transparency rules, if applied to conflict resolution procedures, would not lead to *ex ante* desirable outcomes, especially when the cost of conflict is very high and/or conflict is likely to be indecisive. We have also shown that the availability of obfuscation strategies is particularly relevant when the frequency of hawkish attitudes is intermediate, whereas when the likelihood of hawkish attitudes is either very high or very low, then public communication may suffice.

Beside having clarified the relative appropriateness of different mediation institutions for different international crisis contexts, a contribution of this paper has been to explicitly characterize *how* does shuttle diplomacy achieve the same expected outcomes of arbitration. When the *ex ante* chance of power asymmetry is high, the mediator lowers the reward from mimicking a hawk by not always giving the lion's share to a hawk facing a dove. When the expected intensity or cost of conflict are high, regardless of the expected degree of uncertainty, the mediator reduces the temptation to hide strength by a strong player. The mediator's strategy is to lower the reward from mimicking a dove by sometimes giving an unequal split to two parties reporting to be a low type.

The characterization of what shuttle diplomacy mediators should do, as well as the characterization of the types of international crisis where such procedures dominate communication facilitation with transparency, are all results that are obtained using the full commitment assumption, used almost everywhere in the mechanism design literature. In section 5 we have provided the available evidence that commitment to quit by mediators is indeed a significant factor.

We conclude by noting that, while we have required recommendations to be self-enforcing, they need not be renegotiation-proof, as they might be Pareto-dominated for the players. For instance, when there is common belief that both players are hawks, they would be better off settling for an equal split rather than going to war, although doing so is part of the solution. Yet renegotiation-proofness does not seem to be a first and foremost concern of real world mediators. It is not overly realistic to think that, after the mediator quits, contestants who struggled to find an agreement in the presence of the mediator, will autonomously sit down at the negotiation table again, in search for a Pareto improving agreement. Indeed, while the literature on the causes of conflict underlines that contestants may not be able to individually commit to peaceful conflict resolutions, it may well be the case that they can jointly or even individually commit to belligerent resolutions, when such commitments are *ex ante* valuable. Audience costs, for instance, are recognized to provide an important channel that makes war threats credible (see, for instance, Tomz, 2007).[29]

[29] A technical difficulty for the analysis of renegotiation-proofness in our environment lies in the fact that there is no widely accepted notion of renegotiation-proofness for games with incomplete information. The definitions that apply to our game usually yield non-existence, given the restriction imposed by *ex post* individual rationality and incentive compatibility. Forges (1990), for instance, defines an equilibrium to be renegotiation-proof if it is the case that, for every further (exogenous) proposal that players can simultaneously accept or reject after the mediator's recommendation, players would not unanimously prefer the exogenous proposal. Unfortunately, it can be shown that this requirement is impossible to satisfy in our problem (details available upon request).

References

[1] Aumann, R.J. (1974): "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics* 1, 67–96.

[2] Aumann, R.J. and S. Hart (2003): "Long Cheap Talk," *Econometrica*, **71(6)**, 1619–60.

[3] Banks, J.S. and R.L. Calvert (1992): "A Battle-of-the-Sexes Game with Incomplete Information," *Games and Economic Behavior*, **4**, 347–372.

[4] Bebchik, B. (2002): "The Philosophy and Methodology of Ambassador Dennis Ross as an International Mediator," *International Negotiation*, **7**, 115–131.

[5] Bercovitch, J. (1992): "The Structure and Diversity of Mediation in International Relations," in. J. Bercovitch and J.Z. Rubin (eds) *Mediation in International Relations: Multiple Approaches to Conflict Management*, New York: St. Martin's Press.

[6] Bercovitch, J. (1997): "Mediation in International Conflict: An Overview of Theory, a Review of Practice," in I.W. Zartmann and J.L. Rasmussen (eds) *Peacemaking in International Conflict: Methods and Techniques*, Washington D.C.: United States Institute of Peace Press.

[7] Bercovitch, J., J.T. Anagnoson, and D. Wille (1991): "Some conceptual issues and empirical trends in the study of successful mediation in international relations," *Journal of Peace Research* **28**, 7–17.

[8] Bercovitch, J. and A. Houston (2000): "Why Do They Do It? An Analysis of the Factors Influencing Mediation Behavior in International Conflicts," *Journal of Conflict Resolution*, **44(2)**, 170–202.

[9] Bercovich, J. and R. Jackson (2001): "Negotiation or Mediation? An Exploration of Factors Affecting the Choice of Conflict Management in International Conflict,"*Negotiation Journal*, **17**, 59–77.

[10] Bester, H. and R. Strausz (2000): "Imperfect Commitment and the Revelation Principle: the Multi-Agent Case" *Economics Letters,* **69,** 165-171.

[11] Bester, H. and R. Strausz (2001): "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case", *Econometrica* **69,** 1077-1098.

[12] Bester, H. and R. Strausz (2007): "Contracting with imperfect commitment and noisy communication," *Journal of Economic Theory,* **136(1):** 236-259.

[13] Bester, H. and K. Wärneryd (2006): "Conflict and the Social Contract," *Scandinavian Journal of Economics*, **108(2)**, 231–49.

[14] Blainey, G. (1988): *The Causes of War.* Third ed. New York: The Free Press.

[15] Brecher, M., and J. Wilkenfeld (2000): *A Study of Crisis* Ann Arbor, University of Michigan Press.

[16] Brito, D.L. and M.D. Intriligator (1985): " Conflict, War, and Redistribution, " *American Political Science Review*, **79(4)**, 943–957.

[17] Compte, O. and P. Jehiel (2009): "Veto Constraint in Mechanism Design: Inefficiency with Correlated Types," *American Economic Journal: Microeconomics,* **1**, 182–206.

[18] Cramton P.C. and T.R. Palfrey (1995): "Ratifiable mechanisms: learning from disagreement," Games Economic Behavior, **10**, 255–283.

[19] Curran, D. and J.K. Sebenius (2003): "The Mediator as Coalition Builder: George Mitchell in Northern Ireland," *International Negotiation*, **8(1)**, 111–147.

[20] Fearon, J.D. (1995): "Rationalist explanations for war," *International Organization,* **49:3**, 379–414.

[21] Fey, M. and C. Ramsay (2009): "Mechanism Design Goes to War: Peaceful Outcomes with Interdependent and Correlated Types," *Review of Economic Design*, **13(3)**, 233–250.

[22] Fey, M. and K. Ramsay (2010): "When Is Shuttle Diplomacy Worth the Commute? Information Sharing through Mediation," *World Politics*, 62-4, 529-60.

[23] Fisher, R.J. (1995): "Pacific, Impartial Third Party Intervention in International Conflict: a Review and Analysis." In *Beyond Confrontation: Learning Conflict Resolution in the Post Cold War Era.* Ed. Vasquez, Johnson, Jaffe and Stamato, Ann Arbor: University of Michigan Press, 39–59.

[24] Forges, F. (1990): "Some Thoughts on Efficiency and Information," in *Frontiers of Game Theory,* Binmore, Kirman and Tani (eds.), 133–149.

[25] Forges, F. (1999): "Ex post Individually Rational Trading Mechanisms," in *Current Trends in Economics: Theory and Applications*, ed. Charalambos Aliprantis, Nicholas Yannelis and Ahmet Alkan, Springer, 157–76.

[26] Green, J. and J.J. Laffont (1987): "Posterior Implementation in a Two-Person Decision Problem", *Econometrica* 55, pp. 69-94.

[27] Griffiths, M. (2008): "The Prisoner for Peace: An Interview with Kofi A. Annan," *Centre for Humanitarian Dialogue,* Available from

http://www.reliefweb.int/rw/RWFiles2008.nsf/FilesByRWDocUnidFilename/JBRN-7QMD4C

[28] Hensel, P.R., S. McLaughlin Mitchell, T.E. Sowers II, and C.L. Thyne (2008): "Bones of Contention: Comparing Territorial, Maritime, and River Issues." *Journal of Conflict Resolution* 117-143.

[29] Jackson, M.O. and M. Morelli (2011): "Reasons for War: an Updated Survey," *Handbook on the Political Economy of War* edited by Chris Coyne, Elgar Publishing.

[30] Kennan, J. and R. Wilson (1993): "Bargaining with Private Information", *Journal of Economic Literature,* **31(1),** 45–104

[31] Kleiboer, M. (1996): "Understanding Success and Failure of International Mediation," *Journal of Conflict Resolution*, **40(2)**, 360–89.

[32] Koremenos, B., Lipson C. and D. Snidal (2001): "The Rational Design of International Institutions," *International Organization*, **55(4)**, 761–99.

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

[34] Kydd, A. (2006): "When Can Mediators Build Trust," *American Political Science Review,* **100**, 449–462.

[35] Lindenmayer, E. and J.L. Kaye (2009): "A Choice for Peace? The Story of Forty-One Days of Mediation in Kenya," *International Peace Institute*, Available from http://www.ipacademy.org/publication/policy-papers/detail/221-a-choice-for-peace-the-story-of-forty-one-days-of-mediation-in-kenya.html

[36] Matthews S.A. and A. Postlewaite (1989): "Pre-play communication in two-person sealed-bid double auctions," *Journal Economic Theory* **48**, 238–263.

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

[38] Myerson, R. B. (1982): "Optimal coordination mechanisms in generalized principal-agent problems," *Journal of Mathematical Economics*, **10(1)**, 67–81.

[39] Powell, R. (2006): "War as a Commitment Problem," *International Organization*, **60**, 169–203.

[40] Princen, T. (1992): *Intermediaries in International Conflict*, Princeton, N.J.: Princeton University Press.

[41] Raiffa, H. (1982): *The Art and Science of Negotiation,* Harvard University Press.

[42] Rauchhaus, R. W. (2006): "Asymmetric Information, Mediation and Conflict Management," *World Politics*, **58**, 207–241.

[43] Schwarz, M. and K. Sonin (2008): "A Theory of Brinkmanship, Conflicts, and Commitments," *Journal of Law, Economics, and Organization.* **24(1)**: 161–183.

[44] Stein, K.W. (1999): *Heroic Diplomacy: Sadat, Kissinger, Carter, Begin and the Quest for Arab-Israeli Peace*, Routledge.

[45] Tomz, M. (2007): "Domestic Audience Costs in International Relations: An Experimental Approach," *International Organization*, **61**, 821–40.

[46] Touval, S., and I.W. Zartman (1985): "Introduction: Mediation in theory," In *International mediation in theory and practice*, edited by S. Touval and I.W. Zartman, 7–17. Boulder, CO: Westview.

[47] United Nations General Assembly Security Council (2000): Report A/55/305-S/2000/809.

[48] Wall, J.A. and A. Lynn (1993): "Mediation: A Current Review," *Journal of Conflict Resolution*, **37(1)**, 160–94.

[49] Waltz, K. (1959): *Man, the State and War: A Theoretical Analysis*. New York: Columbia University Press.

[50] Watkins, M. (1998): "Building Momentum in Negotiations: Time-Related Costs and Action-Forcing Events," *Negotiation Journal*, **14(3)**, 241–256.

[51] Wilkenfeld, J., Young, K.J., Quinn, D.M. and V. Asal (2005): *Mediating International Crises*, Oxon, U.K.: Routledge.

[52] WTO Legal Texts: World Trade Organization, "Annex 2: Understanding on the Rules and Procedures Governing the Settlement of Disputes", http://www.wto.org/english/docs

[53] Young, O.R. (1967): *The intermediaries: Third parties in international crises*, Princeton, N.J.: Princeton University Press.

Appendix A – Mediation

Proof of Proposition 1. The proof follows from this Lemma.

Lemma 1 *The solution of the mediator's program with enforcement power is such that: For $\lambda \leq \gamma/2$,*

$$p_M = \frac{1}{\gamma - 2\lambda + 1},\ p_H = 0,\ \text{and } V = \frac{(\gamma+1)}{(\gamma - 2\lambda + 1)(\lambda+1)^2};$$

For $\lambda \geq \gamma/2$,

$$p_M = 1,\ p_H = \frac{2\lambda - \gamma}{(\gamma - \lambda + 1)\lambda},\ \text{and } V = \frac{\gamma+1}{(\gamma - \lambda + 1)(\lambda+1)}.$$

Proof. We first solve the following relaxed program:

$$\min_{b, p_L, p_M, p_H} (1-q)^2 (1-p_L) + 2q(1-q)(1-p_M) + q^2 (1-p_H)$$

subject to high-type *ex interim* individual rationality:

$$(1-q)(p_M b + (1-p_M) p\theta) + q\left(p_H \frac{1}{2} + (1-p_H)\frac{\theta}{2}\right) \geq (1-q) p\theta + q\frac{\theta}{2},$$

to low-type *ex interim* incentive compatibility:

$$(1-q)\left((1-p_L)\frac{\theta}{2}+p_L\frac{1}{2}\right)+q\left((1-p_M)(1-p)\theta+p_M(1-b)\right)\geq$$
$$(1-q)\left((1-p_M)\frac{\theta}{2}+p_M b\right)+q\left((1-p_H)(1-p)\theta+p_H\frac{1}{2}\right),$$

and to

$$p_L\leq 1, p_M\leq 1 \text{ and } p_H\geq 0.$$

First, note that $p_L = 1$ in the solution because p_L appears in the constraints only in the right-hand side of the low-type *ex interim* incentive compatibility constraint, which is increasing in p_L. Second, note that the low-type *ex interim* incentive compatibility must be binding in the relaxed program's solution, or else one could increase p_H thus reducing the value of the objective function, without violating the high-type *ex interim* individual rationality constraint. Third, note that the high-type *ex interim* individual rationality constraint must be binding in the relaxed program's solution, or else one could decrease b and make the low-type *ex interim* incentive compatibility slack.

Solving for b and p_H as a function of p_M in the system defined by the low-type *ex interim* incentive compatibility and high-type *ex interim* individual rationality constraints, and plugging back the resulting expressions in the objective function, we obtain

$$C=-p_M\frac{\gamma+1}{(\lambda+1)(\gamma+1-\lambda)}+K,$$

where K is an inconsequential constant. Hence, the probability of conflict is minimized by setting $p_M = 1$ whenever possible. Substituting $p_M = 1$, in the system defined by the low-type *ex interim* incentive compatibility and high-type *ex interim* individual rationality constraints, we obtain $p_H = \frac{2\lambda-\gamma}{(\gamma-\lambda+1)\lambda}$ and

$$b=\frac{1}{2}\frac{q-\theta-4p^2\theta^2-2p\theta-3q\theta+4p\theta^2+2q\theta^2+4p^2q\theta^2+4pq\theta-6pq\theta^2+1}{q+\theta-2p\theta-2q\theta+2pq\theta}.$$

The quantity p_H is strictly positive for $\lambda\geq\gamma/2$ and always smaller than one. At the same time $1/2 < b < p\theta$ whenever $\lambda\geq\gamma/2$.

Solving for b and p_M as a function of p_H in the system defined by the low-type *ex interim* incentive compatibility and high-type *ex interim* individual rationality constraints, and plugging back the resulting expressions in the objective function, we obtain

$$C=\frac{(\gamma+1)\lambda}{(\gamma-2\lambda+1)(\lambda+1)}p_H+K,$$

where K is another inconsequential constant. The coefficient of p_H is positive for $\lambda\leq\gamma/2$, hence the probability of conflict is minimized by setting $p_H = 0$, which entails $p_M = \frac{1}{\gamma-2\lambda+1}$ and $b = p\theta$. The quantity p_M is strictly positive and smaller than one when $\lambda\leq\gamma/2$.

The proof of Lemma 1 and hence of Proposition 1 is concluded by showing that this solution does not violate the high-type *ex interim* incentive compatibility and low-type *ex interim* individual rationality constraints in the complete program.

Indeed, for $\lambda \geq \gamma/2$, we verify that the slacks of these constraints are, respectively

$$\frac{1}{2}(\gamma-\lambda+1)^{-1}(1-\theta)(\gamma-\lambda)(\gamma+1)>0,$$
$$\text{and } \frac{1}{2}(\gamma-\lambda+1)^{-1}(\gamma+1)(1-\theta)>0.$$

Similarly, for $\lambda \leq \gamma/2$, the slacks are

$$\frac{1}{2}(\gamma-2\lambda+1)^{-1}(\lambda+1)^{-1}(1-\theta)(\gamma-\lambda)(\gamma+1)>0,$$
$$\text{and } \frac{1}{2}(\gamma+1-2\lambda)^{-1}(\lambda+1)^{-1}(\gamma+1)(1-\theta)>0.$$

Proofs of Proposition 2 and 3. These results follow from this Lemma.

Lemma 2 *A solution to the mediator's problem is as follows.*

- *For* $\lambda \leq \gamma/2$*,*

$$q_L+2p_L=1,\ b=p\theta,\ q_H=q_M=0,\ p_M=\frac{1}{1+\gamma-2\lambda},$$

$$p_L \leq \frac{2\lambda}{(\gamma-2\lambda+1)(\gamma-1)} \text{ if } \gamma\geq 1, \text{ and } p_L \geq \frac{(1-\gamma)\lambda}{2\gamma^2}\frac{(\lambda-\gamma)(\gamma+2)}{(\lambda-\gamma-1)} \text{ if } \gamma<1;$$

with ex ante peace probability

$$V=\frac{\gamma+1}{(1+\gamma-2\lambda)(1+\lambda)^2}.$$

- *For* $\lambda \geq \gamma/2$*,*

$$q_L+2p_L=1,\ p_M+q_M=1,\ b=p\theta,\ q_H=\frac{2\lambda-\gamma}{\lambda(\gamma+1-\lambda)},\ q_M=\frac{2\lambda-\gamma}{\gamma(\gamma+1-\lambda)},\ q_L\geq\frac{\lambda(2\lambda-\gamma)}{\gamma^2(\gamma-\lambda+1)}$$

$$p_L \leq 2\frac{(\gamma-\lambda)(\gamma+2)\lambda}{(\gamma-\lambda+1)\gamma(\gamma-1)} \text{ if } \gamma\geq 1, \text{ and } p_L \geq \frac{(1-\gamma)\lambda}{2\gamma^2}\frac{(\lambda-\gamma)(\gamma+2)}{(\lambda-\gamma-1)} \text{ if } \gamma<1;$$

with ex ante peace probability

$$V=\frac{\gamma+1}{(\gamma-\lambda+1)(\lambda+1)}.$$

Proof. Consider the general mechanisms subject to the *ex post* IR and *ex interim* IC* constraints (1)-(4). It is straightforward to observe that the *ex post* IR constraints constraints are stronger than the following (high-type and low-type, respectively) *ex interim* IR constraints

$$\int_0^1 bdF(b|h) \geq \Pr[l,h]p\theta + \Pr[h,h]\theta/2,$$
$$\int_0^1 bdF(b|l) \geq \Pr[h,l](1-p)\theta + \Pr[l,l]\theta/2, \text{ for all } b \in [0,1]$$

and that the *ex interim* IC* constraint are stronger than the *ex interim* IC constraint obtained by substituting the maxima with their first argument (the interim payoff induced by accepting peace recommendations later in the game).

By the revelation principle of Myerson (1979), the optimal *ex ante* probability of peace within the class of mechanisms which satisfy these *ex interim* IC and IR constraints cannot be larger than the *ex ante* probability of peace identified in Lemma 1 in Appendix D. Because the *ex interim* IC and IR constraints are weaker than the *ex interim* IC* and *ex post* IR constraints, it follows that any mechanism subject to the constraints (1)–(4) cannot yield a higher *ex ante* probability of peace than the one identified in Lemma 1.

Hence, to prove the result, it is enough to show that the formulas for the choice variables $(b, p_L, q_L, p_M, q_M, q_H)$ satisfy the constraints (1)-(4) and achieve the same *ex ante* probability of peace as in Lemma 1. Specialized to the mechanisms described by $(b, p_L, q_L, p_M, q_M, q_H)$, the *ex post* IR constraints take the following form, for the high type:

$$bp_M \geq p_M p\theta, \ (qq_H + (1-q)q_M) \cdot 1/2 \geq qq_H\theta/2 + (1-q)q_M p\theta,$$

and for the low type:

$$p_L b \geq p_L\theta/2, \ (qp_M + (1-q)p_L)(1-b) \geq qp_M(1-p)\theta + (1-q)p_L\theta/2,$$
$$(qq_M + (1-q)q_L) \cdot 1/2 \geq qq_M(1-p)\theta + (1-q)q_L\theta/2,$$

whereas the high-type *ex interim* IC* constraint is

$$\begin{aligned}
&q(q_H/2 + (1-q_H)\theta/2) + (1-q)(p_M b + q_M/2 + (1-p_M-q_M)p\theta) \geq \\
&\max\{(qp_M + (1-q)p_L)(1-b), qp_M\theta/2 + (1-q)p_L p\theta\} + \max\{(1-q)p_L b, (1-q)p_L p\theta\} \\
&+ \max\{(qq_M + (1-q)q_L) \cdot 1/2, qq_M\theta/2 + (1-q)q_L p\theta\} \\
&+ q(1-p_M-q_M)\theta/2 + (1-q)(1-2p_L-q_L)p\theta,
\end{aligned}$$

and the low-type *ex interim* IC* constraint is

$$\begin{aligned}
&q(p_M(1-b) + q_M/2 + (1-p_M-q_M)(1-p)\theta) \\
&+ (1-q)(p_L b + p_L(1-b) + q_L/2 + (1-2p_L-q_L)\frac{\theta}{2}) \geq \\
&\max\{(1-q)p_M b, (1-q)p_M\frac{\theta}{2}\} + \max\{(qq_H + (1-q)q_M) \cdot 1/2, qq_H(1-p)\theta + (1-q)q_M\frac{\theta}{2}\} \\
&+ q(1-q_H)(1-p)\theta + q(1-p_M-q_M)\theta/2,
\end{aligned}$$

It is straightforward to verify that the values provided in Lemma 2 are such that the *ex ante* IC* constraint in which the low type does not wage war after misreporting is binding. Also, plugging in our two sets of values for the choice variables gives the same welfare as in Lemma 1. We are left with showing that all other constraints are satisfied. We distinguish the two cases.

Step 1. Suppose that $\lambda < \gamma/2$, so that $q_M = q_H = 0$. After simplification, the low-type IC* constraint becomes

$$q(p_M(1-p\theta) + (1-p_M)(1-p)\theta) + (1-q)\cdot 1/2 \geq$$
$$(1-q)p_M p\theta + q(1-p)\theta + q(1-p_M)\theta/2,$$

which is binding for $p_M = \frac{1}{1+\gamma-2\lambda}$. Consider the high-type IC* constraint

$$q\theta/2 + (1-q)(p_M b + (1-p_M)p\theta) \geq \max\{(qp_M + (1-q)p_L)(1-b), qp_M\theta/2 + (1-q)p_L p\theta\}$$
$$+ \max\{(1-q)p_L b, (1-q)p_L p\theta\} + \max\{(1-q)q_L \cdot 1/2, (1-q)q_L p\theta\} + q(1-p_M)\theta/2,$$

Note that

$$(qp_M + (1-q)p_L)(1-b) \leq qp_M\theta/2 + (1-q)p_L p\theta,$$

as long as either $\gamma > 1$ or $p_L \geq \frac{(1-\gamma)\lambda}{2\gamma}p_M = \frac{(1-\gamma)\lambda}{2\gamma^2}\frac{(\lambda-\gamma)(\gamma+2)}{(\lambda-\gamma-1)}$ for $\gamma < 1$, that

$$(1-q)\,p_L b = (1-q)\,p_L p\theta$$

and that

$$(1-q)q_L \cdot 1/2 \leq (1-q)q_L p\theta.$$

Then we substitute in the high-type IC* constraint (duly simplified):

$$q\theta/2 + (1-q)(p_M b + (1-p_M)p\theta) \geq q\theta/2 + (1-q)\,p\theta,$$

which is clearly satisfied because $b = p\theta$.

Similarly, we find that the two high-type *ex post* constraints

$$p_M b \geq p_M p\theta, \text{ and } (qq_H + (1-q)q_M)\cdot 1/2 \geq qq_H\theta/2 + (1-q)q_M p\theta$$

are satisfied — the second one because both sides equal zero.

We need to show that the low-type *ex post* constraints are satisfied. Indeed:

$$p_L p\theta > p_L\theta/2, \; (1-q)q_L \cdot 1/2 > (1-q)q_L\theta/2,$$

whereas

$$(qp_M + (1-q)p_L)(1-p\theta) \geq qp_M(1-p)\theta + (1-q)p_L\theta/2,$$

as long as $p_L(\gamma-1) = p_L\frac{(\theta+2p\theta-2)}{(1-\theta)} \leq 2\frac{q}{(1-q)}p_M = 2\lambda p_M$. So that if $\gamma \geq 1$, $p_L \leq \frac{2\lambda}{(\gamma-2\lambda+1)(\gamma-1)}$ and if $\gamma < 1$, $p_L \geq 0 \geq \frac{2\lambda}{(\gamma-2\lambda+1)(\gamma-1)}$.

Finally the probability constraints are satisfied. In fact, $0 \leq p_M \leq 1$ requires only that $1 \leq 1+\gamma-2\lambda$, i.e., that $\lambda \leq \gamma/2$.

Step 2. Suppose that $\lambda \geq \gamma/2$. Consider the low-type constraint, first. After simplifying maxima, as the low type always accepts the split if exaggerating strength, the low-type IC* constraint is satisfied as an equality when plugging in the expressions $p_M + q_M = 1$, $b = p\theta$, $q_H = \frac{2\lambda-\gamma}{\lambda(\gamma+1-\lambda)}$, $q_M = \frac{2\lambda-\gamma}{\gamma(\gamma+1-\lambda)}$.

Then we consider the high-type IC* constraint. We proceed in two steps. We first determine the off-path behavior of the high type and show that

$$(qp_M + (1-q)p_L) \cdot (1-b) \leq qp_M\theta/2 + (1-q)p_L p\theta$$

as long as either $\gamma > 1$ or $p_L \geq \frac{(1-\gamma)\lambda}{2\gamma} p_M^+ = \frac{(1-\gamma)\lambda}{2\gamma^2} \frac{(\lambda-\gamma)(\gamma+2)}{(\lambda-\gamma-1)}$ for $\gamma < 1$, that

$$(1-q)p_L b = (1-q)p_L p\theta$$

and that

$$(qq_M + (1-q)q_L)\, 1/2 \leq qq_M\theta/2 + (1-q)q_L p\theta$$

as long as $q_L \geq \frac{1-\theta}{2p\theta-1} \frac{q}{1-q} q_M$, i.e., $q_L \geq \frac{\lambda}{\gamma} q_M = \frac{\lambda(2\lambda-\gamma)}{\gamma^2(\gamma-\lambda+1)}$.

Then we verify that the consequentially simplified high-type IC* constraint is satisfied with equality, when substituting in the expressions $p_M + q_M = 1$, $b = p\theta$, $q_H = \frac{2\lambda-\gamma}{\lambda(\gamma+1-\lambda)}$, $q_M = \frac{2\lambda-\gamma}{\gamma(\gamma+1-\lambda)}$.

We then verify that the two high-type *ex post* constraints

$$p_M b \geq p_M p\theta, \text{ and } (qq_H + (1-q)q_M) \cdot 1/2 \geq qq_H\theta/2 + (1-q)q_M p\theta$$

are satisfied with equality when substituting in the expressions for $b = p\theta$, $q_H = \frac{2\lambda-\gamma}{\lambda(\gamma+1-\lambda)}$, $q_M = \frac{2\lambda-\gamma}{\gamma(\gamma+1-\lambda)}$.

We then show that the low-type *ex post* constraints are satisfied. In fact

$$p_L p\theta > p_L\theta/2, \text{ and } (qq_M + (1-q)q_L) \cdot 1/2 > qq_M(1-p)\theta + (1-q)q_L\theta/2,$$

whereas

$$(qp_M + (1-q)p_L)(1-p\theta) \geq qp_M(1-p)\theta + (1-q)p_L\theta/2,$$

as long as $p_L(\gamma-1) = p_L\frac{(\theta+2p\theta-2)}{(1-\theta)} \leq 2\frac{q}{(1-q)} p_M = 2\lambda p_M$. So that if $\gamma \geq 1$, $p_L \leq 2\frac{(\gamma-\lambda)(\gamma+2)\lambda}{(\gamma-\lambda+1)\gamma(\gamma-1)}$ and if $\gamma < 1$, $p_L \geq 0 \geq 2\frac{(\gamma-\lambda)(\gamma+2)\lambda}{(\gamma-\lambda+1)\gamma(\gamma-1)}$.

We then see that the probability constraints are satisfied. In fact, because $\gamma + 1 - \lambda > 0$, $2\lambda - \gamma - \lambda(\gamma + 1 - \lambda) = (\lambda+1)(\lambda-\gamma) < 0$, and $2\lambda - \gamma - \gamma(\gamma+1-\lambda) = (\gamma+2)(\lambda-\gamma)$, the conditions $0 \leq q_H \leq 1$ and $0 \leq q_M \leq 1$ require only that $2\lambda - \gamma \geq 0$.

Finally, we verify by substitution that the equilibrium interim utility of high and low types are the same in the arbitration and mediation programs. Importantly, this is so because we verify that

$$p_M \cdot p\theta + q_M \cdot 1/2 = \frac{1}{2}\frac{q - \theta - 4p^2\theta^2 - 2p\theta - 3q\theta + 4p\theta^2 + 2q\theta^2 + 4p^2q\theta^2 + 4pq\theta - 6pq\theta^2 + 1}{q + \theta - 2p\theta - 2q\theta + 2pq\theta},$$

where the right hand side is the expression for $b \in (1/2, p\theta)$ for the case of arbitration.

As a consequence of the equality between equilibrium interim utility of high and low types, the welfare is also the same across the two mechanisms, as can be directly verified.

Having proved that the claimed solution satisfies all constraints, the proof of Lemma 2, and hence Propositions 2 and 3 is now concluded.

Appendix B - Unmediated Communication

Proof of Proposition 4 The proof follows from the following Lemma.

Lemma 3 *The best separating equilibrium is characterized as follows.*

1. *Suppose that $\gamma \leq 1$.*

   (a) *When $\lambda < \gamma/(1+\gamma)$, both ex interim IC* constraints bind,*

$$b > p\theta,\ p_H = 0,\ p_M = \frac{1}{(1+\gamma)(1-\lambda)}, \quad \text{and } V = \frac{1+\gamma+\lambda(1-\gamma)}{(1+\gamma)(1-\lambda)(1+\lambda)^2}.$$

   (b) *When $\lambda \in [\gamma/(1+\gamma), \min\{1/(1+\gamma), \gamma\}]$, both IC* constraints bind,*

$$b > p\theta,\ p_M = 1, \quad p_H = 1 - \frac{\gamma}{(1+\gamma)\lambda},\ \text{and } V = 1 - \frac{\gamma\lambda}{(1+\gamma)(1+\lambda)^2}.$$

   (c) *When $\lambda \in [1/(1+\gamma), \gamma)$, only the IC_L^* constraint binds,*

$$b = p\theta,\ p_M = 1,\ p_H = \frac{2\lambda - \gamma}{\lambda(2+\gamma)}, \quad \text{and } V = \frac{2(1+\lambda)+\gamma}{2+\gamma+\lambda(2+\gamma)}.$$

2. *Suppose that $\gamma > 1$.*

   (a) *When $\lambda < \gamma/2$, only the IC_L^* constraint binds,*

$$b = p\theta,\ p_H = 0,\ p_M = \frac{1}{1+\gamma-2\lambda},\ \text{and } V = \frac{1+\gamma}{(1+\gamma-2\lambda)(1+\lambda)^2}.$$

   (b) *When $\lambda \in [\gamma/2, \gamma)$, only the IC_L^* constraint binds,*

$$b = p\theta,\ p_M = 1, p_H = \frac{2\lambda - \gamma}{\lambda(\gamma+2)},\ \text{and } V = 1 - \frac{\gamma\lambda}{(2+\gamma)(1+\lambda)}.$$

The proof of lemma 3 proceeds in two parts.
Part 1 ($\gamma \geq 1$).
We set up the following relaxed problem:

$$\min_{b, p_L, p_M, p_H} (1-q)^2(1-p_L) + 2q(1-q)(1-p_M) + q^2(1-p_H)$$

subject to the high-type *ex post* IR constraints:

$$b \geq p\theta$$

to the probability constraints:

$$p_L \leq 1, p_M \leq 1, 0 \leq p_H$$

and *ex ante* low-type IC* constraint:

$$(1-q)\left((1-p_L)\frac{\theta}{2} + p_L\frac{1}{2}\right) + q\left((1-p_M)(1-p)\theta + p_M(1-b)\right) \geq$$
$$(1-q)\left((1-p_M)\frac{\theta}{2} + p_M b\right) + q\left((1-p_H)(1-p)\theta + p_H\frac{1}{2}\right)$$

Step 1. We want to show that $p_L = 1$. We first note that setting $p_L = 1$ maximizes the LHS of the relaxed low-type IC* constraint and does not affect the RHS. It is immediate to see that the high-type *ex post* constraint is not affected either.

Step 2. We want to show that the relaxed low-type IC* constraint binds. Suppose it does not. It is possible to increase p_H thus decreasing the objective function without violating the constraint (note that there is no constraint that $p_H < 1$ in the relaxed problem).

Step 3. We want to show that the high-type *ex post* constraint binds. Suppose it does not. Then $b > p\theta$, and it is possible to reduce b without violating the *ex post* constraint. But this makes the low-type relaxed IC* constraint slack, because $-b$ appears in the LHS and b in the RHS. Because step 2 concluded that the low-type relaxed IC* constraint cannot be slack in the solution, we have proved that the *ex post* constraint cannot be slack.

Step 4. We want to show that for $\lambda \leq \gamma/2$: $p_H = 0, p_M = \frac{1}{1+\gamma-2\lambda}$ in the relaxed program. The low-type relaxed IC* constraint and *ex post* constraint define the function

$$p_M = \frac{(1-\lambda p_H(\gamma+2))}{(\gamma-2\lambda+1)}, \tag{5}$$

substituting this function into the objective function

$$W = 2(1-q)(1-p_M) + q(1-p_H)$$

duly simplified in light of step 1, we obtain the following expression:

$$W = p_H\frac{(2\lambda+\gamma+3)\lambda}{(\gamma-2\lambda+1)(\lambda+1)} + \frac{2\gamma-3\lambda+\lambda\gamma-2\lambda^2}{(\gamma-2\lambda+1)(\lambda+1)},$$

where we note that, because $\gamma \geq 2\lambda$, the coefficient of p_H is positive and the whole expression is positive. Hence, minimization of W requires minimization p_H. Setting $p_H = 0$ and solving for p_M in (5) yields

$$p_M = \frac{1}{1+\gamma-2\lambda}.$$

Because $\lambda \leq \gamma/2$, it follows that $p_M \leq 1$, as required. We note that the probability of war equals:

$$C = \frac{\left(2\gamma - 3\lambda + \lambda\gamma - 2\lambda^2\right)\lambda}{\left(\gamma - 2\lambda + 1\right)\left(\lambda + 1\right)^2}.$$

Step 5. We want to show that for $\lambda \geq \gamma/2$, $p_M = 1, p_H = \frac{2\lambda-\gamma}{\lambda(\gamma+2)}$ in the relaxed problem. In light of the previous step, the solution $p_H = 0$ yields $p_M > 1$ and is not admissible when $\lambda > \gamma/2$. Because p_M decreases in p_H in (5), the solution requires setting $p_M = 1$ and, from (5), $p_H = \frac{2\lambda-\gamma}{\lambda(\gamma+2)}$. When $\lambda \geq \gamma/2$, $p_H \geq 0$ and hence the solution is admissible. We note that the probability of war equals:

$$C = \frac{\gamma\lambda}{\left(\gamma + 2\right)\left(\lambda + 1\right)}.$$

Step 6. We want to show that the solution constructed satisfies all the program constraints. The low-type *ex post* constraint $1 - b \geq (1-p)\theta$ is trivially satisfied, when $b = p\theta$. Because $b > \theta/2$ and $1/2 > (1-p)\theta$, the low-type *ex ante* IC* constraint coincides with the low-type *ex ante* relaxed IC* constraint. The condition $1 - b = 1 - p\theta \leq \theta/2$ yields $2 - 2p\theta \leq \theta$, i.e., $1 - \theta \leq 2p\theta - 1$, i.e., $\gamma = \frac{2p\theta-1}{1-\theta} \geq 1$. Hence, for $\gamma \geq 1$, we conclude that $1 - b \leq \theta/2$. So, after simplification, the *ex ante* high-type IC* constraint becomes:

$$\begin{aligned}
&\left(1-q\right)p\theta + q\left((1-p_H)\frac{\theta}{2} + p_H\frac{1}{2}\right)\\
&= \left(1-q\right)\left((1-p_M)p\theta + p_M b\right) + q\left((1-p_H)\frac{\theta}{2} + p_H\frac{1}{2}\right) \geq\\
&\left(1-q\right)\left((1-p_L)p\theta + p_L p\theta\right) + q\left((1-p_M)\frac{\theta}{2} + p_M\frac{\theta}{2}\right)\\
&= \left(1-q\right)p\theta + q\theta/2,
\end{aligned}$$

which is satisfied (with slack when $\lambda \geq \gamma/2$). The probability constraints are obviously satisfied.

Part 2 ($\gamma < 1$). We allow for two cases:

Case 1. I will temporarily consider the following relaxed problem:

$$\min_{b,p_L,p_M,p_H} (1-q)^2(1-p_L) + 2q(1-q)(1-p_M) + q^2(1-p_H)$$

subject to the low-type and high-type relaxed IC* constraints:

$$\begin{aligned}
&(1-q)\left((1-p_L)\frac{\theta}{2} + p_L\frac{1}{2}\right) + q\left((1-p_M)(1-p)\theta + p_M(1-b)\right) \geq\\
&(1-q)\left((1-p_M)\frac{\theta}{2} + p_M b\right) + q\left((1-p_H)(1-p)\theta + p_H\frac{1}{2}\right)
\end{aligned}$$

$$(1-q)\left((1-p_M)p\theta + p_M b\right) + q\left((1-p_H)\frac{\theta}{2} + p_H\frac{1}{2}\right) \geq$$
$$(1-q)p\theta + q\left((1-p_M)\frac{\theta}{2} + p_M(1-b)\right).$$

which embed the assumption (to be verified *ex post*) that $1-b \geq \theta/2$, and to the probability constraints:

$$p_L \leq 1, p_M \leq 1, 0 \leq p_H$$

Step 1. As in the previous case, we conclude that $p_L = 1$.

Step 2. We want to show that the low-type relaxed IC* constraint binds. Indeed, if it does not, we can increase p_H without violating either relaxed IC* constraint (note that the LHS of the high-type relaxed IC* constraint increases in p_H).

Step 3. We want to show that the high-type relaxed IC* constraint binds. Suppose not. We can then reduce b because the LHS of the high-type relaxed IC* constraint increases in b and the RHS decreases in b. This makes the low-type relaxed IC* constraint slack, without changing p_M and p_H. But in light of step 2, this cannot minimize the objective function. Hence, the high-type relaxed IC* constraint must bind.

Step 4. We want to show that for $\lambda < \gamma/(1+\gamma)$, $p_H = 0$ and $p_M = \frac{1}{(1+\gamma)(1-\lambda)}$ solve the relaxed problem. The binding relaxed *ex ante* IC* constraints define the function: $[p_M, b]\,(p_H)$, after substituting λ for q and γ for p, we obtain:

$$b = \frac{2\lambda + \gamma - \theta\lambda - \theta\gamma - 2\lambda p_H + \theta\lambda p_H - 3\lambda\gamma p_H + 2\theta\lambda\gamma p_H - \lambda^2 p_H - \lambda\gamma^2 p_H - \lambda^2\gamma p_H + \theta\lambda\gamma^2 p_H + 1}{2\left(1 - \lambda p_H - \lambda\gamma p_H\right)(\lambda+1)}$$
$$p_M = \frac{(1 - \lambda p_H(1+\gamma))}{(\gamma+1)(1-\lambda)}. \tag{6}$$

Substituting p_M into the objective function

$$W = 2(1-q)(1-p_M) + q(1-p_H)$$

duly simplified in light of step 1, we obtain:

$$W = p_H\frac{\lambda}{1-\lambda} + \frac{2\gamma - \lambda - \lambda\gamma - \lambda^2 - \lambda^2\gamma}{(\gamma+1)(\lambda+1)(1-\lambda)},$$

because the coefficient of p_H is positive, this quantity is minimized by setting $p_H = 0$. Then, solving for p_M and b when $p_H = 0$ we obtain:

$$b = -\frac{1}{2\lambda+2}\left(-2\lambda - \gamma + \theta\lambda + \theta\gamma - 1\right)$$
$$p_M = \frac{1}{(\gamma+1)(1-\lambda)}$$

we know that $1 \geq \gamma \geq \lambda$, so $p_M \geq 0$, but the condition $p_M \leq 1$ yields $\frac{1}{(\gamma+1)(1-\lambda)} - 1 \leq 0$, i.e., $\lambda \leq \frac{\gamma}{\gamma+1}$, as stated. We note that the probability of war equals:

$$C = \frac{\left(\lambda - 2\gamma + \lambda\gamma + \lambda^2 + \lambda^2\gamma\right)\lambda}{(\gamma+1)(\lambda+1)(\lambda-1)}.$$

Step 5. We want to show that for $\lambda < \gamma/(1+\gamma)$, $p_H = 0$ and $p_M = \frac{1}{(1+\gamma)(1-\lambda)}$ solve the original problem. Again, the low-type *ex ante* IC* constraint coincides with the relaxed low-type *ex ante* IC* constraint. We need to show that the *ex post* constraint $b \geq p\theta$ is satisfied. In fact, simplification yields:

$$b - p\theta = \frac{1}{2}(\lambda+1)^{-1}(1-\gamma)(1-\theta)\lambda > 0.$$

Finally we show that the high-type IC* constraint coincides with the (binding) relaxed high-type IC* constraint, i.e., that $1 - b \geq \theta/2$. Note in fact, that this implies that the *ex post* constraint $1 - b \geq (1-p)\theta$ is satisfied, because $\theta/2 > (1-p)\theta$. Indeed, after simplification, we obtain:

$$1 - b - \theta/2 = \frac{1}{2}(\lambda+1)^{-1}(1-\gamma)(1-\theta)\lambda \geq 0.$$

Step 6. We want to show that for $\lambda \in [\gamma/(1+\gamma), \min\{1/(1+\gamma), \gamma\}]$, $p_M = 1, p_H = 1 - \frac{\gamma}{(1+\gamma)\lambda}$ solves the relaxed problem. When $\lambda > \gamma/(1+\gamma)$, setting $p_H = 0$ violates the constraint $p_M = 1$. Further, the expression (6) reveals that p_M decreases in p_H. Hence minimization of p_H, which induces minimization of W, requires setting $p_M = 1$. Solving for b and p_H, we obtain:

$$b = -\frac{(-\lambda - 3\gamma + 2\theta\gamma - \lambda\gamma - \gamma^2 + \theta\gamma^2 - 1)}{2\lambda + 2\gamma + 2\lambda\gamma + 2}$$

$$p_H = \frac{\lambda - \gamma + \lambda\gamma}{(\gamma+1)\lambda} = 1 - \frac{\gamma}{(1+\gamma)\lambda}.$$

The condition that $p_H \geq 0$ requires that $\lambda \geq \frac{\gamma}{\gamma+1}$ as stated.

Step 7. We want to show that for $\lambda \in [\gamma/(1+\gamma), \min\{1/(1+\gamma), \gamma\}]$, $p_M = 1, p_H = 1 - \frac{\gamma}{(1+\gamma)\lambda}$ solves the original problem. Again, the low-type *ex ante* IC* constraint coincides with the relaxed low-type *ex ante* IC* constraint. We need to show that the *ex post* constraint $b \geq p\theta$ is satisfied. In fact, simplification yields:

$$b - p\theta = \frac{1}{2}(\gamma+1)^{-1}(\lambda+1)^{-1}(\lambda + \lambda\gamma - 1)(\theta - 1)\gamma$$

and this quantity is positive if and only if $\lambda \leq \frac{1}{\gamma+1}$. Finally we show that the high-type *ex ante* IC* constraint coincides with the (binding) relaxed high-type *ex ante* IC* constraint, i.e., that $1 - b \geq \theta/2$. Note in fact, that this implies that the *ex post* constraint $1 - b \geq (1-p)\theta$ is satisfied, because $\theta/2 > (1-p)\,\theta$. Indeed, after simplification, we obtain:

$$1-b-\theta/2 = \frac{1}{2}\left(\gamma+1\right)^{-1}\left(\lambda+1\right)^{-1}\left(1-\theta\right)\left(\lambda-\gamma+\lambda\gamma-\gamma^2+1\right)$$

and $\lambda-\gamma+\lambda\gamma-\gamma^2+1 \geq 0$ if and only if $\lambda \geq \frac{1}{\gamma+1}\left(\gamma+\gamma^2-1\right)$ but because $\frac{1}{\gamma+1}\left(\gamma+\gamma^2-1\right) < \frac{\gamma}{\gamma+1}$, this condition is less stringent than $\lambda \geq \frac{\gamma}{\gamma+1}$.

Case 2. We want to show that for $\lambda \in [1/(1+\gamma),\gamma)$, $p_M = 1, p_H = \frac{2\lambda-\gamma}{\lambda(2+\gamma)}$ solve the original problem. Consider now the same relaxed problem that we considered in the proof for the case of $\gamma \geq 1$. We know from the analysis for the case $\gamma \geq 1$, that this relaxed problem is solved by $p_H = 0, p_M = \frac{1}{1+\gamma-2\lambda}, b = p\theta$ for $\lambda < \gamma/2$ and by $p_M = 1, p_H = \frac{2\lambda-\gamma}{\lambda(\gamma+2)}, b = p\theta$ for $\lambda \in [\gamma/2,\gamma)$. We now note that

$$\frac{1}{\gamma+1} - \gamma/2 = \frac{1}{2}\left(\gamma+1\right)^{-1}\left(1-\gamma\right)\left(\gamma+2\right)$$

and this quantity is positive when $\gamma \leq 1$. Hence the possibility that $\lambda < \gamma/2$ is ruled out: On the domain $1/(1+\gamma) \leq \lambda \leq \gamma \leq 1$, the solution to the relaxed problem is $p_M = 1, p_H = \frac{2\lambda-\gamma}{\lambda(\gamma+2)}$, with $b = p\theta$. We now need to show that this is also the solution of the original problem. Again, the low-type *ex ante* IC* constraint coincides with the relaxed low-type *ex ante* IC* constraint. Consider the *ex ante* high-type IC* constraint. The condition $1-b = 1-p\theta \geq \theta/2$ yields $\gamma = \frac{2p\theta-1}{1-\theta} \leq 1$. Hence, for $\gamma \leq 1$, we conclude that $1-b \geq \theta/2$, and hence that $1-b \geq (1-p)\,\theta$. So the *ex ante* high-type IC* constraint becomes:

$$(1-q)\left((1-p_M)p\theta + p_M p\theta\right) + q\left((1-p_H)\frac{\theta}{2} + p_H\frac{1}{2}\right) - (1-q)p\theta - q\left((1-p_M)\frac{\theta}{2} + p_M(1-p\theta)\right) \geq 0$$

and indeed, after simplification, the LHS equals:

$$\frac{1}{2}\left(\gamma+2\right)^{-1}\left(\lambda+1\right)^{-1}\left(\lambda+\lambda\gamma-1\right)\left(1-\theta\right)\gamma,$$

a positive quantity as long as $\lambda+\lambda\gamma-1$, i.e., $\lambda > \frac{1}{\gamma+1}$, which is exactly the condition under which we operate.

This concludes the proof of the Lemma 3, and hence of Proposition 4.