Conflict Prevention by Bayesian Persuasion

Raphaela Hennigs*

January 24, 2019

Abstract

Asymmetric information between conflicting parties can lead to war. During mediation, third parties often obtain private information about a conflict. This information can be used to resolve the conflict peacefully. But for this to be possible, the mediator needs to credibly communicate the information to the conflicting parties. If conflicting parties cannot believe the mediator, they do not change their behaviour. This paper solves the credibility requirement by focusing on the generation of private information during mediation. Mediation is modelled as Bayesian Persuasion. The mediator generates evidence about each conflicting party strategically and commits to sharing the obtained evidence with the respective opponent. Players can be convinced not to fight each other. The war probability is reduced and players benefit from mediation. The model stresses the relevance of the information generating process for mediation.

*Max Planck Institute for Tax Law and Public Finance, Munich, Germany. E-Mail: raphaela.hennigs@tax.mpg.de. I thank Kai A. Konrad, Marco Serena, Bharat Goel and Mariana Lopes da Fonseca for helpful comments.

1 Introduction

Understanding conflict resolution and prevention mechanisms remains a central concern of international relations. Mediation is one of the most widely used techniques in international crises management. Mediation occurred in 37 out of the 84 conflicts registered by the International Crisis Behaviour Project in the period from 1990 to 2015. For 27 out of these 37 cases, mediation was either an important or the most important factor for easing tensions.[1] This paper contributes to the theoretical understanding of effective mediation. The paper concentrates on the question how a mediator can use research and intelligence to prevent war. Specifically, a mediator convinces conflicting parties not to fight each other by providing them with evidence about their respective opponent.

To understand how a mediatior can do so, think, for example, of aerial reconnaissance missions. Third parties often use these missions to gather information about a country's military power and activity in a conflict and can share this information with the conflicting parties. For instance, the United States conducted observational flights to monitor the compliance with treaties between Israel and Egypt, and Israel and Syria in the Israel-Arab conflict during the 1970s and 1980s. The obtained information was reported to both parties in each case (Laipson, Kurtzer, and Moore, 1995). During the Falklands War, the United States provided Argentina with photographical coverage of the Falklands Islands and open sea areas, it had been asked for by Argentina based on an agreement with NASA (Freedman, 2007). How can the provision of such information change conflicting parties behaviour? If evidence suggests that a country's military power is strong, the opponent has less incentive to fight. The result of observational flights or satellite pictures might, therefore, change the opponent's behaviour. A mediator can use this effect and employ investigations strategically to convince parties not to fight each other.

This idea is modelled in a conflict game between two players with mediation as persuasive

[1] See Brecher and Wilkenfeld (1997) and Brecher, Michael, Wilkenfeld, Beardsley, James, and Quinn (2017).

communication in light of the Bayesian Persuasion framework developed by Kamenica and Gentzkow (2011). Two players divide a pie of unit size with the default division being an equal split. Players simultaneously decide whether to accept the default division or to fight their opponent. Each player has high or low military strength and does not know the other player's strength. Absent mediation, this uncertainty leads to war. A strong player wages war as the chance to win against a low strength player is sufficiently high. A mediator can reduce players' uncertainty and prevent war by providing them with information about each opponent's strength. A player refrains from fighting, when he learns that he would likely not win against his opponent. I derive how a mediator can optimise his communication to minimise the war probability. The optimal communication takes the form of signals in the Bayesian Persuasion style.

Bayesian Persuasion (Kamenica and Gentzkow, 2011) is part of the growing literature on information design.[2] Bayesian Persuasion studies a sender's problem, who aims at changing a receiver's action. The receiver takes an action in an uncertain environment. Payoffs depend on the taken action and an unknown state of the world. Sender and receiver share a common prior belief about the state of the world. The sender can design an arbitrarily informative signal about the state of the world. The receiver perfectly understands this signal and rationally updates his belief after observing a signal realisation. Given the posterior belief, he chooses an optimal action to maximise his expected utility. Kamenica and Gentzkow (2011) observe that the sender can induce any distribution of posterior beliefs subject to the condition that its expected value equals the prior belief (*"Bayes plausibility"*).

In the conflict game with mediation, the mediator is the sender, and each of both players is a receiver. In the course of mediation, the mediator generates information that is indicative of players' strength levels. Whether a country possesses certain military equipment or weapons is, for example, indicative of the country's military strength. The mediator obtains information of this type and passes it on to the players. Moreover, the mediator decides how

[2]See, for instance, Bergemann and Morris (2018) for an overview of this literature

information is obtained. Spending more effort on investigations allows to generate more precise information. When deciding how precise the information should be, the mediator faces a trade-off. On the one hand, the mediator wants each player to believe that its opponent is strong. When a player believes that the opponent is strong, he does not fight. This effect suggests to bias the information such that it is more often indicative of a strong player. On the other hand, the communication needs to be credible. If the information is not credible, players do not change their behaviour. It is not credible to always assure a player that the opponent is strong. This limits how frequently the mediator can send information indicating a strong player. The mediator commits to pass on the investigation results directly to the players without manipulation. This commitment is crucial to make the communication credible.

Mediation in this model relies on rational belief updating. When a player receives information indicating a strong opponent, his belief about the opponent's strength changes. Fighting becomes unattractive. On the downside, a strong player fights after receiving information indicating a low strength opponent. Albeit this negative effect, the mediator can use the communication mechanism to mitigate the ex-ante war probability. A mediator can use investigation and research efforts to prevent war. Mediation in this form is the more beneficial, the costlier is war and the more likely it is that players fight during absent mediation. When war is costlier, convincing players not to fight is easier. Players benefit from mediation as costly fighting is reduced.

In an extension to the model, I analyse how players' behaviour changes when they anticipate mediation. As mediation reduces the war probability, it makes large militarisation less risky. I analyse how this effect changes players' incentives to militarise by endogenising the probability with which players are of high strength. Players make costly investments to increase their militarisation level. In this setting, mediation has an adverse effect as it increases players' incentives to militarise. Each player benefits the more from mediation, the more he is militarised. Thus, players militarise more when they anticipate mediation than when they do not anticipate mediation. Taking this factor into account, the overall effect of mediation is not straightforward. Mediation can lead to an overall increase in the ex-ante war probability.

Using Bayesian Persuasion to model mediation stresses the relevance of the information generation during mediation. Previous models of mediation can be categorised into two groups concerning the question of how a mediator obtains relevant information about a conflict. The first group assumes that the mediator obtains the relevant information from the conflicting parties. Conflicting parties possess private information which they reveal to the mediator. The second group assumes that the mediator has prior and exogenous access to information. However, a mediator often does not have superior knowledge about the conflict when he takes up his mediation mandate. Instead, he employs fact-finding and investigation groups to obtain information. That is, the mediator does not have prior access to exogenous information, but has means to obtain private information within the course of the mediation. This idea is captured in this paper.

This paper contributes to the growing literature applying Bayesian Persuasion to games. I apply Bayesian Persuasion to a game, in which two receivers interact with each other and uncertainty stems from bilateral private information. In the second part of this paper, I assume that players anticipate the signals used by the sender and that they can influence which signals are sent. The second part of this paper also contributes to the recent literature that analyses how mediation affects players' pre-conflict militarisation.

2 Literature Review

Uncertainty and asymmetric information are often named rationalist explanations for war in the literature on conflict. When conflicting parties have asymmetric information about the cost of fighting (e.g., Fearon, 1995; Powell, 1996) or the distribution of power (e.g., Blainey, 1988; Wittman, 1979), they reach peaceful agreements less easily.[3] War is the consequence. When uncertainty causes fighting, additional information can help to prevent war. Some authors have taken up this idea to model mediation. The common theme of these models is that a third party provides conflicting parties with additional information, such that the conflict can be resolved peacefully.[4,5]

Kydd (2003, 2006) and Rauchhaus (2006) analyse mediation in bargaining models of war in which conflicting parties cannot reach efficient agreements due to uncertainty or mistrust. A privately informed mediator uses cheap-talk to communicate information to the conflicting parties. Inherent to the assumption of cheap-talk, a central question in these models is, under which conditions the mediator's communication is credible. The authors disagree about whether bias towards one party or impartiality facilitates or obstructs credibility. Crescenzi, Kadera, Mitchell, and Thyne (2011) search for an alternative source of credibility. They argue that the mediator can communicate honestly because of mechanisms generated by the global democratic community. Fey and Ramsay (2010) focus on the question of how a mediator obtains information in the first place. According to their argumentation, a successful mediator needs to have exogenous access to information and must not rely on information obtained from conflicting parties. The main distinguishing feature between my approach and models in the cheap-talk framework is the form of the mediator's communication. In my model, the mediator's communication is credible, because he commits ex-ante to a communication strategy.

Commitment is also crucial in the literature using a mechanism design approach to address conflict resolution. Banks (1990) and Fey and Ramsay (2011, 2009) analyse mechanisms

[3]See also Powell (2002) for a review of bargaining theory in the context of international conflict and Ramsay (2017) for a recent review of explanations linking uncertainty to war.

[4]In the often cited trichotomy which categorised mediator behaviour according to its degree of intervention, this type of mediation falls into the category of "communication-facilitation strategies." This category encompasses the least intrusive mediation styles and is the one most widely observed in international mediation (Bercovitch, Anagnoson, and Wille, 1991; Bercovitch and Houston, 2000).

[5]Third parties can also influence conflicts by incentivising peaceful agreements or deterring parties from fighting. See also Kydd (2010) for a literature overview of rationalist approaches to conflict resolution in general.

for peaceful conflict resolution in a situation with asymmetric information about winning probabilities and/or the cost of fighting. They focus on self-enforcing mechanisms. Parties need to agree ex-post on the suggested payoff scheme. According to Hörner, Morelli, and Squintani (2015), this additional constraint does not render mediation less successful, in terms of conflict avoidance, as compared to an enforceable mechanism design scheme. They study a model with uncertainty about winning probabilities. My analysis of mediation can be seen as complementary to the mechanism design view. The mediator does not alter the payoff scheme of the conflict game, but changes its informational environment.

The contributions mentioned so far look at mediation taking the characteristics of the conflict and the conflicting parties as given. Going a step back, the question emerges how conflicting parties react if they anticipate mediation. Meirowitz, Morelli, Ramsay, and Squintani (2018) study how different conflict resolution mechanisms determine militarisation. They identify a perverse effect of unmediated peace talks, as those increase countries' incentives to militarise compared to a situation without any conflict resolution stage. This effect is not present when an impartial mediator engages in conflict resolution. Under optimal mediation, countries' incentives to militarise decrease and the overall effect of mediation is positive. Canidio and Esteban (2018) find an adverse effect of mediation in a model in which players can invest in strength and take wasteful pre-negotiation investments. Optimal mediation is biased towards strong players to reduce wasteful pre-negotiation investments. If they can, players increase their strength to use this bias. Mediation shifts wasteful investments to a prior stage and the positive effect of mediation can dissipate. Although I also identify a possible negative effect of mediation, the reasoning is different as will be laid out at a later point.[6]

Regarding the mediation mechanism applied, this paper relates to literature on Bayesian Persuasion, in particular to the strand of it which analyses games with multiple receivers.

[6]Potentially adverse effects of mediation have been discussed in political science. In particular, moral hazard logic has been used to argue that international peacekeeping institutions might have an adverse effect on peace as they reduce the cost of initiating conflict. (See for example Rauchhaus (2009); Crawford (2008).)

A number of papers in this literature have studied voting games (Wang, 2015; Alonso and Câmara, 2016; Chan, Gupta, Li, and Wang, 2017). Further, Bayesian Persuasion has been applied to investigate environments, in which agents interact strategically.[7] For instance, Ely (2017) analyses the optimal information disclosure to agents in a model of a bank run. Zhang and Zhou (2016) study the use of a public signal in a contest with unilateral private information. Inostroza and Pavan (2018) study Bayesian Persuasion in a global game of regime change and apply it to stress testing. Up to my knowledge, Bayesian Persuasion has not been linked to the context of mediation. Doing so allows to stress the role of the information generating process in mediation.

3 The Model

3.1 Conflict Game

The conflict game is based on Hörner, Morelli, and Squintani (2015). Two players, $i \in \{1, 2\}$, dispute over a pie of unit size. The default division is an equal split and can be taken as the status quo, which prevails unless at least one player chooses to fight for the whole pie. Players are heterogeneous with respect to *strength* $t_i \in \{H, L\}$. Strength is private information and also referred to as a player's type. Player i is of high strength, $t_i = H$, with probability q_i. The probability q_i is common knowledge and referred to as player i's *militarisation level.*

Players take an action $a_i \in \{f, nf\}$ simultaneously. f denotes the choice to fight and nf the one to accept the default division. If both players accept the default division, it is implemented and the game ends. If at least one player chooses to fight, war occurs. War is costly and reduces the pie to θ, $\theta \in (1/2, 1)$.[8] The division of the reduced pie depends on players' types. If players are of equal type, the reduced pie is divided equally. If a strong player encounters a weak player, he receives the whole reduced pie, whereas the weak one

[7]Mathevet, Perego, and Taneva (2017) deal with the general problem of optimal information design in games in which players interact strategically.

[8]Note that for $\theta \leq 1/2$ players would always prefer the default division $(1/2, 1/2)$ and never fight.

goes away empty-handed.[9]

We can think of a country's militarisation level as the sum of all those factors, which influence its military constitution. A largely militarised country is well equipped, has large troops, extensive knowledge about military tactics, etc. The factors that determine militarisation are constant across conflicts. However, once a country gets involved in a particular conflict, its actual military strength is subject to a certain moment of randomness. How well military equipment serves its purpose depends, for example, on the region in which a war is fought. Hence, militarisation determines the ex-ante probability with which a country is of high military strength, whereas the actual strength is the realisation of a random variable.

A low type receives at most $\theta/2$ in war and never benefits from fighting. A high type, on the other hand, gains when defeating a low type, but incurs a loss when fighting with a high type. Whether a high type has an incentive to fight depends on the opponent's militarisation level, which determines the probability with which the opponent is a high or a low type. If the opponent's militarisation is sufficiently small, the gain from defeating a low type outweighs the loss incurred when fighting with a high type in expectation. Specifically, for a high type player i fighting is dominant if $q_j < \bar{q}(\theta)$, $j \neq i$, where the threshold $\bar{q}(\theta)$ is defined such that $\frac{\bar{q}(\theta)}{1-\bar{q}(\theta)} = \frac{\theta-1/2}{1/2-\theta/2}$. The right-hand side of this equation states the ratio of the gain and the loss associated with fighting with a low and a high type, respectively. The left-hand side states the probability ratio of encountering a high and a low type opponent. Throughout, I refer to militarisation levels $q_i \geq \bar{q}(\theta)$ as *deterrent militarisation*, as they might deter the opponent from attacking.

In this section, I derive equilibria for the baseline case without meditation. A pure strategy is defined as $\phi_i : \{H, L\} \rightarrow \{f, nf\}$. I restrict the analysis to pure strategies. I simplify the equilibrium analysis by assuming that upon indifference, a player chooses to accept the default division.[10] Under this assumption, perfect Bayesian Nash equilibria of the

[9] I simplify the model used in Hörner, Morelli, and Squintani (2015) by assuming that a high type receives θ with probability 1 when fighting against a low type.

[10] This assumption can be justified by introducing an $\epsilon > 0$ cost for taking the choice to fight, for example.

game are summarised by the following proposition:

Proposition 1. *A separating equilibrium in symmetric and pure strategies exists in the above described conflict game:*

$$\phi_i(t_i) = \begin{cases} f & if \quad t_i = H \\ nf & if \quad t_i = L \end{cases} \qquad for\ i = 1, 2. \tag{1}$$

Call (1) *the* ***Aggressive Equilibrium****. If $\exists i : q_i < \bar{q}(\theta)$, the Aggressive Equilibrium is unique in the class of pure strategy equilibria. If $\forall i : q_i \geq \bar{q}(\theta)$, a pooling equilibrium exists:*

$$\phi_i(t_i) = \begin{cases} nf & if \quad t_i = H \\ nf & if \quad t_i = L \end{cases} \qquad for\ i = 1, 2. \tag{2}$$

Call (2) *the* ***Deterrent Equilibrium****.*

Proof. Suppose players $i = 1, 2$ stick to the separating strategy as in (1). Incentive compatibility for the high type is satisfied given $\theta \geq 1/2$ and never binds for the low type. Suppose players $i = 1, 2$ pool on $a_i = nf$ as in (2). Incentive compatibility binds for the high type. For player i, sticking to the pooling strategy is incentive compatible provided $q_j \geq \bar{q}(\theta), j \neq i$ with $\bar{q}(\theta) = \frac{\theta - 1/2}{\theta/2}$. Hence, (2) exists for parameters satisfying $\forall i : q_i \geq \bar{q}(\theta)$. Finally, observe that an equilibrium in which both players pool on $a_i = f$ always exists. This equilibrium is ruled out by assuming that upon indifference, a player accepts the default division, as a low type never strictly prefers $a_i = f$. To prove that Proposition 1 constitutes an extensive description of the set of pure strategy equilibria, the existence of equilibria in asymmetric strategies has to be ruled out. Consider the strategy profile in which player 1 sticks to the separating strategy and player 2 sticks to the peaceful pooling strategy. This strategy is not incentive compatible for player 2, as he has an incentive to deviate to the separating strategy. Hence, an asymmetric equilibrium in pure strategies does not exist. □

The Aggressive Equilibrium always exists. The intuition is simple. Consider a high type who expects his opponent to stick to the separating strategy. His action is relevant only for the case that the opponent is a low type. The high type would forego the potential gain from defeating a low type by accepting the default division. Hence, if one player sticks to the separating strategy, the opponent does so as well. This logic applies independently of militarisation levels. Moreover, the Aggressive Equilibrium is unique if fighting conditional on being a high type is dominant for at least one player. The Deterrent Equilibrium, on the other hand, can be supported if this is not the case and militarisation is large. Here, a high type does not have any incentive to fight, given that he expects his opponent to accept the default division. This corresponds to a situation in which military capabilities among countries are of deterrent height, such that no country initiates a war.

For militarisation levels $\forall i : q_i \geq \bar{q}(\theta)$, two equilibria are incentive compatible. I use payoff dominance to refine the set of equilibria:

Equilibrium Refinement. *If multiple perfect Bayesian Nash equilibria exist, players stick to the equilibrium under which ex-ante expected payoffs are maximised.*

Whenever it exists, the Deterrent Equilibrium is the payoff-dominant equilibrium and players stick to it. In this case, peace is maintained.

In the following, I focus on parameters satisfying $\exists i : q_i < \bar{q}(\theta)$, for which the Separating Equilibrium is the unique equilibrium, as the focus of this paper is to study mediation. As a consequence, war occurs if at least one player is a high type. It is useful to state the ex-ante war probability in the baseline case without mediation for later comparisons:

$$\boldsymbol{W}^{NM}(q_1, q_2) = q_1 q_2 + q_1(1 - q_2) + (1 - q_1)q_2.$$

Asymmetric information takes a decisive role in this model. First, under perfect information, high type pairs would not fight. Further, asymmetric information complicates truthful communication between players about their types. A low type always has an incentive to imitate a high type. For example, no truthful equilibrium could be sustained in a cheap-talk communication stage taking place prior to the conflict game.[11]

3.2 Mediation

As uncertainty plays a decisive role in this conflict game, additional information can prevent war. This section exploits this idea by introducing a prior stage of mediation to the game. The mediator is an impartial third player who aims at reducing the ex-ante war probability. He does not recommend a conflict settlement different from the status quo, but convinces players to accept the status quo. To do so, the mediator provides each player with relevant information about his opponent. While the mediator does not have any prior private information about players, he has means to gather additional information independently. The mediator commits on sharing the obtained information with the conflicting parties. He passes on any evidence obtained and cannot misreport on the investigation results. The mediation mechanism relies on the Bayesian Persuasion framework developed in Kamenica and Gentzkow (2011).

The mediator obtains information that is indicative of a player's strength. For instance: Whether or not a country has recently taken large investments into military capabilities; Whether or not a country has taken measures in preparation for a war, by, for example, moving troops close to the boarder; Whether or not a country has a certain type of military equipment, for example, chemical weapons. The opponent would value knowledge about each of these facts. While none of these facts allows to reach a final conclusion about the military strength of a country, each of these facts is helpful for assessing it. As a powerful third country or an international institution, the mediator often has superior access to this type of information. He can send investigation groups to the countries or use intelligence techniques to access information, which players hide from each other. The information is the more extensive and the more precise, the more effort is spent on obtaining it. By deciding how much effort to spend on obtaining the information, the mediator chooses the information's

[11] A sequential game with a simple cheap-talk communication stage is analysed in the appendix.

quality.

Formally, the mediator shapes the informational environment of the conflict game by choosing a signal profile $\boldsymbol{\pi} = \{\boldsymbol{\pi^1}, \boldsymbol{\pi^2}\}$ in the Bayesian Persuasion style. Each of the two independent signals $\boldsymbol{\pi^1}$ and $\boldsymbol{\pi^2}$ is addressed to one of the two players. Signal $\boldsymbol{\pi^1}$ is addressed to player 1 and potentially informative about the strength of player 2. The signal consists of a binary realisation space $\{h, l\}$ with realisations denoted as m_1 and probability distributions $\pi^1(h \mid t_2)$ such that $\pi^1(h \mid t_2) + \pi^1(l \mid t_2) = 1$ for $t_2 = H, L$. $\pi^1(h|H)$ is, for example, the probability with which the signal $\boldsymbol{\pi^1}$ has a high signal realisation, h, given that player 2 is a high type. Signal $\boldsymbol{\pi^2}$ is defined symmetrically. Throughout and without loss, I restrict my attention to signals for which $\pi^i(h \mid H) \geq \pi^i(h \mid L)$ holds.

The timing is as follows: The mediator chooses the signal profile $\boldsymbol{\pi}$. Players learn the chosen signals.[12] Each player observes the signal realisation addressed to him. As each player understands the signal addressed to him perfectly, he uses the observed realisation to update his belief about the opponent's type. Players take actions simultaneously. The taken actions determine payoffs as described in the baseline case.

Players are Bayesian. Player i forms the posterior belief q_j^h (q_j^l) after observing a high (low) signal realisation. q_j^h (q_j^l) denotes the probability with which player j is a high type given that player i received a high (low) signal realisation:

$$
\begin{aligned}
q_j^h &= \text{Prob}(t_j = H \mid m_i = h) = \frac{\pi^i(h \mid H) q_j}{\pi^i(h \mid H) q_j + \pi^i(h \mid L)(1 - q_j)}, \\
q_j^l &= \text{Prob}(t_j = H \mid m_i = l) = \frac{\pi^i(l \mid H) q_j}{\pi^i(l \mid H) q_j + \pi^i(l \mid L)(1 - q_j)}, j \neq i.
\end{aligned}
$$

The nominator of the first formula, for example, is the joint probability of player i receiving a high signal realisation and player j being a high type. The denominator is the total probability with which player i receives a high signal realisation. The ratio gives the probability with which player j is a high type given that player i receives a high signal realisation.

[12]Note that in equilibrium, players can infer the optimally chosen signal profile, as they know the mediator's maximisation problem.

The perfectly informative signal is given by the probability distributions $\pi^i(h \mid H) = 1$ and $\pi^i(h \mid L) = 0$. Upon observing a realisation of the perfectly informative signal, a player perfectly learns his opponent's type. Contrary, an uninformative signal is characterized by $\pi^i(h \mid H) = \pi^i(h \mid L) \in [0, 1]$. A signal $\boldsymbol{\pi^i}$ is informative if a high signal realisation is more likely when player j is a high type than when he is a low type, that is if $\pi^i(h \mid H) > \pi^i(h \mid L)$ holds. The informativeness or precision of the signal $\boldsymbol{\pi^i}$ increases in the difference between both probabilities. The more precise a signal is, the more a player learns by observing a realisation of it. Hence, the difference between the posterior beliefs q_j^h and q_j^l increases with the precision of signal $\boldsymbol{\pi^i}$. Assuming that the mediator chooses an informative signal $\boldsymbol{\pi^i}$, player i's belief about player j's type strictly increases after observing a high realisation and strictly decreases after observing a low realisation.

Players can condition their actions on the observed signal realisation. The pure strategy of player i is now a function of his type and the observed signal realisation: $\phi_i : \{H, L\} \times \{h, l\} \rightarrow \{f, nf\}$. To understand, how different signal profiles can induce different equilibria in the conflict game in the case $\exists i : q_i < \bar{q}(\theta)$, consider a vector of perfectly informative and a vector of uninformative signals as benchmark cases. If the mediator chooses uninformative signals, $\forall i : \pi^i(h \mid H) = \pi^i(h \mid L) \in [0, 1]$, players do not receive any information by observing signal realisations and stick to the Aggressive Equilibrium in the conflict game. On the other hand, if the mediator chooses perfectly informative signals, $\pi^i(h \mid H) = 1$ and $\forall i : \pi^i(h \mid L) = 0$, players perfectly learn their opponents' type upon observing a signal realisation. Hence, the vector of perfectly informative signals changes the conflict game into a game under perfect information. Under perfect information, high types attack low type opponents, but do not attack high type opponents in the payoff-dominant equilibrium. The ex-ante war probability is reduced in the perfect information environment compared to the initial case, as war between high type pairs is avoided.

The question emerges, whether the mediator can use the signals to improve further upon the perfect information case. The following proposition characterises the optimal signal profile to do so and the induced equilibrium behaviour in the conflict game. As previously, the analysis is restricted to pure strategies.

Proposition 2. *The mediator chooses the optimal signal profile* $\boldsymbol{\pi} = \{\boldsymbol{\pi^1}, \boldsymbol{\pi^2}\}$, *which is given by* [13]

$$\pi^i(h \mid H) = 1 \qquad \textit{and} \qquad \pi^i(h \mid L) = \frac{q_j}{1-q_j}\frac{1/2-\theta/2}{\theta-1/2} \qquad \textit{if } q_j < \bar{q}(\theta) \tag{3}$$

and

$$\pi^i(h \mid H) = 1 \qquad \textit{and} \qquad \pi^i(h \mid L) = 1 \qquad \textit{if } q_j \geq \bar{q}(\theta) \tag{4}$$

for $i = 1, 2, j \neq i$. *Players stick to the* ***Mediated Equilibrium****:*

$$\phi_i(t_i, m_i) = \begin{cases} f & \textit{if } t_i = H \textit{ and } m_i = l \\ nf & \textit{otherwise} \end{cases} \qquad \textit{for } i = 1, 2, \tag{5}$$

given that the mediator chooses optimal signal profile.

Proof. The mediator's problem is to find the optimal signal profile $\boldsymbol{\pi}$ to minimise the ex-ante war probability. For each signal profile $\boldsymbol{\pi}$, a set of pure strategy equilibria emerges in the conflict game. Payoff-dominance, as the equilibrium refinement concept, selects a unique equilibrium for each signal profile. This equilibrium allows to pin down the ex-ante war probability associated with each signal profile.

The mediator's problem can be solved backwards. Consider the Mediated Equilibrium as a candidate for the induced equilibrium in the conflict game. I first derive under which conditions a signal profile makes the Mediated Equilibrium incentive compatible. Subsequently, I derive that signal profile, which minimises the ex-ante war probability subject to these conditions being satisfied.

[13] Remember that $\bar{q}(\theta)$ increases with θ and $\lim\limits_{\theta \to 1/2} \bar{q}(\theta) = 0$.

Suppose players $i = 1, 2$ stick to the Mediated Equilibrium (5). It is sufficient to concentrate on incentive compatibility for the high type. Incentive compatibility constraints for a high type player i after observing a low and a high realisation are given by

$$\mathbb{E}_i \Pi_i(f \mid H, l) = 1/2 + (1 - q_j^l)(\theta - 1/2) - q_j^l(1/2 - \theta/2) \geq \\ 1/2 - q_j^l \pi^j(l \mid H)(1/2 - \theta/2) = \mathbb{E}_i \Pi_i(nf \mid H, l) \tag{IC$_{\text{Hl}}$}$$

and

$$\mathbb{E}_i \Pi_i(f \mid H, h) = 1/2 + (1 - q_j^h)(\theta - 1/2) - q_j^h(1/2 - \theta/2) \leq \\ 1/2 - q_j^h \pi^j(l \mid H)(1/2 - \theta/2) = \mathbb{E}_i \Pi_i(nf \mid H, h), \tag{IC$_{\text{Hh}}$}$$

respectively. These constraints hold if the updated beliefs q_j^h and q_j^l induced by the signal $\boldsymbol{\pi^i}$ satisfy

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi^j(h \mid H)} \geq \frac{q_j^l}{1 - q_j^l} = \frac{q_j}{1 - q_j} \frac{\pi^i(l \mid H)}{\pi^i(l \mid L)} \tag{IC$_{\text{Hl}}$}$$

and

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi^j(h \mid H)} \leq \frac{q_j^h}{1 - q_j^h} = \frac{q_j}{1 - q_j} \frac{\pi^i(h \mid H)}{\pi^i(h \mid L)}. \tag{IC$_{\text{Hh}}$}$$

If these conditions are satisfied for players $i = 1, 2$, the Mediated Equilibrium is incentive compatible. The Mediated Equilibrium is then, moreover, the payoff-dominant equilibrium among the set of incentive compatible equilibria. [14] In the Mediated Equilibrium, war occurs if a high type player observes a low signal realisation.

To minimise the ex-ante war probability under the Mediated Equilibrium, the media-

[14] The second incentive compatible equilibrium under these beliefs is the Aggressive Equilibrium. Payoffs under the Aggressive Equilibrium are (strictly) lower than under the Mediated Equilibrium.

tor chooses a signal profile $\boldsymbol{\pi}$, which satisfies these incentive compatibility constraints, to maximise the occurrence of high signal realisations:

$$\underset{\left\{\{\pi^i(h|t_j)\}_{t_j=H,L,j\neq i}\right\}_{i=1,2}}{Max} \quad \left\{q_1q_2\pi^1(h \mid H)\pi^2(h \mid H) + q_1(1-q_2)\pi^1(h \mid L) + (1-q_1)q_2\pi^2(h \mid L)\right\}$$

$$\text{s.t.} \quad \frac{\theta - 1/2}{(1/2-\theta/2)\pi^j(h \mid H)} \geq \frac{q_j}{1-q_j}\frac{\pi^i(l \mid H)}{\pi^i(l \mid L)} \quad j \neq i, i = 1,2 \tag{IC$_{\text{Hl}}$}$$

$$\frac{\theta - 1/2}{(1/2-\theta/2)\pi^j(h \mid H)} \leq \frac{q_j}{1-q_j}\frac{\pi^i(h \mid H)}{\pi^i(h \mid L)} \quad j \neq i, i = 1,2 \tag{IC$_{\text{Hh}}$}$$

$$\{\pi^i(h \mid t_j)\}_{t_j=H,L} \in [0,1] \quad j \neq i, i = 1,2.$$

Consider player 1. For $q_2 \geq \bar{q}(\theta)$, (IC$_{\text{Hh}}$) is satisfied by an uninformative signal and it is optimal so set $\pi^1(h \mid H) = \pi^1(h \mid L) = 1$. For $q < \bar{q}_2$, (IC$_{\text{Hl}}$) is always satisfied and it is sufficient to take care of (IC$_{\text{Hh}}$). Increasing $\pi^1(h \mid H)$ increases the objective function, while it relaxes (IC$_{\text{Hh}}$) for both, player 1 and player 2. It is thus optimal to set $\pi^1(h \mid H) = 1$. This allows to conclude that for any values of q_1 and q_2, $\pi^1(h \mid H) = \pi^2(h \mid H) = 1$ is optimal. For $q < \bar{q}_2$, $\pi^1(h \mid L)$ is chosen optimally such that (IC$_{\text{Hh}}$) binds. Signal $\boldsymbol{\pi^2}$ is found equivalently. The signal profile described in Proposition 2 is thus optimal to induce the Mediated Equilibrium.

To conclude that this signal profile is optimal, it is necessary to show that no different signal profile exists, which induces an equilibrium, that reduces the ex-ante war probability further. As low types accept the default division in any equilibrium, it is sufficient to concentrate on high types' behaviour. Two components of the signal profile influence a high type player i's incentives to fight. First, the information sent to player i about player j: The higher player i's belief about his opponent's strength is, the less incentives he has to fight. Second, the probability with which a high type player j is fought by his opponent j in equilibrium: The higher this probability is, the more incentives the high type player i has to fight. Given the optimal signal profile described in Proposition 2, the high type player i is never attacked by his opponent j, as player j always receives a high signal realisation.

Hence, player i's incentives to fight cannot be reduced by changing the equilibrium behaviour of player j. Finally, observe that it is not possible to persuade player i to always accept the default division if $q_j < \bar{q}(\theta)$. Thus, it is not possible to induce a different equilibrium with a lower ex-ante war probability. $\square$

In the Mediated Equilibrium, a high type player conditions his action on the observed signal realisation. A high type player i fights after observing a low signal realisation, but accepts the default division after observing a high signal realisation. A high signal realisation suggests that the opponent j is a high type and that player i's chance to win is low. Player i does not wage war after observing a high signal realisation in the Mediated Equilibrium. War occurs in the Mediated Equilibrium if exactly one of both players is a high type and receives a low signal realisation.

The mediator uses the signals to persuade players to accept the default division. The optimal signal $\boldsymbol{\pi^i}$ takes one of two forms, depending on the militarisation level of player j. If player j's militarisation is deterrent, $q_j \geq \bar{q}(\theta)$, such that fighting is never dominant for player i, the optimal signal $\boldsymbol{\pi^i}$ is trivially given by (4). In this case, the signal does not need to be informative to convince player i to accept the default division in the Mediated Equilibrium. An uninformative signal is optimal, so that player i always observes a high realisation and never fights.

If player j's militarisation is not deterrent, $q_j < \bar{q}(\theta)$, no signal can induce player i to never fight. Any informative signal $\boldsymbol{\pi^i}$ will induce a posterior belief $\hat{q}_j^l < q_j$, upon which fighting is dominant for a high type player i. The optimal signal in this case, (3), solves a trade-off. The signal needs to be sufficiently informative, such that fighting is no longer dominant for a high type player i after observing a high signal realisation. Meanwhile, the optimal signal maximises the occurrences of high signal realisations, upon which a high type player i accepts the default division. The signal given by (3) solves this trade-off optimally. Player i always observes a high realisation when player j is a high type and sometimes does so when player j is a low type. The information contained in a high realisation is such that a high type player i is exactly indifferent between fighting and accepting the default division in the Mediated Equilibrium. When observing a low realisation, on the other hand, a high type player i perfectly learns that his opponent is a low type and fights.

Comparative statics show how the parameters impact the precision of the informative signal $\boldsymbol{\pi^i}$ given by (3). As player j's militarisation q_j increases from 0 to $\bar{q}(\theta)$, the probability with which player i observes a high signal realisation when player j is a low type, $\pi^i(h \mid L)$, increases from 0 to 1. The signal becomes less precise. The higher player i's prior belief about player j's type is, the less information is necessary to convince player i to accept the default division. Similarly, the probability $\pi^i(h \mid L)$, increases with the cost of war, or, put differently, decreases with the size of the reduced pie θ. The costlier war is and the less player i can win by fighting, the less precise the signal can be to convince player i to accept the default division.

3.3 Mediation Benefit

The signals' precision impacts the benefit of the mediation mechanism. To discuss this benefit, I concentrate on parameters satisfying $\forall i : q_i < \bar{q}(\theta)$, for which fighting is dominant for both high type players. This restriction is taken for ease of exposition only. Given the optimal signal profile, war occurs in equilibrium when exactly one player is a high type and learns that his opponent is a low type. The ex-ante war probability under mediation is given by

$$\boldsymbol{W}^M(q_1, q_2, \theta) = q_1(1 - q_2)\pi^1(l \mid L) + q_2(1 - q_1)\pi^2(l \mid L).$$

Mediation reduces the war probability: High type pairs never fight and with strictly positive probability high-low type pairs do not fight. The difference in ex-ante probabilities of war

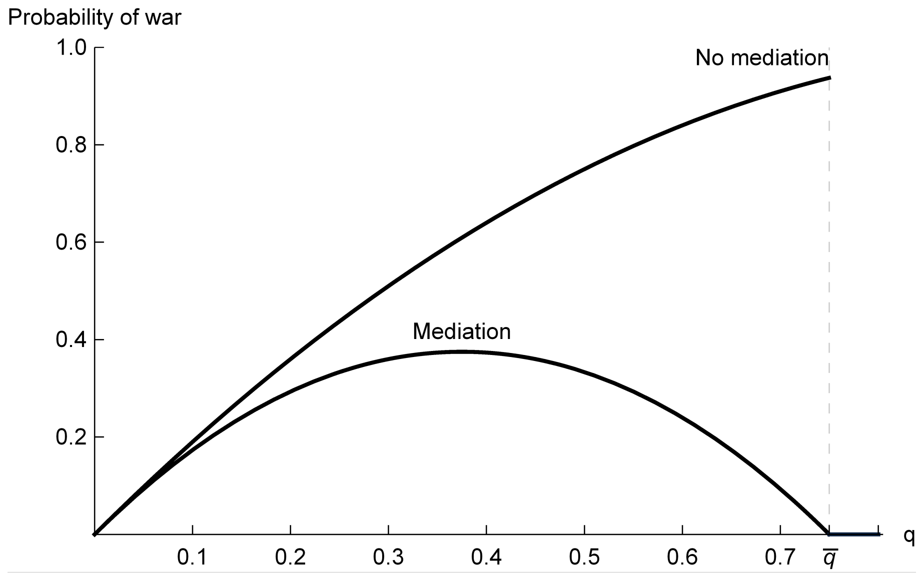


Figure 1: Ex-ante war probability as a function of the militarisation level q for $\theta = 0.8$. Here, players share a common militarisation level q. For $q \geq \bar{q}(\theta)$, the ex-ante war probability is zero.

with and without mediation is referred to as *mediation success* and given by

$$\boldsymbol{S}(q_1, q_2, \theta) \equiv \boldsymbol{W}^{NM}(q_1, q_2) - \boldsymbol{W}^{M}(q_1, q_2, \theta) = q_1 q_2 \frac{1/2}{\theta - 1/2}.$$

Militarisation levels q_1 and q_2 increase mediation success.[15] Two factors contribute to this effect. First, players fight more often during absent mediation as militarisation increases. The scope for mediation is therefore larger. Second, the more player $i's$ opponent is militarised, the less precise the information provided to player i needs to be to convince him to accept the default division. High realisations are more frequent, and high type players fight less often. Figure 1 illustrates both effects. The figure depicts the ex-ante war probability with and without mediation when war reduces the pie to $\theta = 0.8$ and assuming a symmetric militarisation q. For low values of q, the probability that at least one player is a high type is low as is the one of war. As q increases, so does the ex-ante war probability. Without mediation, the probability increases continuously up to the point where militarisation reaches the deterrent threshold $\bar{q}(\theta)$. For values above $\bar{q}(\theta)$, the Deterrent Equilibrium is incentive

[15]The derived expression for mediation success and its comparative statics are conditional on $\forall i : q_i < \bar{q}(\theta)$. Remember that $\bar{q}(\theta)$ increases with θ and $\lim\limits_{\theta \to 1/2} \bar{q}(\theta) = 0$.

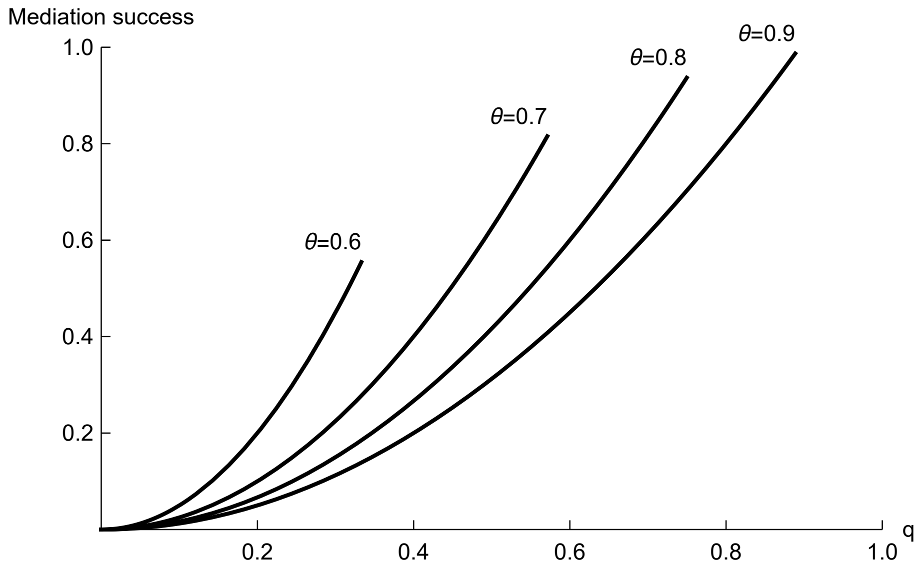


Figure 2: Mediation success as a function of the militarisation level for different values of θ. Here, players share a common militarisation level q.

compatible and played as the payoff-dominant equilibrium. The ex-ante war probability drops to zero. When q is close to $\bar{q}(\theta)$, players engage frequently in war, although each player has little incentive to do so. With mediation, the ex-ante war probability reaches its maximum for a value of q strictly smaller than $\bar{q}(\theta)$. For q close to $\bar{q}(\theta)$, a signal only slightly precise is sufficient to convince players to accept the default division and high realisations are frequent. Mediation can almost eliminate war in this case.

Similarly, costlier war increases mediation success, as it allows for less precise signals. Figure 2 shows this effect. Mediation success increases faster with militarisation q when war is costlier and θ lower. Figure 2 also shows that the range of q for which mediation is relevant shrinks with the cost of war. Taken together, the results suggest that mediation is more beneficial between two countries that are likely to engage in war and when conflict is costlier.

The whole mediation benefit is taken up by low types. If high types benefited, the war probability could be reduced further. Mediation increases a low type player i's expected payoff by

$$q_j \underbrace{\frac{q_i}{1-q_i}\frac{1/2-\theta/2}{\theta-1/2}}_{\pi^j(h|L)} 1/2.$$

This gain increases with militarisation of both, player i and player j. As the opponent's militarisation, q_j, increases, a low type player i is attacked more frequently absent mediation. Player i thus benefits more from mediation. Furthermore, this benefit is larger when player i is more militarised. Large militarisation of player i allows the mediator to send less precise information to player j. Occurrences of a high realisation, upon which j does not fight, become more frequent. For a similar reason, a low type's ex-ante benefit increases in the cost of war.

4 Militarisation as an Investment Choice

Mediation reduces the war probability. The benefit is taken up by low type players and is the larger, the more players are militarised and the costlier war is. How would players prepare for a conflict if they anticipated it to be mediated? Players might change their pre-conflict behaviour by expanding or reducing measures in preparation for a conflict. This change of pre-conflict behaviour would, in turn, influence the likelihood with which a conflict arises. This section addresses this question by endogenising the militarisation parameters q_1 and q_2. Players costly militarise to increase the probability with which they are of high type.[16] Militarisation costs are given by a convex function,

$$C(q_i) = \frac{1}{c}\frac{1}{2}q_i^2 \text{ with } c > 0,$$

[16]Meirowitz, Morelli, Ramsay, and Squintani (2018) offer an alternative interpretation of the choice variable q_i. They use this probability to describe the mixed strategy taken by player i. According to this interpretation, player i militarises at a constant cost to become a strong player with probability q_i and does not militarise and remains a weak player with complementary probability.

and are governed by the cost parameter c. A high level of c implies low militarisation costs. Players militarise simultaneously in stage 1. Chosen militarisation levels q_1 and q_2 are publicly observed. I consider two cases. In the first case, players militarise and encounter each other in the conflict game absent mediation. In the second case, players militarise and encounter each other in the conflict game with mediation.

4.1 Militarisation without Mediation

The sequential game is solved by backward induction. The analysis of the second stage translates directly from the baseline case. Two different subgame perfect equilibria emerge, depending on costs of militarisation and of war. A proof of the Proposition is delegated to the appendix.

Proposition 3. *In stage 2, the Deterrent Equilibrium* (2) *is played if* $\forall i : q_i \geq \bar{q}(\theta)$ *and the Aggressive Equilibrium* (1) *is played if* $\exists i : q_i < \bar{q}(\theta)$. *If* $c < 2$, *the following subgame perfect equilibrium exists:*

(i) Players choose militarisation levels symmetrically in stage 1, s.t. $q_1 = q_2 = q^*(\theta, c)$ *where*

$$q^*(\theta, c) = \frac{c(\theta - 1/2)}{1 - c(1/2 - \theta/2)}; \tag{6}$$

(ii) Players stick to the Aggressive Equilibrium (1) *in stage 2 on equilibrium path.*

Call this the ***Sequential Aggressive Equilibrium***. *This equilibrium is unique if* $c < \tilde{c}^1(\theta)$. *If* $c \geq \tilde{c}^1(\theta)$, *a second equilibrium exists:*

(i) Players choose militarisation levels symmetrically in stage 1, s.t. $q_1 = q_2 = \bar{q}(\theta)$;

(ii) Players stick to the Deterrent Equilibrium (2) *in stage 2 on equilibrium path.*

*Call this the **Sequential Deterrent Equilibrium**. This equilibrium is unique if $c > 2$. The threshold $\tilde{c}^1(\theta)$ is given by*

$$\tilde{c}^1(\theta) = \frac{\theta - \sqrt{-3\theta^2 + 4\theta - 1}}{\theta - 1/2}. \tag{7}$$

.

The Sequential Aggressive Equilibrium exists when militarisation is sufficiently costly with $c < 2$. The symmetric problem solves for a symmetric Nash equilibrium with $q_1 = q_2 = q^*(\theta, c)$. This Nash equilibrium is part of a subgame perfect equilibrium if the chosen militarisation level satisfies $q^*(\theta, c) < \bar{q}(\theta)$ and induces the Separating Equilibrium. For sufficiently high militarisation costs with $c < 2$, militarisation is below the deterrent threshold and $q^*(\theta, c) < \bar{q}(\theta)$ holds. In this equilibrium, players anticipate high types to fight in stage 2 and militarise to increase the probability of being a high type. Players militarise the less, the costlier militarisation and war are. For high costs of war, low θ, being of high type has less value, so that the incentive to militarise decreases with the cost of war. Militarisation costs directly reduce the incentive to militarise. Moreover, militarisation levels q_1 and q_2 are strategic complements as can be seen in Figure 3. Player i responds to a larger militarisation of his opponent q_j with a larger militarisation q_i to avoid loosing against a high type player j.

The Sequential Deterrent Equilibrium exists for low militarisation costs with $c \geq \tilde{c}^1(\theta)$. Players choose deterrent militarisation $\bar{q}(\theta)$ in stage 1 and do not fight in stage 2. This is an equilibrium if a unilateral deviation to some $q_i < \bar{q}(\theta)$ in stage 1 is unprofitable. A deviation saves on militarisation costs, while it implies a loss in the conflict game as it induces the Aggressive Equilibrium. When militarisation is less costly, the induced loss in the conflict game outweighs the saving in militarisation cost. Players have no incentive to deviate from the deterrent investment $\bar{q}(\theta)$. A deviation is beneficial if militarisation costs are high compared to the cost of war and the cost parameter c is below the threshold $\tilde{c}^1(\theta)$.

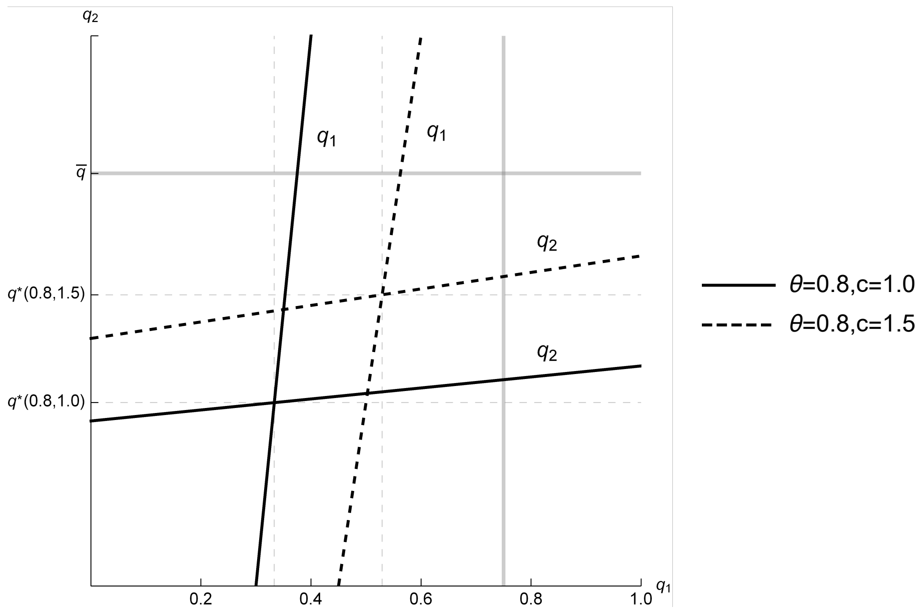


Figure 3: The intersection of the best response functions determines the symmetric militarisation levels in the Sequential Aggressive Equilibrium. The Nash equilibrium choice $q^*(\theta, c)$ increases in c.

This threshold is an increasing function of the reduced pie θ. Figure 5a shows the relation between the threshold $\tilde{c}^1(\theta)$ and the cost of war. When war is very costly and θ close to 0.5, a deviation from deterrent investment $\bar{q}(\theta)$ induces a high loss in the conflict game. A deviation is not profitable even for high militarisation costs. The Deterrent Equilibrium exists for a large range of the cost parameter c. As θ increases, war becomes less costly. Deviating from the deterrent investment $\bar{q}(\theta)$ induces less loss in the conflict game. The Deterrent Equilibrium exists for a smaller range of the cost parameter c.

Two subgame perfect equilibria exist if the cost parameter c takes an intermediate value with $\tilde{c}^1(\theta) \leq c < 2$. I use payoff-dominance as the equilibrium refinement concept in case of multiplicity, as in section 3. Figure 5a depicts the parameter regions for which the Sequential Aggressive and the Sequential Deterrent Equilibrium are played. The Sequential Aggressive Equilibrium is the unique equilibrium in the region below the threshold $\tilde{c}^1(\theta)$. In the region between this threshold and the horizontal line $c = 2$, both sequential equilibria are incentive compatible. The Sequential Aggressive Equilibrium is played as the payoff-dominant equilibrium for parameters below the dotted line, whereas the Sequential Deterrent Equilibrium is played as the payoff-dominant equilibrium for parameters above the dotted line. In the region above $c = 2$, the Sequential Deterrent Equilibrium is the unique. To conclude the discussion of militarisation decisions absent mediation, we can record that players build up large and deterrent militarisation when the cost of militarisation is sufficiently low. When militarisation is less costly, they militarise in preparation for a conflict and high types wages war in stage 2.

4.2 Militarisation with Mediation

With mediation, militarisation has two effects. Militarisation determines the probability with which high and low type players encounter each other in the conflict game, and it determines the optimal signal profile. The sequential game is solved by backward induction. The signal profile described in Proposition 2 remains optimal, as the mediator takes militarisation decisions as given when choosing the signals. When players choose deterrent militarisation levels $\forall i : q_i \geq \bar{q}(\theta)$, the optimal signal profile consists of uninformative signals. Players always observe high realisations and never fight. In this case, the Mediated Equilibrium coincides with the Deterrent Equilibrium. When militarisation levels are not deterrent, the optimal signal profile features at least one signal according to (3). The sequential game with mediation always solves for a unique subgame perfect equilibrium.

Proposition 4. *A unique subgame perfect equilibrium exists:*

(i) Players choose militarisation levels symmetrically in stage 1, s.t.

$$q_1 = q_2 = \begin{cases} q^{**}(\theta, c) & if\ c < \tilde{c}^2(\theta) \\ \bar{q}(\theta) & if\ c \geq \tilde{c}^2(\theta) \end{cases}$$

where

$$q^{**}(\theta, c) = \frac{c(\theta - 1/2)}{1 - c(1/2 - \theta/2)\frac{\theta}{\theta - 1/2}} \tag{8}$$

and

$$\tilde{c}^2(\theta) = \frac{4\theta - 2}{\theta}; \tag{9}$$

(ii) The mediator chooses the optimal signal profile described in Proposition 2;

(iii) Players stick to the Mediated Equilibrium (5) *on equilibrium path.*

Call this the ***Sequential Mediated Equilibrium****. If $c < \tilde{c}^2(\theta)$, the war probability in stage 2 is strictly positive. If $c \geq \tilde{c}^2(\theta)$, players choose deterrent militarisation levels and the Sequential Mediated Equilibrium coincides with the Sequential Deterrent Equilibrium.*

When players expect high types to fight in stage 2, they solve a trade-off between increasing the chance to win a war and saving on militarisation cost, similar to the case without mediation. The symmetric problem solves for a symmetric Nash Equilibrium $q^{**}(\theta, c)$

Mediation makes militarising more profitable. A player benefits the more from mediation, the more he is militarised. With mediation, a low type player i is attacked if his opponent is a high type and receives a low signal realisation. The more player i is militarised, the less frequently the opponent j observes a low signal realisation. This effect is not present in the case without mediation. Figure 4 shows that players militarise more with mediation. The Figure depicts the best response functions with and without mediation. For constant costs of war and militarisation, player i responds with a higher militarisation level q_i to his opponent's choice q_j. The best response functions are strictly steeper for the case with mediation. Militarisation is strictly larger in the Nash equilibrium with mediation, $q^{**}(\theta, c)$, than in the one without mediation, $q^*(\theta, c)$.[17]

[17] The graph also illustrates how the best response function changes once the opponent's militarisation level reaches $\bar{q}(\theta)$. To understand the kink in the best response function of player i at $q_j = \bar{q}(\theta)$, consider the effects an increase in q_j has on player i's incentive to militarise. There is a positive and a negative effect. As q_j increases, player i has a larger incentive to militarise to protect himself from being attacked. This positive effect is present for any value of q_j. A negative effect dampens player i's incentive to militarise. As q_j increases, player i observes less often a low signal realisation and fights less often. Once player j's militarisation level reaches $q_j = \bar{q}(\theta)$, player i does no longer observe low realisations. A further increase in q_j does not longer have the second, negative effect and the overall effect on player i's incentive to militarise is larger.

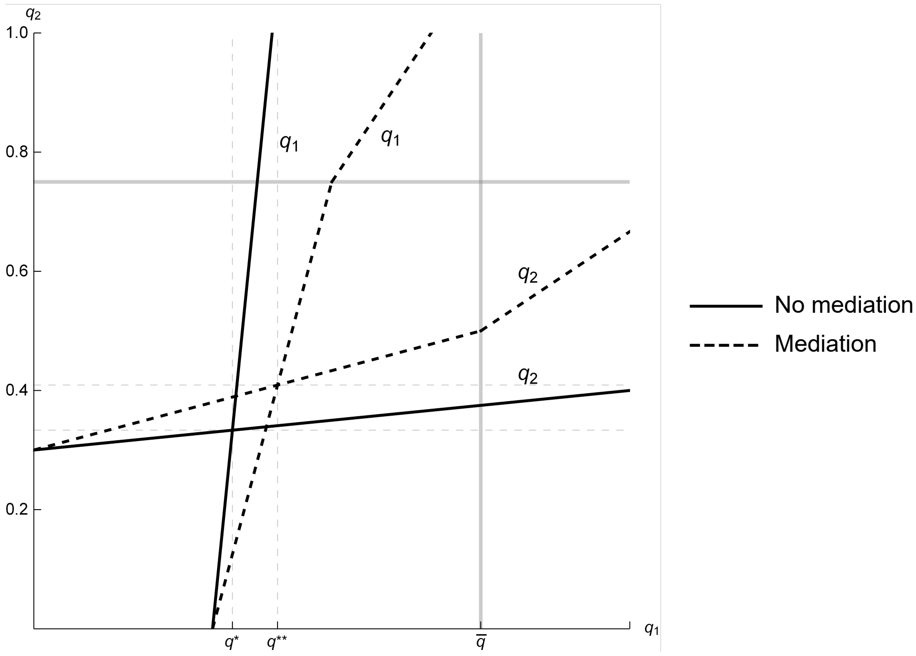
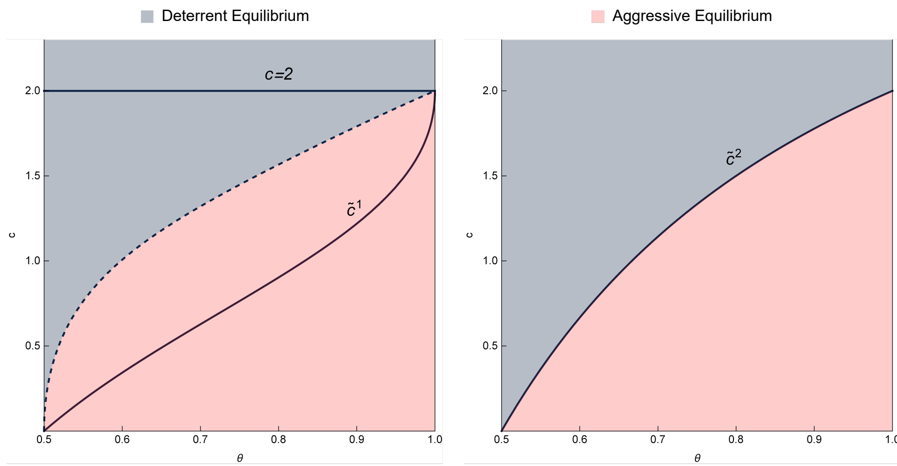


Figure 4: Best response functions with and without mediation for $\theta = 0.8$ and $c = 1.0$. Militarisation levels are strictly larger in the symmetric Nash equilibrium with mediation than in the one without mediation.

The Nash Equilibrium with militarisation level $q^{**}(\theta, c)$ is part of a subgame perfect equilibrium if players fight each other with strictly positive probability in stage 2. This holds if militarisation is not deterrent: $q^{**}(\theta, c) < \bar{q}(\theta)$. Once militarisation reaches the threshold $\bar{q}(\theta)$, the Deterrent Equilibrium is played in stage 2. Players militarise high and deterrent levels $q_1 = q_2 = \bar{q}(\theta)$ if militarisation costs are sufficiently low compared to the cost of war. The costs are sufficiently low if the cost parameter c exceeds the threshold $\tilde{c}^2(\theta)$. Figure 5b shows that this threshold is an increasing function of the size of the reduced pie θ. When war is very costly and θ close to 0.5, less militarisation is deterrent. Players reach deterrent militarisation for higher militarisation costs. Second stage play collapses to the Deterrent Equilibrium for a large range of the cost parameter c. As θ increases and war becomes less costly, militarisation meeds to be larger to be deterrent. The Nash equilibrium militarisation level $q^{**}(\theta, c)$ reaches the deterrent level $\bar{q}(\theta)$ only for lower militarisation costs. The range of c decreases for which second stage play collapses to the Deterrent Equilibrium.



(a) Without Mediation (b) With Mediation

Figure 5: The Sequential Aggressive Equilibrium is played if militarisation in more costly.

4.3 Mediation Benefit with Endogenous Strength

With and without mediation, the sequential game solves for one of two types of equilibria. In the first type of equilibrium, players militarise to increase their chance to win and not to loose a war. In the second type of equilibrium, players militarise to deter their opponent from fighting. Whereas in the former case, players fight with a strictly positive probability, in the latter case, players never fight in equilibrium. Thus, two questions have to be taken into account to assess the benefit of mediation when militarisation levels are endogenous. First, how does mediation affect players' incentives to choose deterrent militarisation levels, such that war is completely avoided? Second, how does mediation affect the war probability when deterrent militarisation is not incentive compatible and players fight with a strictly positive probability in stage 2?[18]

The first question is addressed by comparing the parameter region for which the Deterrent Equilibrium is played in stage 2 so that war is avoided. In both cases, players choose deterrent militarisation if militarisation costs are low and the cost parameter c is large. Mediation increases the parameter region for which second stage play is characterised by the Deterrent

[18] I base the analysis of the mediation benefit on a comparison of the war probability in equilibrium, as I took it as the starting point that mediation aims at reducing the war probability. An alternative would be to compare expected payoffs.

Equilibrium, as can be seen in Figure 5. To understand the reason, it is necessary to recall how mediation affects militarisation when players anticipate war. As discussed, mediation increases the incentive to militarise, because a player benefits the more from mediation, the more he is militarised. Nash equilibrium militarisation is larger when players expect mediation to take place. Consequently, militarisation reaches the deterrent level $\bar{q}(\theta)$ for higher militarisation costs and a lower value of c. Mediation therefore increases the parameter region, for which players choose deterrent militarisation levels in stage 1 and war is avoided. This effect of mediation is positive.

To address the second question, it is necessary to concentrate on that parameter region for which second stage play takes the aggressive form with and without mediation. This is the case if militarisation costs are sufficiently high with $c < \tilde{c}^2(\theta)$. Whether mediation is beneficial here depends on whether it increases or decreases the ex-ante war probability. Mediation affects this probability in a dual way. First, mediation leads to larger militarisation. This effect is indirect and negative. The second effect is the direct mediation benefit, which was discussed in Section 3.3. Once militarisation levels are chosen mediation reduces the war probability. Whether mediation is overall beneficial depends on the size of both effects. Mediation overall reduces the war probability if war is not too costly and θ sufficiently large. Very costly war induces only small militarisation. When players choose small militarisation levels, the mediator needs to use very precise signals, so that players refrain from fighting upon receiving a high signal realisation. When signals are very precise, occurrences of high signal realisations are infrequent and the direct mediation benefit is small. Only when the reduced pie θ is large enough, and, in line with this, militarisation is large enough, the direct benefit of mediation overcompensates the indirect, negative effect.[19]

Assuming that players take mediation into account when militarising allows for a more detailed understanding of mediation. The direct mediation benefit identified in the case with exogenous militarisation persists when militarisation is endogenous. That is, once militari-

[19] A more detailed analysis of this comparison can be found in the appendix.

sation is chosen, mediation decreases the war probability. But mediation also has a perverse effect. Players militarise more when they expect mediation to take place, because a player benefits the more from mediation, the more he is militarised. This indirect effect of mediation increases the probability with which players fight in the conflict stage when deterrent militarisation is not incentive compatible. When war is very costly, this negative effect dominates the direct mediation benefit. In this case, the total effect of mediation is negative.

5 Conclusion

Uncertainty and informational asymmetries are important factors to understand conflict and possible conflict resolution mechanisms. In this paper, I used Bayesian Persuasion to show how altering the informational environment of a conflict situation can prevent war. A mediator generates evidence about the military strength of conflicting parties in the course of mediation, which is valuable to the respective opponent. He passes this information on to players such that they change their behaviour. When a player receives information indicating a strong opponent, he refrains from fighting. I derived conditions for the information to have this effect and how the mediator uses the effect optimally. The effective use of the information rests on the assumptions that the mediator generates information independently and that he commits on transmitting it directly to the conflicting parties. The analysis showed the potential of research and intelligence techniques for preventing war. Mediation in this form is the more effective, the more likely it is that players fight each other and the costlier war is.

In the second part of this paper, mediation was put into a broader context. When players anticipate mediation, they adapt their pre-conflict behaviour. I endogenised players' militarisation by assuming that players costly invest to increase their militarisation. In this setting, players either mutually choose high and deterrent militarisation or they militarise to increase their chance to win a war. With endogenous strength, mediation has two effects on the ex-ante war probability. Mediation increases players' incentives to militarise as players benefit the more from mediation, the more they are militarised. Once militarisation decisions are taken, however, mediation is beneficial, as it reduces the war probability for given militarisation. The overall effect of mediation is not necessarily positive. Mediation can increase the ex-ante war probability when war is very costly. More generally, the result shows that the evaluation of mediation techniques is sensitive to the evaluation's perspective.

Appendix

Proof of Proposition 3 Proof of the Sequential Aggressive Equilibrium: Equilibrium play in stage 2 is characterised by Proposition 1. Suppose players anticipate that militarisation decisions satisfy $\exists i : q_i < \bar{q}(\theta)$ and that the Aggressive Equilibrium is played in stage 2. Under these expectations, player 1 faces the following maximisation problem in stage 1:

$$\underset{q_1}{Max} \left\{ \frac{1}{2} - q_1 q_2 \left(\frac{1}{2} - \frac{\theta}{2} \right) + q_1 (1 - q_2) \left(\theta - \frac{1}{2} \right) - (1 - q_1) q_2 \frac{1}{2} - \frac{1}{2} \frac{1}{c} q_1^2 \right\}$$
$$\text{s.t. } q_1 \in [0, 1]$$

Solving the maximisation problem results in the best response function

$$q_1^*(q_2, \theta, c) = \begin{cases} c\left(\theta - 1/2 + q_2\left(1/2 - \theta/2\right)\right) & \text{for } c \leq \frac{1}{\theta - 1/2 + q_2(1/2 - \theta/2)} \\ 1 & \text{for } c > \frac{1}{\theta - 1/2 + q_2(1/2 - \theta/2)}. \end{cases}$$

The best response function $q_2^*(q_1, \theta, c)$ is symmetric by symmetry of the problem. This allows to derive the symmetric choices $q_1 = q_2 = q^*(\theta, c)$ with

$$q^*(\theta, c) = \begin{cases} \frac{c(\theta - 1/2)}{1 - c(1/2 - \theta/2)} & \text{for } c \leq 2/\theta \\ 1 & \text{for } c > 2/\theta. \end{cases}$$

The Nash equilibrium $q_1 = q_2 = q^*(\theta, c)$ is part of a subgame perfect equilibrium if it is consistent with the expectation under which it is derived, that is if $q^*(\theta, c) < \bar{q}(\theta)$ holds. This condition is satisfied for $c < 2$. Thus, for $c < 2$, the Sequential Aggressive Equilibrium is a subgame perfect equilibrium.[20]

Proof of the Sequential Deterrent Equilibrium: Suppose players anticipate play of the Deterrent Equilibrium in stage 2 when taking militarisation decisions in stage 1. Optimal

[20]Note that if we assumed that players stick to the Aggressive Equilibrium in stage 2 if $\forall i : q_i \geq \bar{q}(\theta)$ (that is, if we did not restrict attention to the payoff-dominant continuation equilibrium), the Sequential Aggressive Equilibrium would be supported for $c \geq 2$.

militarisation is symmetric and given by $q_1 = q_2 = \bar{q}(\theta)$. For these decisions to be part of a subgame perfect equilibrium, a deviation to a different militarisation level in stage 1 must be unprofitable for both players. Consider player 1. A deviation to $q_1 > \bar{q}(\theta)$ cannot be profitable. A deviation to $q_1 < \bar{q}(\theta)$ induces play of the Aggressive Equilibrium in stage 2. Anticipating the Aggressive Equilibrium, player 1 would choose q_1 as an optimal response to $q_2 = \bar{q}(\theta)$. Denote the optimal deviation as $q_1^D(\theta, c) \equiv q_1^*(\bar{q}(\theta), \theta, c)$. It is given by

$$q_1^D(\theta, c) = \begin{cases} \frac{c(\theta - 1/2)}{\theta} & \text{for } c \leq 2 \\ \frac{\theta - 1/2}{\theta/2} & \text{for } c > 2. \end{cases}$$

Computing the expected payoff given the optimal deviation allows to conclude whether a deviation is profitable:

$$\mathbb{E}_1\Pi_1(q_1^D(\theta, c), \bar{q}(\theta)) > \mathbb{E}_1\Pi_1(\bar{q}(\theta), \bar{q}(\theta)) \qquad \text{iff} \qquad c < \tilde{c}^1(\theta) = \frac{\theta - \sqrt{-3\theta^2 + 4\theta - 1}}{\theta - 1/2}.$$

Hence, if $c \geq \tilde{c}^1(\theta)$, a unilateral deviation from $q_i = \bar{q}(\theta)$ is not profitable and the Sequential Deterrent Equilibrium is a subgame perfect equilibrium.

Proof of Proposition 4 Equilibrium play in stage 2 is characterised by Proposition 2. Players anticipate the Mediated Equilibrium to be played in stage 2 when taking militarisation decisions in stage 1. Consider the maximisation problem faced by player 1:

$$\begin{aligned} &\underset{q_1}{Max} \left\{ \frac{1}{2} + q_1(1 - q_2)\pi^1(l \mid L)\left(\theta - \frac{1}{2}\right) - (1 - q_1)q_2\pi^2(l \mid L)\frac{1}{2} - \frac{1}{2}\frac{1}{c}q_1^2 \right\} \\ &\text{s.t. } \pi^1(l \mid L) = max\left\{ \frac{\theta - 1/2 - q_2\theta/2}{(1 - q_2)(\theta - 1/2)}, 0 \right\} \\ &\quad\;\; \pi^2(l \mid L) = max\left\{ \frac{\theta - 1/2 - q_1\theta/2}{(1 - q_1)(\theta - 1/2)}, 0 \right\} \\ &\quad\;\; q_1 \in [0, 1] \end{aligned}$$

The objective function is concave and differentiable in q_1 up to $q_1 = \bar{q}(\theta)$, it has a kink at $q_1 = \bar{q}(\theta)$ and is decreasing and concave afterwards. Remember that for $q_1 \geq \bar{q}(\theta)$, player 2 always receives high realisations about player 1's strength and never fights. For militarisation costs sufficiently high, the objective function reaches its maximum for a value $q_1 < \bar{q}(\theta)$. For militarisation costs below a certain threshold, it is maximised for $q_1 = \bar{q}(\theta)$. Here, militarisation is of deterrent height. The following best response function summarises optimal militarisation for player 1:

$$q_1^{**}(q_2,\theta,c) = \begin{cases} (\theta - 1/2)c + \frac{q_2 c\theta(1-\theta)}{2\theta-1} & \text{for } c \leq -\frac{2(1-2\theta)^2}{\theta(2\theta^2(q_2-2)-2\theta(q_2-2)-1)} \text{ and } q_2 < \bar{q}(\theta) \\ \bar{q}(\theta) & \text{for } c > -\frac{2(1-2\theta)^2}{\theta(2\theta^2(q_2-2)-2\theta(q_2-2)-1)} \text{ and } q_2 < \bar{q}(\theta) \\ \frac{c\theta q_2}{4\theta-2} & \text{for } c \leq \frac{2(1-2\theta)^2}{\theta^2 q_2} \text{ and } q_2 > \bar{q}(\theta) \\ \bar{q}(\theta) & \text{for } c > \frac{2(1-2\theta)^2}{\theta^2 q_2} \text{ and } q_2 > \bar{q}(\theta) \end{cases}$$

The problem faced by player 2 is symmetric and the best response function is symmetrically given. Militarisation levels in the unique Nash equilibrium are given by

$$q_1 = q_2 = \begin{cases} q^{**}(\theta, c) & if\ c < \tilde{c}^2(\theta) \\ \bar{q}(\theta) & if\ c \geq \tilde{c}^2(\theta) \end{cases}$$

where

$$q^{**}(\theta,c) = \frac{c(\theta - 1/2)}{1 - c(1/2 - \theta/2)\frac{\theta}{\theta - 1/2}}$$

and

$$\tilde{c}^2(\theta) = \frac{4\theta - 2}{\theta}.$$

Comparing the ex-ante war probability with Aggressive Play in Stage 2 The ex-ante war probability in the Sequential Aggressive Equilibrium, in which players militarise and there is no mediation is given as[21]

$$\begin{aligned}\boldsymbol{W}^{NM}(q^*, q^*) &= (q^*)^2 + 2q^*(1-q^*) \\ \text{with } q^* &= \frac{c(\theta - 1/2)}{1 - c(1/2 - \theta/2)}.\end{aligned}$$

The ex-ante war probability in the Sequential Mediated Equilibrium, in which players militarise and the mediator uses the optimal mediation mechanism as described in Proposition 3 is given as

$$\begin{aligned}\boldsymbol{W}^{M}(q^{**}, q^{**}, \theta) &= 2q^{**}(1-q^{**})\pi(l \mid L) \\ \text{with } \pi(l \mid L) &= \frac{\theta - 1/2 - q^{**}\theta/2}{(1-q^{**})(\theta - 1/2)} \\ \text{and } q^{**} &= \frac{c(\theta - 1/2)}{1 - c(1/2 - \theta/2)\frac{\theta}{\theta - 1/2}}.\end{aligned}$$

The parameter region for which players fight with a strictly positive probability in both sequential equilibria is relevant for the comparison. This restricts the attention to $c < \tilde{c}^1(\theta)$. Within this parameter region, a region exists, for which the ex-ante war probability is larger with mediation than without mediation:

$$\begin{aligned}&\boldsymbol{W}^{M}(q^{**}, q^{**}, \theta) > \boldsymbol{W}^{NM}(q^*, q^*) \;\; iff \\ &\qquad \theta \in \left(\frac{1}{2}, \frac{3}{4}\right) \;\; and \\ &\qquad c \in \left(0, \frac{4\theta - 1}{\theta} - \sqrt{\frac{16\theta^4 - 24\theta^3 + 13\theta^2 - 4\theta + 1}{(\theta-1)^2\theta^2}}\right).\end{aligned}$$

If costs of war and militarisation are high, low θ and c, mediation has an overall negative

[21]To simplify the notation in this section, I omit the the arguments when referring to the value of a function, that is, I write q^* instead of $q^*(\theta, c)$, for example.

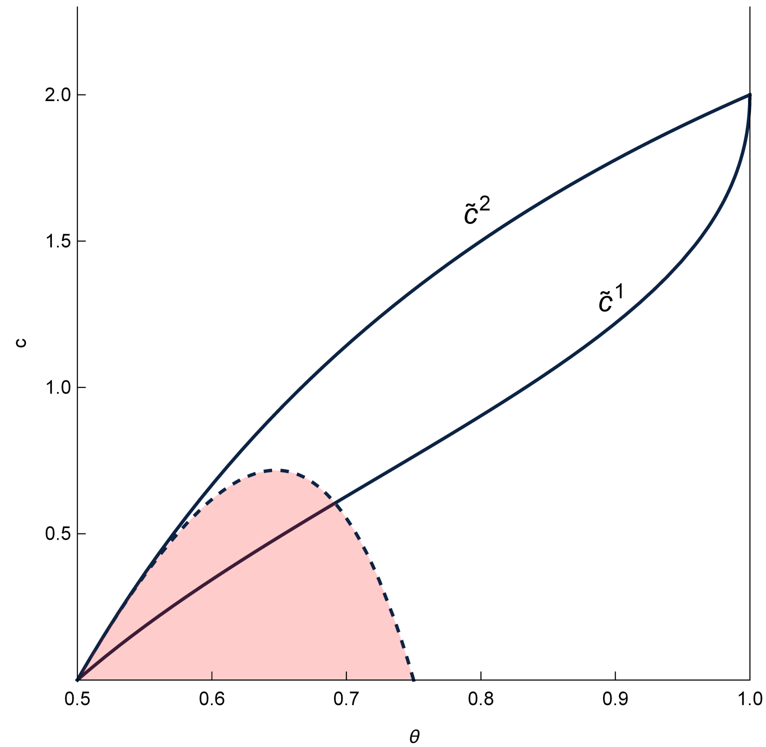


Figure 6: For parameters in the red region, the ex-ante war probability is higher with mediation than without mediation.

effect on the ex-ante war probability. Figure 6 depicts this parameter region together with the thresholds $\tilde{c}^1(\theta)$ and $\tilde{c}^2(\theta)$.

A Note on Cheap-Talk Communication To illustrate the role of asymmetric information in the conflict game, I show that allowing communication between players does not provide a straightforward solution to the problem. Suppose a cheap-talk communication stage takes place prior to the conflict game. Players send messages $m_i \in \{h, l\}$. Messages are observed simultaneously by both players. The game proceeds to stage 2 and players take simultaneously an action $a_i \in \{f, nf\}$. A strategy for player i is now composed of a function determining the message sent in stage 1, $\mu_i : \{H, L\} \to \{h, l\}$ and a function determining the action taken in stage 2, $\phi_i : \{H, L\} \times \{h, l\} \times \{h, l\} \to \{f, nf\}$. This function takes the message sent by player i as its second argument and the message sent by player j the third argument. I continue to assume that players choose $a_i = nf$ upon indifference. As noted before, this implies $\forall m_i, m_j : \phi_i(L, m_i, m_j) = nf$. Suppose the communication strategy used by player 2 was such that player 1 made his action in stage 2 conditional on the observed message. This strategy cannot be incentive compatible for player 2. Player 2 could profit by deviating in stage 1 and always sending that message which induces player 1 to accept the default division. Hence, player 2 cannot employ an informative communication strategy in stage 1. As the same reasoning applies symmetrically, the cheap talk communication stage cannot alter the equilibrium play in stage 2. This reasoning breaks down if we depart from the assumption that players accept the default division upon indifference. Not maintaining this assumption, an equilibrium exists, in which players reveal their types truthfully, and high-low type pairs fight each other in stage 2. This equilibrium is incentive compatible if parameters satisfy $\forall i : q_i \leq \frac{1-\theta}{2-\theta}$.

References

Alonso, R., and O. Câmara (2016): "Persuading Voters," *American Economic Review*, 106(11), 3590–3605.

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

Bercovitch, J., J. T. Anagnoson, and D. L. Wille (1991): "Some Conceptual Issues and Empirical Trends in the Study of Successful Mediation in International Relations," *Journal of Peace Research*, 28(1), 7–17.

Bercovitch, J., and A. Houston (2000): "Why Do They Do It Like This?," *Journal of Conflict Resolution*, 44(2), 170–202.

Bergemann, D., and S. Morris (2018): "Information Design: A Unified Perspective," *Journal of Economic Literature*, forthcoming.

Blainey, G. (1988): *The Causes of War*. Free Press, New York, 3. ed. edn.

Brecher, Michael, J. Wilkenfeld, K. Beardsley, P. James, and D. Quinn (2017): "International Crisis Behavior Data Codebook, Version 12.," .

Brecher, M., and J. Wilkenfeld (1997): *A Study of Crisis*. University of Michigan Press, Ann Arbor, MI.

Canidio, A., and J.-M. Esteban (2018): "Benevolent Mediation in the Shadow of Conflict," Discussion paper, Barcelona Graduate School of Economics.

Chan, J., S. Gupta, F. Li, and Y. Wang (2017): "Pivotal Persuasion," Discussion paper.

Crawford, T. W. (2008): *Gambling on Humanitarian Intervention: Moral hazard, rebellion and civil war*. Routledge, Abingdon.

Crescenzi, M. J., K. M. Kadera, S. M. Mitchell, and C. L. Thyne (2011): "A Supply Side Theory of Mediation1," *International Studies Quarterly*, 55(4), 1069–1094.

Ely, J. C. (2017): "Beeps," *American Economic Review*, 107(1), 31–53.

Fearon, J. D. (1995): "Rationalist Explanations for War," *International Organization*, 49(03), 379.

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

——— (2010): "When Is Shuttle Diplomacy Worth the Commute? Information Sharing through Mediation," *World Politics*, 62(04), 529–560.

——— (2011): "Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict," *American Journal of Political Science*, 55(1), 149–169.

Freedman, L. (2007): *The official history of the Falklands campaign*. Routledge, London, rev. and updated ed. edn.

Hörner, J., M. Morelli, and F. Squintani (2015): "Mediation and Peace," *The Review of Economic Studies*, 82(4), 1483–1501.

Inostroza, N., and A. Pavan (2018): "Persuasion in Global Games with Application to Stress Testing," Discussion paper, Northwestsern University.

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

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

Kydd, A. H. (2006): "When Can Mediators Build Trust?," *American Political Science Review*, 100(03), 449–462.

——— (2010): "Rationalist Approaches to Conflict Prevention and Resolution," *Annual Review of Political Science*, 13(1), 101–121.

Laipson, E., D. Kurtzer, and J. L. Moore (1995): "Intelligence and the Middle East: What Do We Need To Know?," .

Mathevet, L., J. Perego, and I. Taneva (2017): "On Information Design in Games," Discussion paper, New York University.

Meirowitz, A., M. Morelli, K. W. Ramsay, and F. Squintani (2018): "Dispute Resolution Institutions and Strategic Militarization," *Journal of Political Economy*, forthcoming.

Powell, R. (1996): "Bargaining in the Shadow of Power," *Games and Economic Behavior*, 15(2), 255–289.

——— (2002): "Bargaining Theory and International Conflict," *Annual Review of Political Science*, 5(1), 1–30.

Ramsay, K. W. (2017): "Information, Uncertainty, and War," *Annual Review of Political Science*, 20(1), 505–527.

Rauchhaus, R. W. (2006): "Asymmetric Information, Mediation, and Conflict Management," *World Politics*, 58(02), 207–241.

——— (2009): "Principal-Agent Problems in Humanitarian Intervention: Moral Hazards, Adverse Selection, and the Commitment Dilemma," *International Studies Quarterly*, 53(4), 871–884.

Wang, Y. (2015): "Bayesian Persuasion with Multiple Receivers," Discussion paper, Wang Yanan Institute for Studies in Economics (WISE), Xiamen University.

Wittman, D. (1979): "How a War Ends," *Journal of Conflict Resolution*, 23(4), 743–763.

Zhang, J., and J. Zhou (2016): "Information Disclosure in Contests: A Bayesian Persuasion Approach," *The Economic Journal*, 126(597), 2197–2217.