

Hennigs, Raphaela

Article — Published Version

Conflict prevention by Bayesian persuasion

Journal of Public Economic Theory

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Hennigs, Raphaela (2021) : Conflict prevention by Bayesian persuasion, Journal of Public Economic Theory, ISSN 1467-9779, Wiley, Hoboken, NJ, Vol. 23, Iss. 4, pp. 710-731, <https://doi.org/10.1111/jpet.12511>

This Version is available at:

<https://hdl.handle.net/10419/284843>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Conflict prevention by Bayesian persuasion

Raphaela Hennigs 

Department of Public Economics, Max Planck Institute for Tax Law and Public Finance, Munich, Germany

Correspondence

Raphaela Hennigs, Department of Public Economics, Max Planck Institute for Tax Law and Public Finance, Marstallplatz 1, Munich 80539, Germany.

Email: raphaela.hennigs@tax.mpg.de

Funding information

Max-Planck-Gesellschaft

Abstract

Drawing upon the Bayesian persuasion literature, I show that a mediator can provide conflicting parties strategically with information to decrease the ex-ante war probability. In a conflict between two parties with private information about military strength, the mediator generates information about each conflicting party's strength and commits to sharing the obtained information with the respective opponent. The conflicting parties can be convinced not to fight each other. The conflicting parties benefit from mediation, as the ex-ante war probability is reduced. The benefit is taken up by weak conflicting parties. This benefit is larger when war is costlier and when the war probability absent mediation is higher.

1 | INTRODUCTION

Understanding conflict resolution and prevention is 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 named as either an “important” or the “most important” factor for easing tensions (see Brecher & Wilkenfeld, 1997; Brecher et al., 2017). This paper contributes to the theoretical understanding of effective mediation by showing how a mediator can use research and intelligence to reduce the ex-ante war probability. Specifically, a mediator can convince conflicting parties not to fight each other by strategically providing them information about their respective opponent.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of Public Economic Theory* published by Wiley Periodicals LLC

Illustrative examples for this idea are aerial reconnaissance missions. Mediators conduct observational flights or take satellite pictures to obtain information about a conflicting party's military power and activity in a conflict and can provide the opponent with this information. 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 pictures were reported to both sides in each case (Laipson et al., 1995). During the Falklands War, the United States provided Argentina with photographic coverage of the Falklands Islands and open sea areas Argentina had asked for based on an agreement with NASA (Freedman, 2007). Potentially, these pictures could have shown the movements of the British fleet. How can the provision of such information change conflicting parties' behavior? If the information indicates that a conflicting party is strong, the opponent has less incentive to fight. On the downside, if the information indicates that the conflicting party is weak, this may increase the opponent's incentive to fight. The result of observational flights or satellite pictures might, therefore, prevent or trigger war. I show that a mediator can provide conflicting parties strategically with information to decrease the ex-ante war probability. I do so in a game-theoretic setting under full rationality, meaning that the conflicting parties are aware of the mediator's incentive to provide information strategically, so as to lower the ex-ante war probability.

This idea is modeled in a conflict game between two conflicting parties with mediation as information provision using a Bayesian persuasion framework (see Kamenica & Gentzkow, 2011). Two conflicting parties divide a pie of unit size with the default division being an equal split. The conflicting parties simultaneously decide whether to fight their opponent. Each conflicting party is strong or weak and does not know whether its opponent is strong or weak. The probability with which a conflicting party is strong is given by its militarization level, which is common knowledge. Absent mediation, this uncertainty can lead to war. As a strong conflicting party always wins against a weak opponent, it fights if the probability to encounter a weak opponent is sufficiently high. A mediator can decrease the ex-ante war probability by decreasing conflicting parties' uncertainty and providing them with information indicating whether each opponent is strong or weak. In the context of international relations, such information could be for instance: whether or not a conflicting party has recently made large investments into military capabilities; whether or not a conflicting party has taken measures in preparation for a war, by, for example, moving troops close to the border; whether or not a conflicting party has a certain type of military equipment, such as specific missiles. The mediator can decide how precise the provided information is.

I derive how a mediator provides information optimally to minimize the ex-ante war probability. Mediation decreases the ex-ante war probability the more, the costlier war is and the more likely it is that conflicting parties fight when the mediator does not provide any information. The conflicting parties benefit from mediation as a costly war is less likely. Using Bayesian persuasion to model mediation stresses a strategic aspect of the information-collecting process during mediation.

2 | LITERATURE REVIEW

The modeling approach in this paper draws on Bayesian persuasion as developed by Kamenica and Gentzkow (2011). Their contribution was one among the first in the growing literature on information design.¹ Information design asks how a sender optimally provides information to

¹See Bergemann and Morris (2019) for an overview of this literature.

one or multiple receivers to induce a certain equilibrium behavior.² Kamenica and Gentzkow (2011) focus on the case of one sender and one receiver with symmetric information.

Further work has extended Bayesian persuasion to games with multiple receivers.³ Few papers address these games in a general form. (See Bergemann & Morris, 2019; Mathevet et al., 2020 and Taneva, 2019.) More often, specific games with multiple receivers have been analyzed, a large group being voting games/games of collective decision making. Alonso and Câmara (2016) study a voting game between a sender and a group of voters with heterogeneous preferences, in which the sender can use a public signal to influence voters' choice over different policy options. They show that the sender can exploit the heterogeneity of voters' preferences by choosing a signal, which targets different winning coalitions. Whereas Alonso and Câmara (2016) focus on a public signal, Chan et al. (2019) allow for private signals in a voting game, in which voters are heterogeneous with respect to voting costs. They show that the sender benefits from using private and correlated signals. In both papers, voters share a common prior belief about the state of the world, being a single random variable. In contrast, I analyze a setting with uncertainty about two random variables, that is, conflicting parties' strengths, and each of both conflicting parties being privately informed about its own strength.⁴

To the best of my knowledge, only Balzer and Schneider (2019) have studied conflict prevention from an information design perspective. They consider a general class of conflicts between two conflicting parties with private information about strengths. Conflict can either be solved peacefully or escalates to war. In the case of war, payoffs are determined endogenously by actions taken by the conflicting parties. Balzer and Schneider (2019) show the equivalence between formulating these conflicts as arbitration problems and as information design problems. While Balzer and Schneider (2019) look at a class of conflicts, I focus on a specific conflict game. This allows for an intuitive understanding of the information design problem and its solution. Moreover, I provide comparative statics which illustrate the effect of mediation.

Arbitration with exogenous conflict escalation has been analyzed extensively. With exogenous escalation, payoffs are determined exogenously once war occurs, implying that information revealed during arbitration does not affect payoffs in case of war. Arbitration games have been analyzed, among others, in Bester and Warneryd (2006), Bester and Warneryd (2006), Fey and Ramsay (2009, 2011), and Hörner et al. (2015). As the literature on arbitration is also interested in peaceful conflict resolution the arbitrator can be interpreted as a mediator. Modeling mediation as arbitration assumes that the mediator has means to control the payoff scheme of a conflict. Formulating mediation as an information design problem does not rest upon this assumption, but assumes that the mediator can control the information the parties in a conflict hold. The different assumption distinguished the present paper from the arbitration literature.

The present paper also adds to the literature addressing mediation in sender-receiver games. These games typically model the mediator as a self-interested player in a multi-stage game, who is equipped with some private information. Potentially, the use of this information allows a peaceful conflict resolution. The central question in this setting is if or under which conditions

²The use of terminology in this literature is not consistent. For instance, the "sender" is also referred to as "information designer," "mediator," or "principal," and a "receiver" is also referred to as "agent" or "player."

³Kamenica (2019) provides a recent survey on the literature of Bayesian persuasion, covering some of the extensions.

⁴Further applications of Bayesian persuasion which are related to the present paper look at contest models (see Chen, 2019; Feng & Lu, 2016; Zhang & Zhou, 2016).

the mediator can credibly communicate with the conflicting parties. Communication is modeled as cheap-talk. The models by Kydd (2003), Rauchhaus (2006), and Kydd (2006) stress the question of whether bias or impartiality enhance mediation success. Smith and Stam (2003) address this question in a multiperiod model of war and argue that biased mediators cannot communicate credibly. Fey and Ramsay (2010) elaborate on the cheap-talk framework by asking how a mediator can obtain private information in the first place. Mediation in the present paper differs from mediation in sender-receiver games. Whereas the above models assume cheap-talk communication, I assume that the mediator commits to a signal which provides information to the conflicting parties. The commitment assumption guarantees the credibility of the provided information. The present analysis therefore focuses on the question how to optimally obtain information about conflicting parties instead of how to distribute given information.

From a more general perspective, this paper also contributes to literature applying models of information design and strategic information transmission to real world examples. Applications have been made, for instance, in the literature on public goods (See, among others, Agastya et al., 2007; Barbieri & Malueg, 2019; Palfrey & Rosenthal, 1991.), in the contest literature (e.g. Colef & Garcia, 2017) or in contexts in which senders care about their reputation (Garrido et al., 2020; Ottaviani & Sørensen, 2006; Zimper & Molefinyane, 2020).

3 | THE MODEL

3.1 | Setup

3.1.1 | Conflict stage

The game consists of a mediation stage and a subsequent conflict stage. The conflict stage follows Hörner et al. (2015). The three players of the game are two conflicting parties $i \in \{1, 2\}$ and a mediator. Throughout, I refer to conflicting party j as i 's opponent, assuming $i \neq j$. All players are risk neutral. The two conflicting parties are heterogeneous with respect to *strength* $t_i \in T = \{H, L\}$, where H denotes high strength and L denotes low strength. I refer to a conflicting party with high strength as *strong* and to a conflicting party with low strength as *weak*. The probability with which conflicting party i is strong is given by its *militarization level* q_i .

Strength t_i is private information of conflicting party i . Militarization levels q_1 and q_2 are commonly known by both conflicting parties and the mediator.

The two conflicting parties dispute over a pie of unit size. The conflicting parties simultaneously take an action $a_i \in \{f, nf\}$, where f denotes *fight* and nf denotes *not fight*. If at least one conflicting party fights, war occurs. If both conflicting parties do not fight, no war occurs.

If no war occurs, each conflicting party receives $1/2$. If war occurs, the unit pie shrinks to the *reduced pie* θ , with $\theta \in (1/2, 1)$.⁵ The cost of war $(1 - \theta)$ decreases in the size of the reduced pie θ . If the two conflicting parties are of equal strength, the reduced pie is divided equally among them. If a strong conflicting party encounters a weak conflicting party,

⁵Note that for $\theta \leq 1/2$, each conflicting party would always prefer to receive $1/2$ and never fight.

the strong one receives the whole reduced pie, whereas the weak one walks away empty-handed.⁶ Let $u_i(a_i, a_j, t_i, t_j)$ denote the payoff, conflicting party i receives.

Note that the cost of war $(1 - \theta)$ is the same regardless of whether only one or both conflicting parties fight. Further, the conflicting parties do not incur any private cost upon fighting. Hence, if party i fights, its payoff $u_i(a_i, a_j, t_i, t_j)$ does not depend on the opponent's action a_j .

3.1.2 | Mediation stage

Before the conflicting parties' strengths realize, the mediator shapes the informational environment of the conflict stage by choosing a signal profile $\pi = \{\pi_1, \pi_2\}$ publicly and at no cost. Signal π_i is addressed at conflicting party i and can provide information about j 's strength. The signal π_i consists of a binary realization space $S = \{h, l\}$, with elements denoted as s_i , and a pair of conditional probability distributions $\{\pi_i(\cdot | t_j)\}_{t_j=H,L}$ over S . I refer to $s_i = h$ as a *high signal realization* and to $s_i = l$ as a *low signal realization*. Signal π_i prescribes, for example, that conflicting party i receives a high signal realization with probability $\pi_i(h|H)$ given that its opponent j is strong. Throughout and without loss of generality, I restrict attention to signals for which $\pi_i(h|H) \geq \pi_i(l|H)$ holds. The mediator receives a payoff of (-1) if war occurs and a payoff of 0 if no war occurs. Let $v(a_i, a_j)$ denote the mediator's payoff.

I restrict the analysis to a profile of uncorrelated signals. This is at no cost, as the mediator could not improve upon the obtained results by choosing correlated signals.⁷ The reason why the two signals can be chosen independently is twofold: As noted above, the payoff of a strong conflicting party who chooses to fight is independent of its opponent's action. Second, by choosing the two uncorrelated signals optimally, the mediator can reduce the probability of war when both conflicting parties are strong to zero.

Upon receiving a signal realization s_i of the signal π_i , conflicting party i forms the posterior belief $q_j^{s_i}$ about its opponent's strength. The posterior belief q_j^h is the probability with which j is strong given that i receives a high signal realization, and q_j^l is the probability with which j is strong given that i receives a low signal realization:

$$q_j^h = \text{Prob}(t_j = H | s_i = h) = \frac{\pi_i(h|H)q_j}{\pi_i(h|H)q_j + \pi_i(h|L)(1 - q_j)},$$

$$q_j^l = \text{Prob}(t_j = H | s_i = l) = \frac{\pi_i(l|H)q_j}{\pi_i(l|H)q_j + \pi_i(l|L)(1 - q_j)}.$$

Consider, for example, the formula which describes the posterior belief q_j^h : The denominator states the total probability with which i receives a high signal realization given signal π_i . The numerator states the joint probability of i receiving a high signal realization and j being strong, given as the prior probability of j being strong multiplied by the conditional probability of a high signal realization given j being strong. The ratio is, thus, the posterior probability of j being strong given that i receives a high signal realization.

⁶ I simplify the model used in Hörner et al. (2015) by assuming that a strong conflicting party receives $\theta \leq 1/2$ with probability 1 when fighting against a low type, while Hörner et al. (2015) allow this probability to take any interior value greater than $1/2$.

⁷ In Appendix A, I show this formally by generalizing the analysis and allowing for correlated signals.

Two special cases of a signal are the fully informative signal and an uninformative signal. The fully informative signal π_i features $\pi_i(h|H) = 1$ and $\pi_i(h|L) = 0$. Upon receiving a realization of the fully informative signal, conflicting party i perfectly learns its opponent's strength. A signal is uninformative if $\pi_i(h|H) = \pi_i(h|L) \in [0, 1]$. Upon receiving the realization of an uninformative signal, conflicting party i does not learn anything about its opponent's strength. Denote the profile of fully informative signals as π^{FI} , and a profile of uninformative signals as π^{NI} . I will refer to π^{FI} and π^{NI} in the later analysis. Signal π_i is informative if $\pi_i(h|H) > \pi_i(h|L)$. If signal π_i is informative, conflicting party i 's belief about its opponent's strength strictly increases after receiving a high signal realization and strictly decreases after receiving a low signal realization. The precision of signal π_i strictly increases in the difference between $\pi_i(h|H)$ and $\pi_i(h|L)$. The more precise signal π_i is, the more conflicting party i learns by receiving a realization of it, meaning that the difference between the posterior beliefs q_j^h and q_j^l is larger.

3.1.3 | Timing

The timing is as follows:

- Stage 1 The mediator publicly chooses the signal profile π .
- Stage 2 The conflicting parties strengths t_1 and t_2 realize.
- Stage 3 Signal realizations s_1 and s_2 realize according to π , each conflicting party i receives the signal realization s_i and forms the posterior belief $q_j^{s_i}$.
- Stage 4 The conflicting parties take actions a_1 and a_2 and payoffs realize.

3.1.4 | Militarization levels

Consider the expected payoff of conflicting party i , assuming i is strong and assigns the probability q_j to its opponent being strong. Then i prefers war over no war if

$$1/2 + (1 - q_j)(\theta - 1/2) - q_j(1/2 - \theta/2) \geq 1/2.$$

This condition can be reformulated to

$$\frac{q_j}{1 - q_j} \leq \frac{\theta - 1/2}{1/2 - \theta/2}. \tag{1}$$

The right-hand side of inequality (1) states the ratio of the benefit when fighting a weak opponent and the loss when fighting a strong opponent in war. The left-hand side states the probability ratio of encountering a strong and a weak opponent. Let $\bar{q}(\theta)$ be defined such that $q_j = \bar{q}(\theta)$ solves inequality (1) with equality. If $q_j < \bar{q}(\theta)$, i strictly prefers war over no war. If $q_j > \bar{q}(\theta)$, i strictly prefers no war over war. I refer to militarization level q_j as *deterrent* if $q_j \geq \bar{q}(\theta)$. If both parties have deterrent militarization levels, war can always be avoided and mediation is needless. Therefore, I restrict the further analysis:

Assumption 1 (Parameter Region). At least one conflicting party does not have a deterrent militarization level.

4 | ANALYSIS

4.1 | Strategies and expected payoffs

A pure strategy of conflicting party i is defined as $\phi_i: T \times S \rightarrow \{f, nf\}$. For a given signal profile π , a Bayesian Nash equilibrium in pure strategies in the conflict stage consists of a strategy profile $\{\phi_1^*, \phi_2^*\}$ such that for all i , $a_i = \phi_i^*(t_i, s_i)$ maximizes the expected payoff $E_i[u_i(a_i, a_j, t_i, t_j)|t_i, s_i]$ for all t_i and s_i given ϕ_j^* .

I refine the set of equilibria by focusing on payoff-undominated equilibria: If multiple Bayesian Nash equilibria exist, parties never stick to a payoff-dominated one. Consequently, if exactly one payoff-undominated equilibrium exists, parties stick to it.

The expected payoff $E_i[u_i(a_i, a_j, t_i, t_j)|t_i, s_i]$ is defined at the end of stage 3 given the strength t_i of conflicting party i and the received signal realization s_i . Expectations are taken over the opponent's strength t_j and the opponent's equilibrium action $a_j = \phi_j^*(t_j, s_j)$ using the posterior belief $q_j^{s_i}$. If a pure strategy equilibrium $\{\phi_1^*, \phi_2^*\}$ exists in the conflict stage for a given signal profile π , the ex-ante expected payoff of a conflicting party i can be defined at the end of stage 1 as $E[u_i(a_i, a_j, t_i, t_j)|\pi, \phi_i^*, \phi_j^*]$. Ex-ante refers to the realization of the conflicting parties' strengths and the signals realizations and the ex-ante expected payoff is derived by taking expectations over the realizations of conflicting parties' strengths and the signal realizations. I abbreviate the notation by writing $E[u_i|\pi, \phi_i^*, \phi_j^*] \equiv E[u_i(a_i, a_j, t_i, t_j)|\pi, \phi_i^*, \phi_j^*]$. If the equilibrium $\{\phi_1^*, \phi_2^*\}$ is unique given signal profile π , I further abbreviate the notation to $E[u_i|\pi]$.

The expected payoff of the mediator is denoted as $E[v(a_1, a_2)|\pi, \phi_1^*, \phi_2^*]$ and equals the negative of the ex-ante war probability. Analogously, I abbreviate to the notation to $E[v|\pi, \phi_1^*, \phi_2^*]$, and to $E[v|\pi]$ if the equilibrium $\{\phi_1^*, \phi_2^*\}$ is unique given signal profile π .

Two observations regarding the conflicting parties' strategies simplify the analysis.

Observation 1. A weak conflicting party never strictly prefers to fight in the conflict stage, independently of the belief it holds about its opponent's strength.

Observation 2. No signal exists that induces a strong conflicting party to never fight unless its opponent has a deterrent militarization level.

To understand Observation 2, remember that a strong conflicting party strictly prefers war over no war unless the opponent has a deterrent militarization level and that any informative signal π_i induces a posterior belief $q_j^i < q_j$.

I begin the analysis by considering two benchmark cases, being a profile of uninformative signals π^{NI} (Section 4.2) and the profile of fully informative signals π^{FI} (Section 4.3). In Section 4.4, I derive the signal profile which maximizes the mediator's ex-ante expected payoff (i.e. minimizes the ex-ante war probability).

4.2 | No information

Assume the mediator does not provide any information and chooses a profile of uninformative signals π^{NI} . As conflicting parties do not learn anything by receiving a signal realization, $q_j^{s_i} = q_j$ holds for all s_i and all i . A unique equilibrium exists in the conflict stage:

Proposition 1 (No Information Case). *If the mediator does not provide any information to the conflicting parties about their opponents' strengths, the following symmetric equilibrium always exists in the conflict stage: A strong conflicting party fights and a weak conflicting party does not fight. Call this the Aggressive Equilibrium. Under Assumption 1 (Parameter Region) the Aggressive Equilibrium is the unique payoff-undominated equilibrium in the no information case.*

Proof. I prove (a) the existence of the Aggressive Equilibrium, and (b) show that the Aggressive Equilibrium is the unique payoff-undominated equilibrium under Assumption 1.

(a) In the Aggressive Equilibrium, both conflicting parties stick to the separating strategy

$$\phi^*(t_i, s_i) = \begin{cases} f & \text{if } t_i = H \\ nf & \text{if } t_i = L \end{cases} \quad \forall s_i. \tag{2}$$

Given $\pi = \pi^{NI}$ and $\phi_j = \phi^*$, the expected payoff of a weak conflicting party i is given by

$$\begin{aligned} & E_i[u_i(a_i, a_j, L, t_j)|s_i] \\ &= \begin{cases} 1/2 - (1 - q_j)(1/2 - \theta/2) - q_j/2 & \text{if } a_i = f \\ 1/2 - q_j/2 & \text{if } a_i = nf \end{cases} \end{aligned}$$

for all s_i . As $E_i[u_i(nf, a_j, L, t_j)|s_i] > E_i[u_i(f, a_j, L, t_j)|s_i]$ holds given $\theta < 1$, $a_i = nf$ is optimal for a weak conflicting party. Further, the expected payoff of a strong conflicting party i is given by

$$\begin{aligned} & E_i[u_i(a_i, a_j, H, t_j)|s_i] \\ &= \begin{cases} 1/2 + (1 - q_j)(\theta - 1/2) - q_j(1/2 - \theta/2) & \text{if } a_i = f \\ 1/2 - q_j(1/2 - \theta/2) & \text{if } a_i = nf \end{cases} \end{aligned}$$

for all s_i . As, $E_i[u_i(f, a_j, H, t_j)|s_i] > E_i[u_i(nf, a_j, H, t_j)|s_i]$ holds given $\theta > 1/2$, $a_i = f$ is optimal for a strong conflicting party. Hence $\phi_i = \phi^*$ is the optimal reply to $\phi_j = \phi^*$ and the Aggressive Equilibrium exists given $1/2 < \theta < 1$. This reasoning holds independently of the militarization levels q_1 and q_2 .

(b) The Aggressive Equilibrium is further the unique payoff-undominated equilibrium:
(b.1)

An asymmetric equilibrium in pure strategies does not exist: Consider an asymmetric strategy profile in which i sticks to the separating strategy ϕ^* and j either never fights or always fights. It follows from part (a) of the proof that this is not an equilibrium, as $\phi_j = \phi^*$ is the optimal reply to $\phi_i = \phi^*$. Further, assume i never fights and j always fights. This cannot be an equilibrium, as j would have an incentive to deviate to nf if being weak.

(b.2)

An equilibrium in which both conflicting parties never fight is ruled out by Assumption 1: Assume j sticks to the pooling strategy

$$\phi(t_j, s_j) = \begin{cases} nf & \text{if } t_j = H \\ nf & \text{if } t_j = L \end{cases} \quad \forall s_j. \quad (3)$$

Given $\pi = \pi^{NI}$ and j sticks to 3, the expected payoff of a strong conflicting party i is given by

$$E_i[u_i(a_i, nf, H, t_j)|s_i] = \begin{cases} 1/2 + (1 - q_j)(\theta - 1/2) & \text{if } a_i = f \\ -q_j(1/2 - \theta/2) & \\ 1/2 & \text{if } a_i = nf \end{cases}$$

for all s_i . It is optimal for i to choose $a_i = nf$ if and only if $q_j \geq \bar{q}(\theta)$. By Assumption 1, there is at least one opponent j with $q_j < \bar{q}_j$, such that there is at least one conflicting party i for which the pooling strategy is not optimal.⁸

(b.3)

An equilibrium in which both conflicting parties always fight is payoff dominated by the Aggressive Equilibrium. The same applies to any equilibrium in nondegenerate mixed strategies.

□

To understand the intuition for the Aggressive Equilibrium, consider a strong conflicting party i which expects a strong opponent j to fight and a weak opponent not to fight. Hence, i can trigger war only if the opponent j is weak. If a strong conflicting party i did not fight, it would forego the potential benefit from fighting a weak opponent. Hence, a strong conflicting party i replies optimally by fighting. This reasoning holds independently of the militarization levels. Moreover, the Aggressive Equilibrium is unique if at least one conflicting party strictly prefers war over no war if it is strong. This holds under Assumption 1.

In the Aggressive Equilibrium, war occurs if at least one conflicting party is strong. The mediator's ex-ante expected payoff for choosing an uninformative signal profile π^{NI} is given by

$$E[v|\pi^{NI}] = -[q_1q_2 + q_1(1 - q_2) + (1 - q_1)q_2]. \quad (4)$$

4.3 | Full information

Assume the mediator chooses π^{FI} , the profile of fully informative signals. Each conflicting party perfectly learns its opponent's strength after receiving a signal realization and the conflict stage is changed to a game of full information.

Proposition 2 (Full Information Case). *If the mediator provides full information to the conflicting parties, the following equilibrium always exists: A strong conflicting party fights upon receiving a low signal realization and does not fight upon receiving a high signal realization. A weak conflicting party does not fight. Call this the Mediated Equilibrium.*

⁸ If both conflicting parties have deterrent militarization levels, such a pooling equilibrium always exists and is moreover the payoff-dominant equilibrium.

The Mediated Equilibrium is the unique payoff undominated equilibrium in the full information case.

A formal proof is omitted. In the full information case, a strong conflicting party conditions its action on the received signal realization. The ex-ante war probability is reduced compared to the no information case, as war between two strong conflicting parties is avoided. The mediator's ex-ante expected payoff for choosing the profile of fully informative signals π^{FI} is given by

$$E[v|\pi^{FI}] = -[q_1(1 - q_2) + (1 - q_1)q_2]. \tag{5}$$

4.4 | Optimal information

The question emerges whether the mediator can choose a signal profile π to improve upon the full information case and if he can do so, which signal profile minimizes the ex-ante war probability. Observation 1 and Observation 2, stated in the beginning of the analysis, help to simplify the mediator's problem: The mediator aims at inducing strong conflicting parties to choose not to fight as often as possible. The following proposition characterizes the optimal signal profile to do so together with the equilibrium it induces in the conflict stage.

Proposition 3 (Optimal Information Case). *To minimize the ex-ante war probability, the mediator provides each conflicting party strategically with information about its opponent's strength. Signals of the optimal signal profile $\pi^* = \{\pi_1^*, \pi_2^*\}$ take the following form:⁹*

$$\pi_i^*(h|H) = 1 \quad \text{and} \quad \pi_i^*(h|L) = \frac{q_j}{1 - q_j} \frac{1/2 - \theta/2}{\theta - 1/2} \quad \text{if} \quad q_j < \bar{q}(\theta) \tag{6}$$

and

$$\pi_i^*(h|H) = 1 \quad \text{and} \quad \pi_i^*(h|L) = 1 \quad \text{if} \quad q_j \geq \bar{q}(\theta) \tag{7}$$

for all i . Given that the mediator chooses the optimal signal profile, the Mediated Equilibrium is the unique payoff undominated equilibrium in the conflict stage: A strong conflicting party fights upon receiving a low signal realization and does not fight upon receiving a high signal realization. A weak conflicting party does not fight.

Proof. It follows from Observation 2 that no equilibrium can be induced in which war never occurs and it follows from Observation 1 that it comes at no cost to induce weak conflicting parties to never fight. Hence, it is optimal to induce the Mediated Equilibrium and to minimize the probability with which strong conflicting parties fight under the Mediated Equilibrium. I derive (a) conditions under which a signal profile π induces the Mediated Equilibrium in the conflict stage, and derive (b) the signal profile which minimizes the ex-ante war probability in the Mediated Equilibrium subject to these conditions being satisfied.

(a) In the Mediated Equilibrium, both conflicting parties stick to the separating strategy

$$\phi^*(t_i, s_i) = \begin{cases} f & \text{if } t_i = H \quad \text{and} \quad s_i = l \\ nf & \text{otherwise.} \end{cases} \tag{8}$$

⁹To see that $\theta \leq 1/2$ holds for any $\theta \leq 1/2$, it helps to recall that $\theta \leq 1/2$ increases with $\theta \leq 1/2$ and $\theta \leq 1/2$.

It is sufficient to concentrate on the behavior of a strong conflicting party. Given π and $\phi_j = \phi^*$, the expected payoff of a strong conflicting party i upon receiving a high signal realization is given by

$$E_i[u_i(a_i, a_j, H, t_j)|s_i = h] = \begin{cases} 1/2 + (1 - q_j^h)(\theta - 1/2) - q_j^h(1/2 - \theta/2) & \text{if } a_i = f \\ 1/2 - q_j^h \pi_j(l|H)(1/2 - \theta/2) & \text{if } a_i = nf. \end{cases}$$

If the posterior belief q_j^h satisfies

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h|H)} \leq \frac{q_j^h}{1 - q_j^h}, \quad (C1_i)$$

$E_i[u_i(nf, a_j, H, t_j)|s_i = h] \geq E_i[u_i(f, a_j, H, t_j)|s_i = h]$ and it is optimal for a strong conflicting party to choose $a_i = nf$. Further, the expected payoff of a strong conflicting party upon receiving a low signal realization is given by

$$E_i[u_i(a_i, a_j, H, t_j)|s_i = l] = \begin{cases} 1/2 + (1 - q_j^l)(\theta - 1/2) - q_j^l(1/2 - \theta/2) & \text{if } a_i = f \\ 1/2 - q_j^l \pi_j(l|H)(1/2 - \theta/2) & \text{if } a_i = nf. \end{cases}$$

If the posterior belief q_j^l satisfies

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h|H)} > \frac{q_j^l}{1 - q_j^l}, \quad (C2_i)$$

$E_i[u_i(f, a_j, H, t_j)|s_i = l] > E_i[u_i(nf, a_j, H, t_j)|s_i = l]$ and it is optimal for a strong conflicting party to choose $a_i = f$.

Thus, if the posterior beliefs q_j^h and q_j^l satisfy (C1_i) and (C2_i), $\phi_i = \phi^*$ is the optimal reply to $\phi_j = \phi^*$. A signal profile π induces the Mediated Equilibrium if signal π_i induces posterior beliefs q_j^h and q_j^l satisfying (C1_i) and (C2_i) for all i .

Substituting for q_j^h and q_j^l , the constraints (C1_i) and (C2_i) can be restated as

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h|H)} \leq \frac{q_j}{1 - q_j} \frac{\pi_i(h|H)}{\pi_i(h|L)} \quad (C1_i)$$

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h|H)} > \frac{q_j}{1 - q_j} \frac{\pi_i(l|H)}{\pi_i(l|L)}. \quad (C2_i)$$

- (b) The mediator chooses the signal profile π to maximize the expected payoff $E[v|\pi]$ under the Mediated Equilibrium, subject to (C1_i) and (C2_i) being satisfied for all i . The choice of a signal profile π corresponds to that of the four probabilities $\{\{\pi_i(h|t_j)\}_{t_j=H,L}\}_{i=1,2}$. The mediator's objective function can be equivalently formulated as the ex-ante probability of no war:

$$q_1 q_2 \pi_1(h|H) \pi_2(h|H) + q_1(1 - q_2) \pi_1(h|L) + (1 - q_1) q_2 \pi_2(h|L) + (1 - q_1)(1 - q_2).$$

The first term corresponds to the case that both conflicting parties are strong and receive a high signal realization each. The second and the third term each corresponds to the case that exactly one conflicting party is strong and receives a high signal realization. The fourth term corresponds to the case that both conflicting parties are weak. Dropping the fourth term, which is constant given q_1 and q_2 , allows to write the mediator's maximization problem as

$$\begin{aligned} \text{Max}_{\{\pi_i(h|t_j)\}_{i,j=H,L}\}_{i=1,2}} & \{q_1 q_2 \pi_1(h|H) \pi_2(h|H) + q_1(1 - q_2) \pi_1(h|L) \\ & + (1 - q_1) q_2 \pi_2(h|L)\} \\ \text{s.t.} & \frac{\theta - 1/2}{(1/2 - \theta/2) \pi_j(h|H)} \leq \frac{q_j}{1 - q_j} \frac{\pi_i(h|H)}{\pi_i(h|L)} \end{aligned} \tag{C1_i}$$

$$\begin{aligned} & \frac{\theta - 1/2}{(1/2 - \theta/2) \pi_j(h|H)} > \frac{q_j}{1 - q_j} \frac{\pi_i(l|H)}{\pi_i(l|L)} \tag{C2_i} \\ & \{\pi_i(h|t_j)\}_{t_j=H,L} \in [0, 1] \\ & \text{for all } i. \end{aligned}$$

The objective function increases in $\pi_i(h|H)$ and $\pi_i(h|L)$ for all i . Assume $\pi_i(h|H) = 1$ for all i . The constraints reduce to

$$\frac{\theta - 1/2}{(1/2 - \theta/2)} \leq \frac{q_j}{1 - q_j} \frac{1}{\pi_i(h|L)} \tag{C1'_i}$$

$$\frac{\theta - 1/2}{(1/2 - \theta/2)} > 0. \tag{C2'_i}$$

(C2'_i) holds trivially. If $q_j \geq \bar{q}(\theta)$, (C1'_i) holds for any $\pi_i(h|L)$ and it is optimal to set $\pi_i(h|L) = 1$. If $q_j < \bar{q}(\theta)$, it is optimal to set $\pi_i(h|L)$ such that (C1'_i) binds:

$$\pi_i(h|L) = \frac{q_j}{1 - q_j} \frac{1/2 - \theta/2}{\theta - 1/2}.$$

Now consider whether $\pi_i(h|H) = 1$ is optimal. $\pi_i(h|H) = 1$ is optimal unless choosing $\pi_i(h|H) < 1$ allows to increase $\pi_i(h|L)$ or $\pi_j(h|L)$. However, decreasing $\pi_i(h|H)$ decreases the right hand side of (C1_i) and increases the left hand side of (C1_j). This implies that if $\pi_i(h|H)$ was reduced, $\pi_i(h|L)$ and $\pi_j(h|L)$ would need to decrease to satisfy (C1_i) and (C1_j). Thus, $\pi_i(h|H) = 1$ for all i is optimal. To conclude, optimal signals are given by (6) and (7). □

As in the full information case, a strong conflicting party conditions its action on the received signal realization in the optimal information case. Different to the full information case, a high signal realization is at most indicative of a strong opponent. The mediator improves

upon the full information case by choosing a signal which sometimes has a high realization although the opponent is weak.

The optimal signal takes one of two forms, depending on the opponent's militarization level. If the opponent has a deterrent militarization level, the signal does not need to provide any information to convince a strong conflicting party not to fight. In this case, the uninformative signal given by (7) is optimal such that the conflicting party always receives a high signal realization and never fights.

If the opponent does not have a deterrent militarization level, the signal does need to provide information to convince a strong conflicting party not to fight. In this case, the signal given by (6) is optimal. The signal has always a high realization if the opponent is strong and sometimes does so if the opponent is weak. To maximize the occurrence of high signal realizations, the probability with which the signal has a high realization if the opponent is weak is chosen such that a strong conflicting party is indifferent between fighting and not fighting upon receiving a high signal realization.

The more incentive a strong conflicting party has to fight without receiving any further information about the opponent's strength, the more information is necessary to convince it not to fight. For this reason, the probability $\pi_i(h|L)$ decreases from 1 to 0 and the signal π_i becomes more precise as the opponent's militarization level decreases from $\bar{q}(\theta)$ to 0. When the opponent's militarization level q_j is close to the deterrent threshold $\bar{q}(\theta)$, a strong conflicting party has little incentive to fight. Little information is therefore necessary to convince it not to fight and an imprecise signal with a high probability $\pi_i(h|L)$ is optimal. As q_j decreases, a strong conflicting party has more incentive to fight and more information is necessary to convince it not to fight. The probability $\pi_i(h|L)$ decreases such that the signal becomes more precise.

Parallel reasoning applies regarding the size of the reduced pie θ . The probability $\pi_i(h|L)$ decreases from 1 to 0 and the signal π_i becomes more precise as the size of the reduced pie increases from 1/2 to 1. For θ close to 1/2, a strong conflicting party i has little incentive to fight and little information is necessary to convince it not to fight. As θ increases, a strong conflicting party has more incentive to fight and a more precise signal is necessary to convince it not to fight.

Given the mediator chooses the optimal signal profile and war occurs in the Mediated Equilibrium if exactly one conflicting party is strong and receives a low signal realization. The mediator's expected payoff for choosing the optimal signal profile π^* is given by

$$E[v|\pi^*] = \begin{cases} -[q_1(1 - q_2)\pi_1^*(l|L) + q_2(1 - q_1)\pi_2^*(l|L)] & \text{if } \forall i: q_i < \bar{q}(\theta) \\ -[q_1(1 - q_2)\pi_1^*(l|L)] & \text{if } q_2 < \bar{q}(\theta) \leq q_1 \\ -[q_2(1 - q_1)\pi_2^*(l|L)] & \text{if } q_1 < \bar{q}(\theta) \leq q_2. \end{cases} \quad (9)$$

5 | EFFECT OF MEDIATION

5.1 | War probability

I discuss how optimal information affects the ex-ante war probability (Sections 5.1 and 5.2) and the conflicting parties' ex-ante expected payoff (Section 5.3). For ease of exposition, I restrict the discussion by assuming that no conflicting party has a deterrent militarization level.

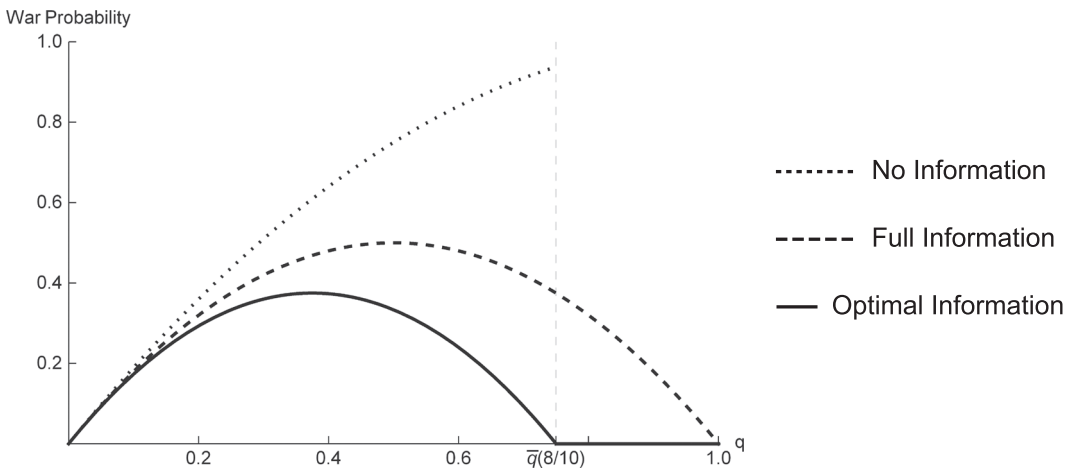


FIGURE 1 Here, the ex-ante war probability for no information, full information, and optimal information is depicted as a function of a symmetric militarization level q for $\theta = 0.8$. In the no information case, the ex-ante war probability strictly increases in the militarization level q . In the optimal information case, the ex-ante war probability increases for low militarization levels and decreases for high militarization levels. For $q \geq \bar{q}(0.8)$, the ex-ante war probability is zero in the no information case and in the optimal information case

Optimal information reduces the ex-ante war probability compared with the no information case and the full information case: No war occurs between two strong conflicting parties and the probability of war between a strong and a weak conflicting party is strictly less than one. Figure 1 depicts the ex-ante war probability in the no information case, the full information case and in the optimal information case for $\theta = 0.8$ assuming a symmetric militarization level q . In the no information case, the ex-ante war probability strictly increases in the militarization level q up to the deterrent threshold $\bar{q}(0.8)$. As q increases, so does the probability of at least one conflicting party being strong which fights in the no information case. Once the militarization levels reach $\bar{q}(0.8)$, war can always be avoided in the no information case. In the full information case, war occurs if exactly one conflicting party is strong. The probability of this being the case is concave in q , reaching its maximum at $q = 1/2$.¹⁰

Optimal information improves further upon full information. In the optimal information case, war occurs if exactly one conflicting party is strong, but does so with a probability strictly less than one. Figure 1 shows that the difference between the ex-ante war probability in the full information case and in the optimal information case increases in q . As the opponent's militarization level increases, less information is necessary to convince a strong conflicting party not to fight. The optimal signal is less precise, implying a higher probability of high signal realizations. When q is close to $\bar{q}(0.8)$, the probability with which the optimal signals have a high realization is close to one and war can almost always be avoided. Once militarization levels reach $\bar{q}(\theta)$, the mediator can use uninformative signals, such that both conflicting parties always receive high signal realizations and never fight.

¹⁰It can be seen that full information provokes war when there is no war in the no information case if both conflicting parties have deterrent militarization levels. However, this case is not the focus of this paper and was ruled out by Assumption 1.

5.2 | Mediation success

Let $\Delta E[v|\pi^*]$ be defined as the increase in the mediator's expected payoff over the no information case achieved by choosing the optimal signal profile π^* :

$$\Delta E[v|\pi^*] \equiv E[v|\pi^*] - E[v|\pi^{NI}].$$

I refer to $\Delta E[v|\pi^*]$ as *mediation success*, meaning that the higher $\Delta E[v|\pi^*]$, the more successful mediation is. Assuming that for all i , $q_i < \bar{q}(\theta)$,

$$\Delta E[v|\pi^*] = q_1 q_2 \frac{1/2}{\theta - 1/2}. \quad (10)$$

Mediation success increases in militarization levels q_1 and q_2 and decreases in the size of the reduced pie θ . The relationship between militarization levels and mediation success follows directly from the discussion in Section 5.1. The negative relationship between the size of the reduced pie θ and mediation success stems from the effect an increase in θ has on the optimal signals. As θ increases, more information about the opponent's strength is needed to convince a strong conflicting party not to fight. The optimal signals get more precise and the probability of high signal realizations decreases, implying a higher probability of war between a strong and a weak conflicting party. Thus, mediation success increases faster in the militarization level q the smaller the reduced pie θ is, as can be seen in Figure 2. Figure 2 also shows that the parameter range of q , for which mediation is relevant is the smaller the smaller θ is.

5.3 | Conflicting parties' benefit

Each conflicting party benefits from optimal information. Let $\Delta E[u_i|\pi^*]$ be defined as the increase in conflicting party i 's ex-ante expected payoff in the optimal information case compared to the no information case:

$$\Delta E[u_i|\pi^*] \equiv E[u_i|\pi^*] - E[u_i|\pi^{NI}].$$

I refer to $\Delta E[u_i|\pi^*]$ as conflicting party i 's *mediation benefit*. Assuming that for all i , $q_i < \bar{q}(\theta)$,

$$\Delta E[u_i|\pi^*] = q_j q_i \frac{1/2 - \theta/2}{\theta - 1/2} 1/2. \quad (11)$$

The mediation benefit of conflicting party i increases in i 's militarization level q_i , in the opponent's militarization level q_j and decreases in the size of the reduced pie θ . Conflicting party i 's own militarization level has two countervailing effects on i 's mediation benefit. The probability with which i is weak and benefits from optimal information decreases in q_i . This effect is negative. On the other hand, the extent to which a weak conflicting party i benefits from optimal information increases in q_i , as the higher q_i , the less precise the signal π_j^* is. The second effect dominates, implying an overall positive relationship between i 's militarization level q_i and i 's mediation benefit.

The relationship between the opponent's militarization level and i 's mediation benefit is more straightforward. As q_j increases, so does the probability with which i faces a strong opponent and can, therefore, benefit from optimal information.

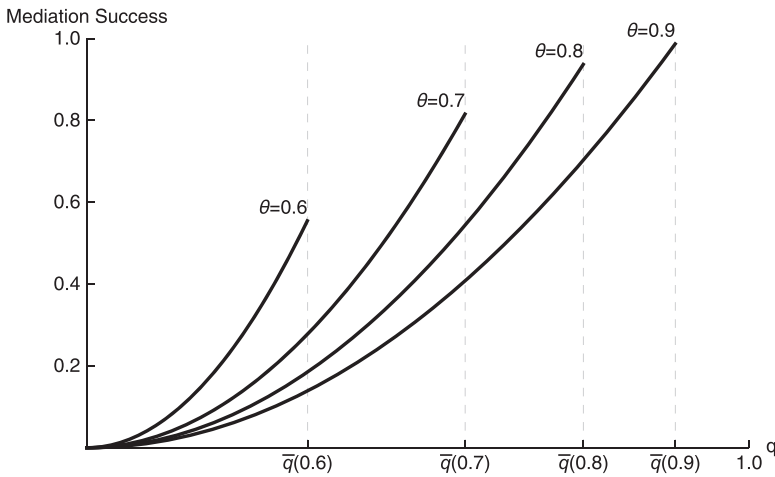


FIGURE 2 Here, the mediation success is depicted as a function of the militarization level q for different values of θ , assuming a symmetric militarization level q . The mediation success increases with q and is the steeper, the costlier war is

Lastly, the relationship between θ and the mediation benefit is negative, because the larger the reduced pie θ is, the more precise optimal signals are and the less often the opponent receives high signal realizations.

Decomposing the mediation benefit by conflicting party i 's strength reveals that the whole benefit is taken up by weak conflicting parties:

$$\Delta E [u_i | \pi^*] = (1 - q_i) \left[q_j \frac{q_i}{1 - q_i} \frac{1/2 - \theta/2}{\theta - 1/2} \frac{1}{\pi_j^*(h|L)} \frac{1}{2} \right] + q_i 0. \tag{12}$$

Strong conflicting parties are as good off in the optimal information case as in the no information case. This is intuitive: If a strong conflicting party benefited from optimal information, the ex-ante war probability could be reduced further.

6 | CONCLUSION

Uncertainty and informational asymmetries are important factors to understand conflict and conflict resolution. In this paper, I use Bayesian persuasion (see Kamenica & Gentzkow, 2011) to show how a mediator can reduce the ex-ante war probability by providing conflicting parties strategically with information about the respective opponent's strength. When a conflicting party receives information indicating that it faces a strong opponent, it refrains from fighting. On the downside, if a conflicting party receives information indicating a weak opponent, it fights. I derive under which conditions the provided information has this effect and how the mediator uses this effect optimally. The effective use of information to prevent war rests on the assumption that the mediator obtains the information independently and commits on sharing this information with the conflicting parties. Mediation in

this form is the more successful, the higher the ex-ante war probability is absent mediation and the costlier war is.

ACKNOWLEDGMENTS

I thank Laura Arnemann, Jana Cahlíková, Florian Englmaier, Bharat Goel, Kai A. Konrad, Mariana Lopes da Fonseca, Jonas Send, Marco Serena, and all my remaining colleagues at the Max Planck Institute for Tax Law and Public Finance for helpful comments and feedback.

ORCID

Raphaela Hennigs  <http://orcid.org/0000-0003-1924-2432>

REFERENCES

- Agastya, M., Menezes, F., & Sengupta, K. (2007). Cheap talk, efficiency and egalitarian cost sharing in joint projects. *Games and Economic Behavior*, 60(1), 1–19.
- Alonso, R., & Câmara, O. (2016). Persuading voters. *American Economic Review*, 106(11), 3590–3605. <https://doi.org/10.1257/aer.20140737>
- Balzer, B., & Schneider, J. (2019). *Belief management and optimal arbitration*, (Discussion Paper). University of Technology Sydney, Sydney.
- Barbieri, S., & Malueg, D. A. (2019). On the voluntary provision of “weakest-link” public goods: The case of private information. *Journal of Public Economic Theory*, 21(5), 866–894.
- Bergemann, D., & Morris, S. (2019). Information design: A unified perspective. *Journal of Economic Literature*, 57(1), 44–95.
- Bester, H., & Warneryd, K. (2006). Conflict and the social contract. *Scandinavian Journal of Economics*, 108(2), 231–249.
- Brecher, M., Wilkenfeld, J., Beardsley, K., James, P., & Quinn, D. (2017). *International crisis behavior data codebook, Version 12*. <http://sites.duke.edu/icbdata/data-collections/>
- Brecher, M., & Wilkenfeld, J. (1997). *A study of crisis*. University of Michigan Press.
- Chan, J., Gupta, S., Li, F., & Wang, Y. (2019). Pivotal persuasion. *Journal of Economic Theory*, 180, 178–202.
- Chen, Z. (2019). *Information disclosure in contests: Private versus public signals*, (Discussion Paper). Harbin Institute of Technology, Shenzhen.
- Colef, J., & Garcia, D. (2017). Information provision in procurement auctions. *Journal of Public Economic Theory*, 19(2), 426–444.
- Feng, X., & Lu, J. (2016). The optimal disclosure policy in contests with stochastic entry: A bayesian persuasion perspective. *Economics Letters*, 147, 103–107.
- Fey, M., & Ramsay, K. W. (2009). Mechanism design goes to war: Peaceful outcomes with interdependent and correlated types. *Review of Economic Design*, 13(3), 233–250.
- Fey, M., & Ramsay, K. W. (2010). When is shuttle diplomacy worth the commute? Information sharing through mediation. *World Politics*, 62(04), 529–560.
- Fey, M., & Ramsay, K. W. (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: Volume 2: War and diplomacy* (rev. ed.). Routledge.
- Garrido, D., Espínola-Arredondo, A., & Munoz-Garcia, F. (2020). Can mandatory certification promote greenwashing? A signaling approach. *Journal of Public Economic Theory*, 22(6), 1801–1851.
- Hörner, J., Morelli, M., & Squintani, F. (2015). Mediation and peace. *The Review of Economic Studies*, 82(4), 1483–1501.
- Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics*, 11(1).
- Kamenica, E., & Gentzkow, M. (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.

- Laipson, E., Kurtzer, D., & Moore, J. L. (1995). *Intelligence and the middle east: What do we need to know?* The Washington Institute for Near East Policy.
- Mathevet, L., Perego, J., & Taneva, I. (2020). On information design in games. *Journal of Political Economy*, 128(4), 1370–1404.
- Ottaviani, M., & Sørensen, P. N. (2006). Reputational cheap talk. *The RAND Journal of Economics*, 37(1), 155–175.
- Palfrey, T. R., & Rosenthal, H. (1991). Testing for effects of cheap talk in a public goods game with private information. *Games and Economic Behavior*, 3(2), 183–220.
- Rauchhaus, R. W. (2006). Asymmetric information, mediation, and conflict management. *World Politics*, 58(2), 207–241.
- Smith, A., & Stam, A. (2003). Mediation and peacekeeping in a random walk model of civil and interstate war. *International Studies Review*, 5(4), 115–135.
- Taneva, I. (2019). Information design. *American Economic Journal: Microeconomics*, 11(4), 151–185.
- Zhang, J., & Zhou, J. (2016). Information disclosure in contests: A Bayesian persuasion approach. *The Economic Journal*, 126(597), 2197–2217.
- Zimper, A., & Molefinyane, M. (2020). Bargaining over loan contracts with signaling. *Journal of Public Economic Theory*. <https://doi.org/10.1111/jpet.12480>

How to cite this article: Hennigs, R. (2021). Conflict prevention by Bayesian persuasion. *J Public Econ Theory*. 23, 710–731. <https://doi.org/10.1111/jpet.12511>

APPENDIX A: CORRELATED SIGNAL

I generalize the analysis and allow the mediator to use correlated signals. Let $t = (t_1, t_2)$ denote the profile of conflicting parties' strengths and let \mathbf{T} denote the set of strength profiles:

$$\mathbf{T} = \left\{ \begin{pmatrix} H \\ H \end{pmatrix}, \begin{pmatrix} H \\ L \end{pmatrix}, \begin{pmatrix} L \\ H \end{pmatrix}, \begin{pmatrix} L \\ L \end{pmatrix} \right\}$$

Let $s = (s_1, s_2)$ denote the profile of signal realizations, with s_i being the signal realization observed by i and let \mathbf{S} denote the set of signal realization profiles:

$$\mathbf{S} = \left\{ \begin{pmatrix} h \\ h \end{pmatrix}, \begin{pmatrix} h \\ l \end{pmatrix}, \begin{pmatrix} l \\ h \end{pmatrix}, \begin{pmatrix} l \\ l \end{pmatrix} \right\}$$

The signal $\mathbf{\Pi}$ consists of \mathbf{S} and a set of conditional probability distributions $\{\pi(\cdot|t)\}_{t \in \mathbf{T}}$ over \mathbf{S} . To simplify the notation, I write

$$\pi \left(\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \middle| \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \right) \equiv \pi \left(\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \middle| \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \right) = \text{Prob} \left(s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \middle| t = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \right).$$

The mediator chooses the signal $\mathbf{\Pi}$ publicly. Each conflicting party i observes signal realization s_i privately.

Upon observing signal realization s_i , conflicting party i forms the posterior belief $q_j^{s_i, t_i}$ about its opponent's strength. As opposed to the baseline model, the posterior belief $q_j^{s_i, t_i}$ depends on the observed signal realization and on conflicting party i 's strength. For example, conflicting party 1 forms the posterior belief $q_2^{h, H}$ after observing $s_1 = h$ and given $t_1 = H$:

$$q_2^{h,H} = \text{Prob}(t_2 = H | s_1 = h, t_1 = H) \\ = \frac{\left[\pi \left(\begin{matrix} h & H \\ h & H \end{matrix} \right) + \pi \left(\begin{matrix} h & H \\ l & H \end{matrix} \right) \right] q_2}{\left[\pi \left(\begin{matrix} h & H \\ h & H \end{matrix} \right) + \pi \left(\begin{matrix} h & H \\ l & H \end{matrix} \right) \right] q_2 + \left[\pi \left(\begin{matrix} h & H \\ h & L \end{matrix} \right) + \pi \left(\begin{matrix} h & H \\ l & L \end{matrix} \right) \right] (1 - q_2)}.$$

The numerator states the joint probability of conflicting party 1 receiving a high signal realization and conflicting party 2 being strong, whereas the denominator states the total probability of 1 receiving a high signal realization. Any signal induces posterior beliefs $q_j^{l,t_i} \leq q_j \leq q_j^{h,t_i}$ for $t_i = H, L$, as the posterior belief $q_j^{s_i,t_i}$ needs to equal the prior belief q_j in expectation. The above described posterior beliefs allow to analyze the mediator's problem. The following proposition summarizes this analysis:

Proposition A1 (Correlated Signals). *The mediator does not benefit from using correlated signals instead of a profile of two independent signals. The optimal signal $\mathbf{\Pi}^*$ is equivalent to the signal profile $\mathbf{\pi}^*$ described in Proposition 3.*

Proof. To prove the Proposition, I derive a signal $\mathbf{\Pi}^*$ which solves the mediator's maximization problem I show that $\mathbf{\Pi}^*$ is equivalent to the signal profile $\mathbf{\pi}^*$ described in Proposition 3. The derivation of the signal $\mathbf{\Pi}^*$ is parallel to that of $\mathbf{\pi}^*$ in Section 4.4. I derive (a) conditions under which a signal $\mathbf{\Pi}$ induces the Mediated Equilibrium in the conflict stage, and derive (b) a signal which minimizes the ex-ante war probability in the Mediated Equilibrium subject to these conditions being satisfied.

(a) In the Mediated Equilibrium, both conflicting parties stick to the separating strategy

$$\phi^*(t_i, s_i) = \begin{cases} f & \text{if } t_i = H \text{ and } s_i = l \\ nf & \text{otherwise.} \end{cases} \quad (\text{A1})$$

It is sufficient to concentrate on the behavior of a strong conflicting party. Given $\mathbf{\Pi}$ and $\phi_2 = \phi^*$, the expected payoff of conflicting party 1 upon receiving a high signal realization is given by

$$E_1[u_1(a_1, a_2, H, t_2) | s_1 = h] \\ = \begin{cases} (1 - q_2^{h,H})\theta + q_2^{h,H}\theta/2 & \text{if } a_1 = f \\ (1 - q_2^{h,H})1/2 + q_2^{h,H} \left[\pi \left(\begin{matrix} h & H \\ h & H \end{matrix} \right) 1/2 + \pi \left(\begin{matrix} h & H \\ l & H \end{matrix} \right) \theta/2 \right] & \text{if } a_1 = nf. \end{cases}$$

If the posterior belief $q_2^{h,H}$ satisfies

$$\frac{q_2^{h,H}}{1 - q_2^{h,H}} \left[\pi \left(\begin{matrix} h & H \\ h & H \end{matrix} \right) 1/2 + \pi \left(\begin{matrix} h & H \\ l & H \end{matrix} \right) \theta/2 - \theta/2 \right] \geq \theta - 1/2, \quad (\text{CC1})$$

$E_1[u_1(nf, a_2, H, t_2) | s_1 = h] \geq E_1[u_1(f, a_2, H, t_2) | s_1 = h]$ and it is optimal for 1 to choose $a_1 = nf$. Further, the expected payoff of conflicting party 1 upon receiving a low signal realization is given by

$$\begin{aligned}
 & E_1[u_1(a_1, a_2, H, t_2)|s_1 = l] \\
 &= \begin{cases} (1 - q_2^{l,H})\theta + q_2^{l,H}\theta/2 & \text{if } a_1 = f \\ (1 - q_2^{l,H})1/2 + q_2^{l,H} \left[\pi\left(\frac{l}{h} \middle| \frac{H}{H}\right)1/2 + \pi\left(\frac{l}{l} \middle| \frac{H}{H}\right)\theta/2 \right] & \text{if } a_1 = nf. \end{cases}
 \end{aligned}$$

If the posterior belief $q_2^{l,H}$ satisfies

$$\begin{aligned}
 & \frac{q_2^{l,H}}{1 - q_2^{l,H}} \left[\pi\left(\frac{l}{h} \middle| \frac{H}{H}\right)1/2 + \pi\left(\frac{l}{l} \middle| \frac{H}{H}\right)\theta/2 - \theta/2 \right] \\
 & < \theta - 1/2, \quad (CC2_1)
 \end{aligned}$$

$E_1[u_1(f, a_2, H, t_2)|s_1 = l] > E_1[u_1(nf, a_2, H, t_2)|s_1 = l]$ and it is optimal for conflicting party 1 to choose $a_1 = f$.

Thus, if the posterior beliefs $q_2^{h,H}$ and $q_2^{l,H}$ satisfy (CC1₁) and (CC2₁), $\phi_1 = \phi^*$ is the optimal reply to $\phi_2 = \phi^*$. Equivalent conditions (CC1₂) and (CC2₂) can be derived for conflicting party 2. Hence, a signal $\mathbf{\Pi}$ induces the Mediated Equilibrium if it induces posterior beliefs satisfying (CC1_{*i*}) and (CC2_{*i*}) for all *i*. These four inequalities restrict only the posterior beliefs conflicting parties form when they are strong, whereas no restrictions are posed on those posterior beliefs formed when conflicting parties are weak.

Substituting for $q_2^{h,H}$ and $q_2^{l,H}$, the constraints (CC1₁) and (CC2₁) can be restated as

$$\frac{\pi\left(\frac{h}{h} \middle| \frac{H}{H}\right) + \pi\left(\frac{h}{l} \middle| \frac{H}{H}\right)}{\pi\left(\frac{h}{h} \middle| \frac{H}{L}\right) + \pi\left(\frac{h}{l} \middle| \frac{H}{L}\right)} \frac{q_2}{1 - q_2} \left[\pi\left(\frac{h}{h} \middle| \frac{H}{H}\right)1/2 + \pi\left(\frac{h}{l} \middle| \frac{H}{H}\right)\theta/2 - \theta/2 \right] \geq \theta - 1/2, \quad (CC1_1)$$

$$\frac{\pi\left(\frac{l}{h} \middle| \frac{H}{H}\right) + \pi\left(\frac{l}{l} \middle| \frac{H}{H}\right)}{\pi\left(\frac{l}{h} \middle| \frac{H}{L}\right) + \pi\left(\frac{l}{l} \middle| \frac{H}{L}\right)} \frac{q_2}{1 - q_2} \left[\pi\left(\frac{l}{h} \middle| \frac{H}{H}\right)1/2 + \pi\left(\frac{l}{l} \middle| \frac{H}{H}\right)\theta/2 - \theta/2 \right] < \theta - 1/2, \quad (CC2_1)$$

(b) As in the baseline case, the mediator's objective function can be equivalently formulated as the ex-ante probability of no war:

$$\begin{aligned}
 & q_1 q_2 \pi\left(\frac{h}{h} \middle| \frac{H}{H}\right) + q_1(1 - q_2) \left[\pi\left(\frac{h}{h} \middle| \frac{H}{L}\right) + \pi\left(\frac{h}{l} \middle| \frac{H}{L}\right) \right] \\
 & + (1 - q_1) q_2 \left[\pi\left(\frac{h}{h} \middle| \frac{L}{H}\right) + \pi\left(\frac{l}{h} \middle| \frac{L}{H}\right) \right] + (1 - q_1)(1 - q_2).
 \end{aligned}$$

The first term corresponds to the case that both conflicting parties are strong and receive a high signal realization each. The second and the third term each corresponds to the case that exactly one conflicting party is strong and receives a high signal realization. (It is irrelevant which signal realization the weak conflicting party receives in these cases.) The fourth term corresponds to the case that both conflicting parties are weak. Dropping

the fourth term, which is constant given q_1 and q_2 , allows to write the mediator's maximization problem as

$$\begin{aligned} \text{Max}_{\{\pi(s|t)\}_{t \in T, s \in S}} & \left\{ q_1 q_2 \pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right) + q_1 (1 - q_2) \left[\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ L \end{matrix} \right) + \pi \left(\begin{matrix} h \\ l \end{matrix} \middle| \begin{matrix} H \\ L \end{matrix} \right) \right] \right. \\ & \left. + (1 - q_1) q_2 \left[\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} L \\ H \end{matrix} \right) + \pi \left(\begin{matrix} l \\ h \end{matrix} \middle| \begin{matrix} L \\ H \end{matrix} \right) \right] \right\} \\ \text{s.t.} & \frac{\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right) + \pi \left(\begin{matrix} h \\ l \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)}{\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ L \end{matrix} \right) + \pi \left(\begin{matrix} h \\ l \end{matrix} \middle| \begin{matrix} H \\ L \end{matrix} \right)} \frac{q_2}{1 - q_2} \left[\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)^{1/2} + \pi \left(\begin{matrix} h \\ l \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)^{\theta/2 - \theta/2} \right] \geq \theta - 1/2, \quad (\text{CC}_1) \end{aligned}$$

$$\frac{\pi \left(\begin{matrix} l \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right) + \pi \left(\begin{matrix} l \\ l \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)}{\pi \left(\begin{matrix} l \\ h \end{matrix} \middle| \begin{matrix} H \\ L \end{matrix} \right) + \pi \left(\begin{matrix} l \\ l \end{matrix} \middle| \begin{matrix} H \\ L \end{matrix} \right)} \frac{q_2}{1 - q_2} \left[\pi \left(\begin{matrix} l \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)^{1/2} + \pi \left(\begin{matrix} l \\ l \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)^{\theta/2 - \theta/2} \right] < \theta - 1/2, \quad (\text{CC}_{2_1})$$

$$\frac{\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right) + \pi \left(\begin{matrix} l \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)}{\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} L \\ H \end{matrix} \right) + \pi \left(\begin{matrix} l \\ h \end{matrix} \middle| \begin{matrix} L \\ H \end{matrix} \right)} \frac{q_1}{1 - q_1} \left[\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)^{1/2} + \pi \left(\begin{matrix} l \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)^{\theta/2 - \theta/2} \right] \geq \theta - 1/2, \quad (\text{CC}_{1_2})$$

$$\frac{\pi \left(\begin{matrix} h \\ l \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right) + \pi \left(\begin{matrix} l \\ l \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)}{\pi \left(\begin{matrix} h \\ l \end{matrix} \middle| \begin{matrix} L \\ H \end{matrix} \right) + \pi \left(\begin{matrix} l \\ l \end{matrix} \middle| \begin{matrix} L \\ H \end{matrix} \right)} \frac{q_1}{1 - q_1} \left[\pi \left(\begin{matrix} h \\ l \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)^{1/2} + \pi \left(\begin{matrix} l \\ l \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)^{\theta/2 - \theta/2} \right] < \theta - 1/2, \quad (\text{CC}_{2_2})$$

$$\begin{aligned} \sum_{s \in S} \pi(s|t) &= 1 \quad \forall t \in T, \\ \pi(s|t) &\in [0, 1] \quad \forall t \in T, s \in S \end{aligned}$$

The objective function increases in $\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right)$ and we can set $\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ H \end{matrix} \right) = 1$ without tightening any of the constraints. Further, we can simplify the maximization problem by setting $\pi \left(\begin{matrix} h \\ l \end{matrix} \middle| \begin{matrix} H \\ L \end{matrix} \right) = \pi \left(\begin{matrix} l \\ h \end{matrix} \middle| \begin{matrix} L \\ H \end{matrix} \right) = 0$ without any loss. The constraints reduce to

$$\frac{1}{\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} H \\ L \end{matrix} \right)} \frac{q_2}{1 - q_2} \left[1/2 - \theta/2 \right] \geq \theta - 1/2, \quad (\text{CC}'_{1_1})$$

$$0 < \theta - 1/2, \quad (\text{CC}'_{2_1})$$

$$\frac{1}{\pi \left(\begin{matrix} h \\ h \end{matrix} \middle| \begin{matrix} L \\ H \end{matrix} \right)} \frac{q_1}{1 - q_1} \left[1/2 - \theta/2 \right] \geq \theta - 1/2, \quad (\text{CC}'_{1_2})$$

$$0 < \theta - 1/2, \quad (\text{CC}'_{2_2})$$

(CC2₁) and (CC2₂) hold trivially. If $q_2 \geq \bar{q}(\theta)$, (CC1₁) holds for any $\pi \left(\begin{smallmatrix} h & H \\ h & L \end{smallmatrix} \right)$ and it is optimal to set $\pi \left(\begin{smallmatrix} h & H \\ h & L \end{smallmatrix} \right) = 1$. If $q_2 < \bar{q}(\theta)$, it is optimal to set $\pi \left(\begin{smallmatrix} h & H \\ h & L \end{smallmatrix} \right)$ such that (CC1₁) binds:

$$\pi \left(\begin{smallmatrix} h & H \\ h & L \end{smallmatrix} \right) = \frac{q_2}{1 - q_2} \frac{1/2 - \theta/2}{\theta - 1/2}.$$

Equivalent reasoning applies to $\pi \left(\begin{smallmatrix} h & L \\ h & H \end{smallmatrix} \right)$.

Hence, it is optimal to choose Π^* with

$$\begin{aligned} \pi^* \left(\begin{smallmatrix} h & H \\ h & H \end{smallmatrix} \right) &= 1 \\ \pi^* \left(\begin{smallmatrix} h & H \\ l & H \end{smallmatrix} \right) &= \pi^* \left(\begin{smallmatrix} l & H \\ h & H \end{smallmatrix} \right) = \pi^* \left(\begin{smallmatrix} l & H \\ l & H \end{smallmatrix} \right) = 0 \\ \pi^* \left(\begin{smallmatrix} h & H \\ h & L \end{smallmatrix} \right) &= \begin{cases} 1 & \text{if } q_2 \geq \bar{q}(\theta) \\ \frac{q_2}{1 - q_2} \frac{1/2 - \theta/2}{\theta - 1/2} & \text{if } q_2 < \bar{q}(\theta) \end{cases} \\ \pi^* \left(\begin{smallmatrix} h & L \\ h & H \end{smallmatrix} \right) &= \begin{cases} 1 & \text{if } q_1 \geq \bar{q}(\theta) \\ \frac{q_1}{1 - q_1} \frac{1/2 - \theta/2}{\theta - 1/2} & \text{if } q_1 < \bar{q}(\theta) \end{cases} \\ \pi^* \left(\begin{smallmatrix} l & L \\ l & L \end{smallmatrix} \right) &= 1 \\ \pi^* \left(\begin{smallmatrix} h & L \\ l & L \end{smallmatrix} \right) &= \pi^* \left(\begin{smallmatrix} l & L \\ h & L \end{smallmatrix} \right) = \pi^* \left(\begin{smallmatrix} h & L \\ h & L \end{smallmatrix} \right) = 0 \end{aligned}$$

The signal Π^* is equivalent to the signal profile described in Proposition 3. Thus, the mediator does not benefit from using correlated signals.¹¹ □

¹¹However, different to the baseline case, the signal $\theta \leq 1/2$ is not the unique solution to the mediator's maximization problem. For example, $\theta \leq 1/2$ can be chosen arbitrarily for $\theta \leq 1/2$.