Which Side Are You On? Bias, Credibility, and Mediation

Andrew Kydd Harvard University

Mediators are often thought to be more effective if they are unbiased or have no preferences over the issue in dispute. This article presents a game theoretic model of mediation drawing on the theory of "cheap talk" which highlights a contrary logic. Conflict arises in bargaining games because of uncertainty about the resolve of the parties. A mediator can reduce the likelihood of conflict by providing information on this score. For a mediator to be effective, however, the parties must believe that the mediator is telling the truth, especially if the mediator counsels one side to make a concession because their opponent has high resolve and will fight. An unbiased mediator who is simply interested in minimizing the probability of conflict will have a strong incentive to make such statements even if they are not true, hence the parties will not find the mediator credible. Only mediators who are effectively "on your side" will be believed if they counsel restraint.

Consider two cases of international mediation, one failure and one success. In 1982 after Argentina invaded the Falkland/Malvinas Islands, the United States attempted to mediate between Britain and Argentina. Secretary of State Alexander Haig attempted to persuade the Argentine Generals to withdraw:

> *In one final attempt to convince them, I sent Dick Walters to see Galtieri alone and tell him in crystal clear terms, in the Spanish language, that if there was no negotiated settlement, the British would fight and that the United States would support Britain. Galtieri listened and replied: "Why are you telling me this? The British won't fight." (Haig 1984, 280)*

As it turned out, Haig's statements to Argentina were truthful, yet the mediation effort failed.

In early 1999, the conflict between the Serbian government and the Kosovo Albanians spun out of control. In April, Serbia rejected a deal brokered at Rambouillet, France by NATO. Serbian forces swept into Kosovo and began a wholesale effort to displace the Albanian population, and NATO began a bombing campaign that grew in intensity for three months. In June, Russian envoy Victor Chernomyrdin went to Belgrade for talks with Milosevic. The Russian Foreign Minister recalled, "By late May we had reliable information that preparation for a ground invasion was in full swing" while another Russian source commented, "We told Milosevic that he needed to take a ground attack seriously."[1] Milosevic accepted the latest NATO proposals and the mediation was a success.

These two mediation attempts were similar in that the mediator attempted to persuade one side to make a concession because the other side would fight if no concession was forthcoming. This kind of advice, if it is believed, can be crucial in resolving conflicts. In bargaining situations, conflict is often avoidable if the parties have complete information but may erupt if there is uncertainty about resolve or other factors (Fearon 1995; Powell 2002). For instance, Argentina presumably would not have started the crisis if it knew that the British would fight rather than acquiesce in a military takeover of the islands. Argentina's

Andrew Kydd is Assistant Professor of Government, Harvard University, 1033 Massachusetts Ave. 318B, Cambridge, MA 02138 (akydd@wcfia.harvard.edu).

Earlier versions of this article were presented at the Research Frontiers in International Relations Conference, La Jolla, CA, at the PIPES Workshop, University of Chicago, at the 2000 meetings of the Midwest Political Science Association, Chicago, at the University of California, Davis, at the DVPW in Halle, Germany, and at the Olin Institute National Security Seminar Series, and I thank the participants for their comments. I would also like to thank Shaun Bowler for the conversation that started the whole project, and Jasen Castillo, Jörg Faust, Page Fortna, Erik Gartzke, Barbara Koremenos, Lisa Martin, Robert Powell, Holger Schmidt, Naunihal Singh, Monica Toft, Benjamin Valentino, and Barbara Walter for helpful feedback along the way.

[1] Newsweek 7/26/99. See also Daalder and O'Hanlon (2000, 203–5).

American Journal of Political Science, Vol. 47, No. 4, October 2003, Pp. 597–611

 ISSN 0092-5853

uncertainty about British resolve led it to take a chance on invading. Haig attempted to provide Argentina with information about British resolve, to reduce its uncertainty, but he was unpersuasive and conflict ensued. Chernomyrdin's analogous communication was more successful.

Why did Chernomyrdin succeed where Haig failed? The answer has to do with the mediator's motivations and corresponding incentives to tell the truth. A mediator who is biased in favor of one's opponent will have a strong incentive to claim that the opponent has high resolve whether or not this is true and will therefore not be believed. Haig acknowledged that his "sympathy was with the British," though he attempted to act impartially out of a belief that neutral mediators are more effective (Haig 1984, 266). However, if Haig preferred a solution involving an Argentine withdrawal from the islands, and believed that Argentina would withdraw if it thought Britain would fight, he would have every incentive to tell Argentina that Britain would fight, even if he thought Britain was unlikely to do so. Thus a mediator who is biased against you will have no credibility on the subject of the other side's resolve.

Perhaps more surprisingly, the same logic applies in the case of a mediator who is unbiased or has no policy preferences at all over the issue in dispute. An unbiased mediator who is simply interested in minimizing the probability of war will also have a strong incentive to tell each side that their opponent has high resolve and will fight unless they receive a concession. If the mediator is believed, each side will make a concession, increasing the likelihood of a peaceful settlement. Since this is what the mediator wants, the mediator has an incentive to make such statements even if they are not true. Hence the parties will not find such statements credible.

Only a mediator who is effectively "on your side" will be believed if she counsels restraint. A mediator who shares your policy preferences to some extent could be trusted to tell you if she thought the adversary had low resolve and was likely to back down even without a concession. She could therefore be trusted if she counseled the reverse, that the adversary has high resolve and will fight. This is the central reason behind Chernomyrdin's success. Russian preferences were known to be aligned with Serbia and therefore Russia was a credible mediator. If Russia said NATO would invade, Serbia could believe it.[2]

In this article I develop these arguments with a game theoretic model of mediation. The article has three sections. First, I place mediation in the context of what economists call "cheap talk," or costless but strategic communication. Second, I present a formal model of bargaining and mediation that shows when mediators are credible. Finally, I draw out the empirical implications of the model. An appendix follows with mathematical details of the model.

[2]For a conterargument that Russian preferences were no longer aligned with Milosevic and that the threat of an invasion was not important, see Stigler (2002/03).

Mediation and Cheap Talk

Mediation is often seen as one of the primary tools of conflict resolution, in both civil and international conflicts. Given the importance of mediation, it is not surprising that there is a large academic literature on the subject (for reviews, see Wall 1981; Wall and Lynn 1993; Kleiboer 1996; Wall, Stark, and Standifer 2001; for collections of essays, Touval and Zartman 1985; Mitchell and Webb 1988; Kressel and Pruitt 1989; Bercovitch and Rubin 1992; Vasquez et al. 1995; Bercovitch 1996; Crocker, Hampson, and Aall 1999). Despite an extensive literature on the subject, however, the theory of how mediation works is not well developed. Many unresolved questions remain about what makes for successful mediation. One of the most salient of these debates is about the impact of the mediator's motivations on the mediation process. Mediators are often thought to be more effective if they are unbiased or impartial (Fisher 1995). Young argues that, "the existence of a meaningful role for a third party will depend on the party's being perceived as an impartial participant (in the sense of having nothing to gain from aiding either protagonist). . ." (Young 1967, 81). Many practitioners agree, in Haig's view, "the honest broker must, above all, be neutral" (Haig, 1984, 266).

Some scholars, however, have questioned the importance of mediator impartiality (Touval 1982; Bercovitch and Houston 1996, 26; Carnevale and Arad 1996; Zartman and Touval 1996). Touval and Zartman argue that "mediators are seldom indifferent to the terms being negotiated. Even when they seek peace in the abstract, they try to avoid terms not in accord with their own interests" (Touval and Zartman 1989, 118). They see mediation as an exercise in power politics: "leverage is the ticket to mediation" (Touval and Zartman 1989, 129). Mediators apply leverage to one side and or the other to extract concessions. Bias does not prevent mediators from being successful, "impartiality is neither an indispensible condition of their acceptability, nor a necessary condition for the successful performance of an intermediary's functions" (Touval 1975, 56). The divergence between this conception of mediation and the more traditional image of a neutral facilitator has led some scholars to posit two different categories of mediator. Princen, for instance, distinguishes the "neutral" mediator who is both weak and impartial from the "principal" mediator who

is powerful and interested or biased (1992, 18).[3] While impartiality is appropriate for the weak mediator, bias is seen as acceptable, perhaps inevitable, for the powerful mediator.

This debate within the mediation literature is echoed in the literature on international institutions as well. Realist perspectives on mediation tend to view it as either epiphenomenal or as an opportunity for powerful and interested third parties to impose settlements to their liking in regional conflicts via a combination of carrots and sticks (Gelpi 1999, 117). Realists (and, in the context of the European Union, liberal intergovernmentalists) maintain that information providing mediators should have no impact on dispute resolution, whether they are biased or impartial (Moravcsik 1999, 278–9). In contrast, institutionalists maintain that international institutions can exert influence by providing information, even if they are powerless in a traditional sense (Keohane 1984; Koremenos, Lipson, and Snidal 2001). Neutral mediators are often thought to act in just this way, providing information that facilitates conflict resolution. These debates remain unresolved in part because of a lack of a firm theoretical understanding of the various roles that mediators play in the bargaining process. This problem can best be remedied by incorporating mediators directly into bargaining models of conflict.[4]

Here I focus on one of these roles, the provision of information about the resolve of the negotiating parties, and argue that certain aspects of this role can be thought of as an instance of what economists call cheap talk. Cheap talk is communication in strategic contexts that does not affect the payoffs directly, but may affect them indirectly if it conveys information that can cause the players to modify their behavior.[5] Cheap talk is often contrasted with "costly signals," gestures which have a direct impact on payoffs, and derive their credibility from this link. The classic illustration of a costly signal in international relations is when one party in a dispute mobilizes military forces during a crisis to demonstrate resolve; by increasing the risk of war the mobilization demonstrates a willingness to fight. An example of cheap talk in this case would be one party simply telling the other that it will fight unless the other side makes a concession. In itself, the communication is not costly, but if it changes the other side's beliefs about the first party's resolve, it could affect the outcome of the negotiation.

Cheap-talk models typically focus on a "sender" with private information who can communicate to a "receiver" who then takes an action that affects the payoffs of both parties. The central result of such models is that successful communication requires a certain amount of common interest, or at least lack of a conflict of interest. This means that cheap talk is usually quite effective in coordination games (Morrow 1994). If two people are trying to coordinate on a meeting place, there is no point in the sender misleading the receiver about where she will be. Truth-telling equilibria can be supported even if the sender is completely indifferent to what action the receiver takes. If a stranger asks the time of day, there is no positive reason to mislead her, so truth telling is an equilibrium.

In bargaining situations, where the interests of the parties are more directly opposed, cheap talk is more problematic because of a credibility problem. In the international relations context, states have an incentive to say they are "tough" or have low costs for fighting, because this will persuade the other side in a crisis to back down or concede the issue at stake. This incentive holds as much for states which are actually unwilling to fight as it does for those willing to fight, thus there is an incentive to bluff, or misrepresent one's true costs for fighting. One party cannot credibly just tell another that it has high resolve and will fight if it does not receive a concession, since they face an incentive to say this even if it is not true (Fearon 1995).[6]

If we consider mediation as a form of cheap talk, the first thing that becomes apparent is that we cannot use a mediator to solve the credibility problem that plagues cheap talk in bargaining directly. If the mediator is known to be credible to the other party, and the mediator will believe what she is told, then each party has every incentive to tell the mediator that it has high resolve and will fight unless it receives a concession, regardless of whether this is true. The mediator will then convey this to the other side; the other side will then believe that it faces a high-resolve type and will make a concession. Thus if there is an incentive to bluff to the other side, there will be an incentive to bluff to the mediator, and the mediator will have to discount statements about resolve. Interjecting a mediator, therefore, does

[3] See Smith (1985) for a similar distinction between "traditional" and "international" mediation and Touval (1985) for a response.

[4] For another recent game theoretic analysis of mediation, see Maoz and Tellis (2002).

[5] For a good introduction see Farrell and Rabin (1996); for the origins of the literature, see Crawford and Sobel (1982); for refinements and developments, see Farrell and Gibbons (1989a), Rabin (1990), Matthews, Okuno-Fujiwara, and Postlewaite (1991), Farrell (1993), Blume and Sobel (1995), and Austen-Smith and Banks (2000).

[6] An exception to this logic is analyzed by Farrell and Gibbons (1989b). While this model somewhat artificially generates a breakdown in bargaining if one or both sides claim to be tough, it is suggestive of how cheap talk might matter if there are opportunity costs of negotiation.

not solve the bluffing problem that plagues cheap talk in bargaining.[7]

This problem highlights two questions that any theory of mediation as information provision must answer. First, how does the mediator get the information that she is to provide? Second, when can the mediator credibly communicate this information to the negotiating parties? I do not focus on the first question here, except to note that there are many sources of information about the resolve of a state other than that state's communications. Once a mediator gets involved in a negotiation, she typically spends considerable effort learning about the dispute and the parties involved, researching the issue, talking to third parties, etc. Some state mediators, such as the United States or Russia, have intelligence capabilities that enable them to form their own estimates from clandestine sources. Suffice it to say that third parties usually exist who have information or opinions to contribute that could alter the parties' beliefs if they were credible.

Assuming the mediator has information to provide, when can they credibly provide it? Intuition might suggest that only an unbiased mediator could credibly provide such information because only an impartial source could be trusted. The theory of cheap talk suggests that the sender's interests must be either aligned with the receiver's, or the sender must be indifferent to the receiver's interests for a truth-telling equilibrium to be sustained. In the mediation context, the parties interests are opposed so the mediator clearly cannot be aligned with both of them at once.[8] However if the mediator were completely indifferent to the issue in dispute, the mediator would seem to be credible to both sides. This would seem to justify the focus on impartiality in the mediation literature; an impartial mediator can be trusted by both sides because she has no interest in favoring or harming either side. In fact, as outlined in the introduction, this intuiton does not turn out to be correct. To see why, it is necessary to consider the problem formally.

[7] A related strand in economics is the mechanism design literature (Myerson 1979, 1991; Banks and Calvert 1992). These models posit two actors with private information who send messages to a central "mediator" which processes the messages and issues instructions to the players that form an equilibrium. A fundamental problem with this literature, in an international relations context, is that the mediator here is really an automaton, not an actor with preferences and beliefs (for an exception, see Baliga, Corchon, and Sjöström 1997). In particular, in order to encourage honest revelation of private information, the mediator must deliberately induce a certain probability of conflict after receiving information sufficient to prevent this. This is not credible for a mediator in international relations who is a strategic actor who wishes, among other things, to prevent conflict.

[8] An interesting exception would be if there are multiple issues and the mediator is aligned with party A on issue 1 and with party B on issue 2.

The Model

Before laying out the model, I will discuss the scope of the analysis and define a key term, mediator bias.

The Scope of the Model

The model applies to situations with the following characteristics. First, the parties to the dispute are engaged in a negotiation such that if they reach an agreement they expect to suffer lower costs from conflict than they expect if they fail to reach an agreement. The lower expected level of conflict with an agreement results from the fact that at least one issue dividing the parties, the subject of the negotiations, has been at least temporarily settled. This may not reduce the level of violence right away, but the parties, and mediator, must believe that the agreement will help reduce the violence eventually, or bring it to an end sooner. The model makes no specific assumptions about whether the two sides are currently at peace and attempting to prevent a war or are negotiating the end of an ongoing conflict. The conflict resulting from bargaining failure may range from all-out war to low-intensity conflict, guerrilla activity, or terrorist strikes. An agreement may not lead to absolute and permanent peace; it could simply mean a temporary reduction in the level or intensity of conflict. The assumption behind the model is simply that bargaining success leads to a reduction in the expected level of mutual costs from fighting in comparison to what would have happened if the bargaining had failed.[9]

I make two assumptions about the mediator's preferences. First, I assume that the mediator prefers that there be an agreement rather than that the bargaining fail. More precisely, the mediator suffers a cost if the bargaining breaks down, so that whatever her other preferences, there are at least some agreements that the mediator would prefer to see made rather than have the negotiations collapse. All this assumption rules out is mediators who engage in mediation in order to prevent a resolution to the conflict because they gain some positive benefit from prolonging it. Second, I assume that mediators may or may not have

[9] For simplicity, the model focuses on a single issue which if resolved produces peace and if not resolved leads to conflict. However, parameters could easily be added representing the payoffs from conflicts over other unresolved issues without altering the results, so long as settling the issue at hand does not make settling the others harder, and hence increase the costs of conflict related to the other issues.

some other interests which lead them to have preferences over the different possible issue resolutions to the dispute. These preferences may be weak or strong, directly or indirectly related to the issues in dispute, and may cause the mediator to favor the interests of one party or the other. For instance, if Kissinger preferred to have Egypt get the Sinai back after the 1973 war, it was not because the United States cared about who owned the Sinai per se, but because he felt it would enable the United States to cultivate Egypt as a regional ally and erode Soviet influence in the Middle East (Quandt 1993, 186, 230). Thus the assumptions about mediator motivations are fairly broad, they have some desire to see negotiations succeed, and they may have some other preferences over the outcomes or possible deals.

Mediator bias is defined in terms of the preferences of the mediator. If the mediator's preferences are aligned with one party or the other, she is said to be biased in favor of that party. More precisely, for any two possible issue resolutions *A* and *B* over which the players have a conflict of interest such that player 1 prefers *A* to *B* and player 2 has the reverse preferences, if the mediator prefers *A* to *B*, I say the mediator is biased in favor of player 1, and if the mediator has the reverse preferences, she is biased in favor of player 2. If the mediator is indifferent between *A* and *B*, then she is defined to be unbiased or neutral. This definition of mediator bias follows from Young's conception of impartiality.

The most restrictive aspect of the model is that it focuses on only one of the many roles of mediators: the provision of information about the resolve of the parties in an attempt to extract concessions. Flesh and blood mediators do many other things; they make proposals, manipulate the agenda, cajole and flatter, and, if they represent powerful countries, offer carrots and threaten sticks in order to move the parties to agreement.[10] I am certainly not arguing that the role I focus on is more important than the other roles or that the other roles are not worthy of much study. I focus on information provision about resolve for two reasons. First, I argue that most mediators, at some point in their mediation experience, in the midst of all their other tasks and strategies, do attempt to extract concessions from the negotiating parties by warning them that without a concession the negotiations will fail and conflict will continue (or begin). Hence it is well worth knowing when this kind of tactic will be successful. Second, we have a set of theoretical tools which are well adapted to shed light on this question: models which have developed in the literature on crisis bargaining and cheap talk. So while the model does single out one tactic among many that mediators employ, I argue that that tactic is an important one which is used in most mediation efforts and merits careful study.[11]

The Structure of the Model

Often in international bargaining contexts one side is "satisfied" while the other is potentially "dissatisfied," where satisfied means prefering the status quo to conflict, and dissatisfied means prefering conflict to the status quo (Powell 1999, 88). Once the Argentines had executed their fait acompli and taken the islands, Argentina was satisfied because it would prefer to keep the islands rather than fight further. Britain was potentially dissatisfied, because it might have prefered to fight rather than live with the new status quo (as it turned out, Britain was dissatisfied and did fight). The key strategic problem is for the satisfied power to judge how likely the potentially dissatisfied power is to really be dissatisfied. To reduce the risk of conflict, the satisfied power may wish to make some concession to the other side, hoping to buy them off. The danger is that the concession will not be big enough and that the other side will decide to attack (or keep fighting if the war is already in progress) rather than accept it or live with the status quo.

The model captures this strategic dilemma by positing two players, 1 and 2, where player 1 is satisfied and player 2 is potentially dissatisfied. Player 1 is not sure if player 2 is dissatisfied or not and may make a concession in an effort to buy him off. The sequence of events is as follows. First, Nature determines player 2's type and the information that the mediator has about player 2's level of resolve. Next, the mediator can communicate with player 1 about player 2's level of resolve. Finally, player 1 and player 2 bargain together.[12]

The bargaining game is illustrated in Figure 1. Two players are bargaining over a pie represented by the unit interval. Player 1 proposes a division x. Player 2 can accept the proposal or reject it and fight. If player 2 accepts the proposal, player 1 gets x and player 2 gets $y = 1 - x$. If

[10] See Bercovitch and Houston (2000) for a discussion of the range of mediator strategies.

[11] It may be argued that isolating one tactic for study may distort our understanding by obscuring how the different tactics interact. While this may be true, I think an adequate understanding of the whole must be built up from a careful study of the parts, first taken in isolation, and then carefully combined to examine their interactions.

[12] Structured in this way, the model may seem to depict an advisor rather than a mediator. However, mediators often act like advisors in this sense; mediators frequently advise the negotiating parties, based on their sense of the other party's resolve, to make a concession to increase the chances of concluding an agreement.

FIGURE 1 The Bargaining Game

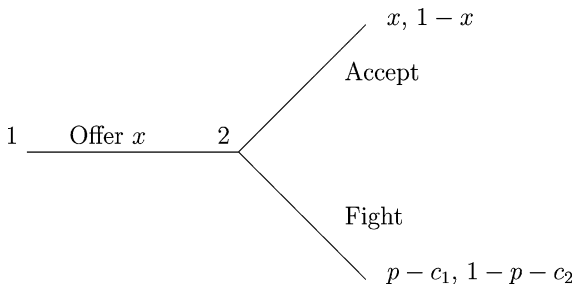


TABLE 1 Notation in the Model

x, (y)	Player 1's (2's) payoff from a deal
p	Player 1's chance of winning in a conflict
c_i	Player i's cost of conflict
β	The mediator's degree of bias
k	The upper bound on player 2's costs of fighting ($c_2 \in [0, k]$)
s	The status quo division of the issue
ϵ	The likelihood the mediator's information is in error
h	The likelihood player 2 has high costs ($c_2 > s - p$)
A, B	The boundaries of k between which player 1's offer depends on k

player 2 decides to fight, player 1's conflict payoff is $p - c_1$, while player 2 receives $1 - p - c_2$, where p is player 1's chance of winning the conflict and c_i is player i's cost of fighting. The mediator also has a preference function over the issue space βx and pays a cost if there is a conflict c_m. The mediator's payoff for conflict is therefore $\beta p - c_m$. If $\beta > 0$ the mediator is said to be biased in favor of player 1 because she prefers issue resolutions to the right (bigger values of x) as does player 1. If $\beta < 0$, the mediator is biased in favor of player 2, and if $\beta = 0$, the mediator is unbiased. Note if the mediator is unbiased, her payoff for any accepted offer is 0 and her payoff for conflict is $-c_m$, so the mediator is indifferent over the possible settlements and just wants to secure a deal.[13] Notation in the game is summarized in Table 1.

[13]Calvert (1985) and Myers (1998) present models demonstrating the usefulness of biased advice but their conception of bias is different from the above so their results speak to different issues. In the Calvert set-up, the advisors are not strategic actors with preferences, as the mediator is here, but simply provide recommendations of a course of action in accordance with certain probabilities, which may be biased toward one of the possible choices. Hence the credibility of the advisor—his willingness to tell the truth—is not at stake; rather the question is what kind of advice could be useful to the decision maker by actually influencing the decision. Decision makers prefer biased advisors in this set up because if the advisor says something against his known bias, it is highly informative. In my set-up, in contrast, players prefer biased mediators because only biased mediators are honest.

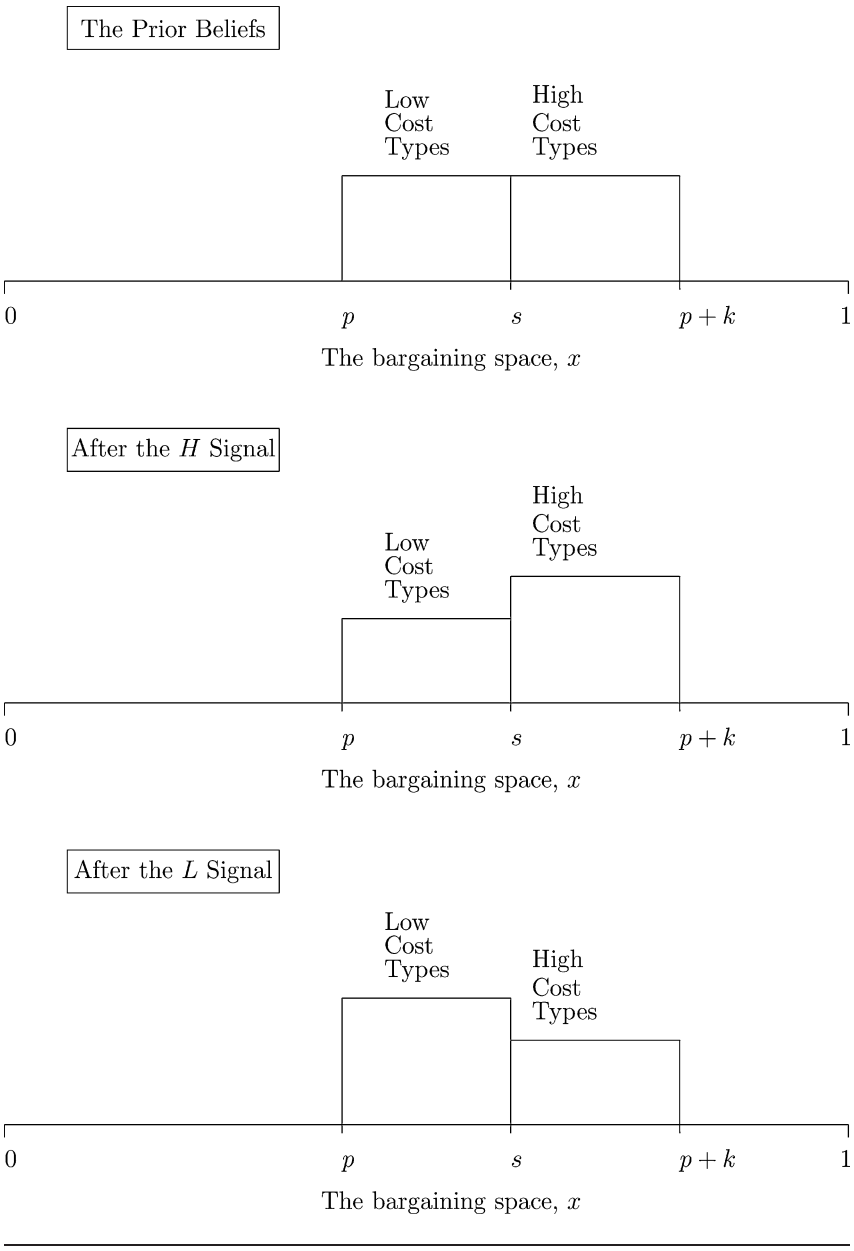
Beliefs

The uncertainty in the model is over player 2's cost for fighting, or resolve. Nature starts the game by choosing player 2's cost from a uniform distribution $c_2 \in [0, k]$. Thus player 2's reservation value in the bargining game, or the minimum offer it will accept rather than fight, ranges from p to $p + k$ as illustrated in Figure 2. The horizontal lines represent the bargaining space, x, ranging between zero and 1. Player 1 prefers issue resolutions to the right; player 2 prefers deals to the left. Player 1's likelihood of winning a conflict, p, is less than the status quo, s, hence player 1 prefers the status quo to conflict or is satisfied, while player 2 may not be, depending on player 2's cost of fighting. Types of player 2 with $c_2 \in [0, s - p)$ are dissatisfied or have "low" costs for fighting, types with $c_2 \in [s - p, k]$ are satisfied or have "high" costs for fighting. The height of the bars reflects the probability density for each type, or the likelihood of facing a type with the corresponding cost for fighting. In the top of the figure, the prior beliefs are illustrated, the uniform distribution previously mentioned. The prior belief that player 2 has high costs (the area under the right bar) is denoted h.

To model the mediator's additional information, after Nature chooses player 2's type we have her send a signal to the mediator about this fact. That is, Nature can send the mediator a signal saying, "Player 2 has high costs," or H for short, or she can tell player 1, "Player 2 has low costs," or L. These messages may not be correct, however. I do not assume that the mediator is perfectly informed about player 2's type, merely that the mediator has some additional information that might be useful to player 1. The likelihood that a message is in error is ϵ, and the corresponding likelihood that the message is accurate is $1 - \epsilon$ where $1 - \epsilon > \epsilon$. These messages reflect any information that the mediator has been able to gather about player 2's resolve.

The mediator's subsequent beliefs about whether player 2 has high or low costs can be derived using Bayes rule as shown in the appendix. The mediator's posterior beliefs given that she received the message H are illustrated in the middle of Figure 2. Her confidence that player 2 has high costs has increased, (the high cost bar is taller) and her belief that player 2 has low costs has declined. Her new belief that player 2 has high costs is denoted h_H. Conversely, if the mediator receives the L signal, her beliefs are

FIGURE 2 The Bargaining Space and Distribution of Reservation Values



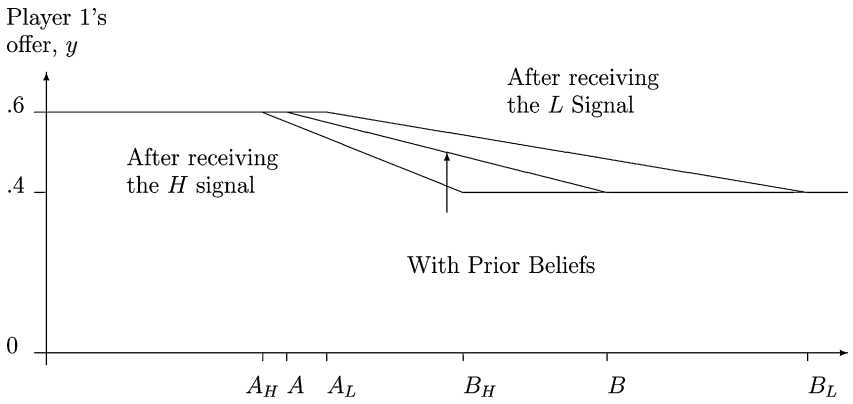
as shown in the bottom of Figure 2. Her confidence that player 2 has low costs has increased; her belief that player 2 has high costs has gone down. Her new belief that player 2 has high costs is h_L. The posterior belief that player 2 has high costs after receiving the high signal, h_H, is greater than the prior belief, h, which is in turn greater than the posterior belief after receiving the low signal, h_L, that is, $h_H > h > h_L$.[14]

[14]In Figure 2, the parameter values are $p = 0.4$, $s = 0.6$, $k = 0.4$, and $\epsilon = 0.4$, so that $h = 0.5$, $h_H = 0.6$, and $h_L = 0.4$.

I assume that the parties are not informed about what signal the mediator received and hence do not know for certain what the mediator's beliefs are. Thus after Nature moves, player 2 is informed of her type, player 1 is not, the mediator has some additional information that she can share with player 1, but the players are not sure what this information is.

In order to understand the intuition behind the results, it will help to think about how the equilibrium offer made by player 1 depends on her beliefs and how this influences the likelihood of conflict. The likelihood of

FIGURE 3 The Equilibrium Offer



conflict will in turn influence the mediator's incentive to be truthful.

The Equilibrium Offer

If player 1 knew player 2's type, she would make no concession to a high-cost type (offer $y = 1 - s$), and just buy off a low-cost type (offer $y = 1 - p - c_2$). In neither case would conflict occur. Thus with complete information, bargaining is successful, conflict does not occur, and so there is no need for a mediator.

With uncertainty about player 2's cost of fighting, things are different. The equilibrium offer is illustrated in Figure 3. The horizontal axis is the upper bound on the distribution of player 2's cost for fighting, k. Thus this dimension can be thought of as a measure of how likely player 2 is to have high costs, ex ante. The greater k is, the more likely player 2 is to have high costs, and be unwilling to fight. The smaller k is, the greater the likelihood that player 2 has low costs, and will fight if not bought off. The vertical axis is the equilibrium offer from player 1, y.

First, consider the case where player 1 is operating with her prior beliefs, perhaps because the mediator is untrustworthy or is not conveying any information. This case is represented by the middle of the three downward sloping lines. For very low values of k, player 1 thinks player 2 is very likely to be dissatisfied, and so player 1 makes the maximum concession which suffices to buy off all types of player 2. This offer is $y = 1 - p$, or 0.6 in the example used so far. As the likelihood that player 2 has high costs goes over a certain threshold, denoted A and derived in the appendix, player 1 starts making a smaller offer. Player 1's offer declines the more likely player 2 is to have high costs, until a second threshold, B, is reached, at which it reaches a minimum. For k greater than B, player 1 thinks player 2 is very likely to have high costs and be unwilling to fight. Player 1 therefore makes no concession at all, proposing that the status quo remain in place, offering $y = 1 - s$, or 0.4.

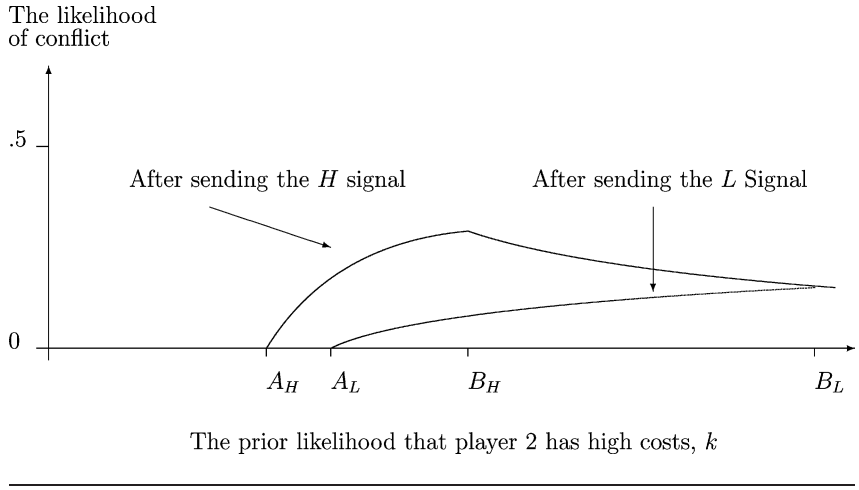
Now consider the case where player 1 has received the H signal from the mediator and believed it, as illustrated in the left-most line. Since player 1 has become more convinced that player 2 has high costs, and will not fight, she decides to make a smaller offer in equilibrium. This has the effect of shifting the curve to the left. Player 1 stops making a full concession earlier at A_H (which is less than A), and her offer decreases more quickly, until player 1 is so confident that she is facing a weak type that she makes no concession at all, when $k > B_H$.

Finally, on the right is the case where player 1 has received the L signal and believes it. Player 1 is now more convinced that player 2 has low costs and will fight. This makes player 1 more generous with her offer, which shifts the curve to the right. Player 1 must be more convinced that player 2 has high costs before she will stop making a full concession, at A_L, and must be very convinced before insisting on the status quo at B_L.

The essential point is that if player 1 receives the H message from the mediator and believes it, she will make a smaller offer to player 2 than she would if she received the L message and believed it.[15] To tell player 1 that player

[15]Note, for extreme values of the prior beliefs—at the left and right hand side of Figure 3—no information from the mediator can change player 1's course of action. On the left, she is so convinced that player 2 is tough that she will make a full concession even if the mediator tells her player 2 is weak, and on the right she is so convinced that player 2 is weak that she will make no concession even if the mediator tells her he is resolved.

FIGURE 4 The Likelihood of Conflict (Mediator's perspective)



2 is likely to have high costs, therefore, is to encourage her to lower her offer. To tell her that player 2 has low costs is to encourage her to raise her offer. This fact has ramifications for the likelihood of bargaining failure and conflict.

The Likelihood of Conflict

If the mediator is credible, player 1 will condition her offer on what the mediator says. The size of her offer, in turn, will influence the probability that negotiations will fail and that conflict will ensue. Essentially, the smaller the offer, the higher the probability of conflict. The smaller the offer, the more low-cost types of player 2 will reject it and decide to fight. For the mediator, therefore, to tell player 1 that player 2 has high costs is to increase the likelihood of conflict.

This relationship is illustrated in Figure 4. The horizontal axis is the same as in Figure 3, the upper bound on the prior distribution of player 2's costs, k. The vertical axis is the probability of bargaining failure and conflict. The figure takes the perspective of a mediator who has received the H signal from Nature, so is more persuaded that player 2 has high costs and will not fight. The mediator can now either communicate this to Player 1, causing her to make a lower offer, or lie and say she got the L signal, causing her to make a higher one.

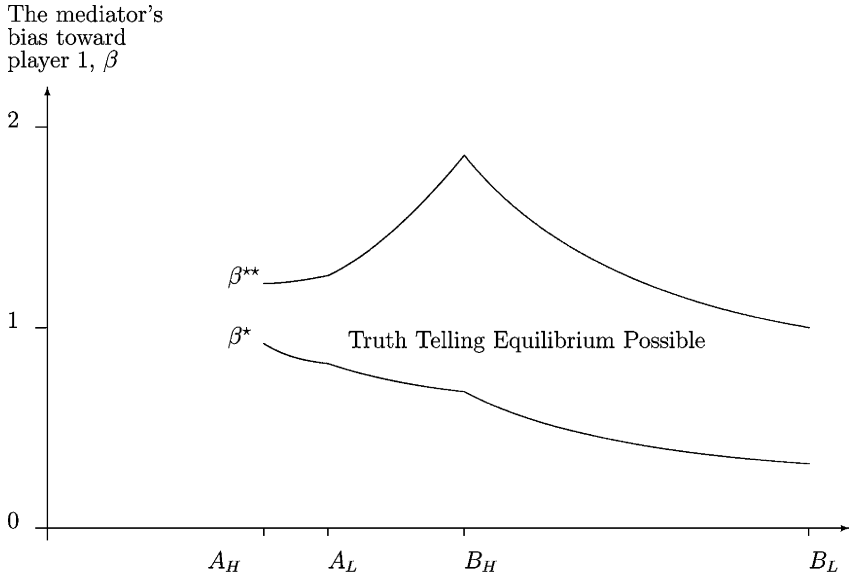
The top curve, starting at A_H, is the probability of conflict if the mediator truthfully sends the H signal. It rises as player 1's offer declines, until it reaches a maximum at B_H where player 1's offer reaches its minimum at $y = 1 - s$. From then on it declines as k increases, reflecting the fact that player 2 is growing increasingly unlikely to have low costs and be willing to fight. The bottom curve, starting at A_L, is the probability of conflict if the mediator lies instead, telling player 1 that she received the L signal rather than the H signal. This increases until it intersects the top curve, and then is identical to it, trailing off after B_L.

The top curve is everywhere above the bottom curve. This means that, regardless of the prior likelihood that player 2 has high costs, k, the likelihood of conflict is always higher if the mediator tells player 1 that player 2 has high costs than it is if she says that player 2 has low costs. If the mediator thinks player 2 does have high costs, being truthful will increase the likelihood of conflict, whereas lying will reduce it. For a mediator who wants, among other things, to facilitate an agreement that reduces the likelihood of conflict, this presents a troubling incentive to lie. It can easily be shown in a figure similar to Figure 4 that if the mediator received the L signal and thinks player 2 is likely to have low costs, sending the L signal (in this case truthfully) also reduces the likelihood of conflict. In the end, regardless of what the mediator actually believes, telling player 1 that player 2 has high costs increases the chance of conflict, telling her that player 2 has low costs will reduce it. This has crucial implications for the mediator's incentives to be truthful.

The Credibility of the Mediator

For a mediator to be credible, it must be the case that she has an incentive to tell the truth. Assume that player 1 believes the mediator honestly announces whether she received the H or L signal, and will act accordingly. The mediator must then have an incentive to do just that,

FIGURE 5 The Credibility of the Mediator



communicate honestly about the signal. That is, if she receives the *H* signal, she must have an incentive to report this to player 1 rather than say she got the *L* signal, and if she got the *L* signal, she must prefer to pass that on rather than pretend she got the *H* signal. These two conditions lead to constraints on the preferences the mediator can have if she is to be credible, constraints which are illustrated in Figure 5. Recall that the mediator's bias is measured by β, if β is positive, the mediator is biased in favor of player 1, if β is negative, the mediator is biased in favor of player 2, and if β is zero, the mediator is unbiased.

If the mediator receives the low-costs signal, *L*, she must prefer to tell player 1 this, rather than lie and pretend she got the *H* signal instead. Being truthful in this case causes player 1 to make a larger offer than she would have if the mediator had lied and sent the *H* signal, which decreases the likelihood of conflict. Given that the mediator suffers a cost if conflict occurs, this increases the mediator's payoff. This condition basically puts an upper bound on how strongly the mediator can prefer higher-issue resolutions, an upper bound on β denoted β**. An unbiased mediator would be happy to be truthful in this case, since being truthful leads to a lower likelihood of conflict. If the mediator's preferences are identical to player 1's, if $c_m = c_1$ and $\beta = 1$, she is also happy to be truthful. A mediator would only wish to lie in this case if her preferences over the issue were so strong that she would prefer that player 1 make no concession even if player 1, with the same information, would prefer to make one. To prefer to lie, the mediator would have to be more royalist than the king. Thus this constraint puts a (fairly high) upper bound on how strongly the mediator wants to minimize concessions to player 2.

If the mediator receives the high costs signal, *H*, she faces more of a dilemma. Telling the truth will encourage player 1 to make a relatively small offer, which will precipitate a higher chance of conflict. Lying will convince player 1 to make a larger offer which is more likely to be accepted. Thus if the mediator is unbiased and simply wants to minimize the chance of conflict, she will face an insuperable incentive to lie. Lying will make peace more likely; telling the truth will make conflict more likely. Thus the unbiased mediator faces a serious credibility problem. To have an incentive to tell the truth in these circumstances, the mediator must be at least somewhat biased in favor of player 1. If the mediator prefers an issue resolution associated with a smaller offer to that associated with a larger one (β is positive), then she has an incentive to encourage player 1 to make a smaller offer if she thinks player 2 is likely to accept it. This policy preference outweighs the downside of making the smaller offer, the increased chance of conflict. Thus this constraint places a lower bound on how biased in favor of player 1 the mediator can be in a truthtelling equilibrium, a positive lower bound on β denoted β*.

The upper and lower bounds on the mediator's bias are illustrated in Figure 5. The horizontal axis is once again the upper bound on the prior distribution of player

2's costs, k. The vertical axis is β, the mediator's bias. The lower curve is the lower bound, $\beta^{\star}$, and the upper curve is the upper bound, $\beta^{\star\star}$.[16] Between the lines, a truth-telling equilibrium is possible, above and below; only babbling equilibria exist.[17] Immediately apparent is the fact that the mediator must be biased in favor of player 1 to have the proper incentives to tell the truth regardless of the signal she has received.

Empirical Implications

The model shows that the effect of mediation is to reduce the likelihood of conflict and encourage states to make concessions. Dixon (1996), in a study of 640 postwar interstate disputes, supports this implication and finds that mediation and communication are associated with preventing escalation and promoting peaceful resolution of conflicts. Walter, in a study of civil wars, finds that "governments and rebels are 39% more likely to bargain successfully with the help of a mediator than on their own" (Walter 2002, 82).[18]

Beyond this, the main empirical implication of the model can be stated as follows.

Hypothesis 1 *Mediators who attempt to persuade one side to make a concession because the other side has high resolve must be biased in favor of the side they are communicating with in order to be successful.*

Supporting this hypothesis, in a recent study of third-party interventions, Regan finds that neutral interventions are associated with longer conflicts than biased interventions (Regan 2002).

In addition, the model implies that biased mediators succeed by getting the side toward which they are biased to make concessions. This suggests the following hypothesis.

Hypothesis 2 *Within the sample of successful mediation efforts, in the cases in which the mediator is biased toward one of the parties, that party will make a larger concession in the negotiation in comparison with what the average party does in cases in which the mediator is unbiased.*

This hypothesis has been tested by Gelpi on a dataset of mediation efforts by state leaders in crises between 1918 and 1988 (based on the International Crisis Behavior dataset). Gelpi studies the effect of assymetric alliance ties on mediation. Building on the work of institutionalist scholars, he argues that mediators which have alliance ties to one side act to constrain that side, and get them to make a concession. Alliances, that is, act as instruments of control as well as signaling devices toward adversaries. He tests the hypothesis that an alliance tie to the mediator should make the challenger do worse in a crisis and finds it to be statistically significant and substantively important (Gelpi 1999, 136–7).[19]

A final implication of the model has to do with why good mediators are rare. The effective mediator in the model is one who is biased toward one side, but has information to contribute about the other side, the side against whom she is biased. It may be the case that the more biased you are in favor of one side's position, the less new information you have to contribute about the other side. Haig, for instance, had better judgment about whether the British would fight over the Falklands due to his long association with British officials, but this long association undoubtedly also contributed to his bias in their favor, making him useless to the Argentines as a mediator. In some cases it may be difficult to find the requisite mediator who has enough exposure to the other side to acquire new information, yet not so much as to come to share their preferences, and hence lose their usefulness.

Conclusion

Mediation can indeed be useful in bargaining situations. By providing information about the resolve of the parties, mediators can reduce the likelihood of conflict. However, paradoxically, in order to be believed when they attempt to provide this information, mediators must be biased. An unbiased mediator who simply wants to prevent conflict will not be credible to the parties invovled in the negotiations because she could not be trusted to send messages that might increase the likelihood of conflict. If she cannot be trusted to send messages that increase the likelihood of conflict, she cannot be trusted when she sends messages that decrease it either, and hence she will have no credibility. Only a biased mediator that shares the preferences

[16] The parameters are the same as before, with the addition of $c_m = 0.3$ and $c_1 = 0.3$.

[17] In babbling equilibria the sender sends a message at random, and the receiver consequently receives no information and does not condition her behavior on the signal. Babbling equilibria are also possible between the lines, and for values of k below A_H and above B_L. In cheap-talk games, babbling equilibria are always possible, the question of interest is under what conditions truthful communication is possible.

[18] Most quantitative studies take it for granted that mediation works and focus on what features of the mediator or dispute make for more or less successful mediation, i.e., Bercovitch and Langley (1996), Bercovitch and Houston (1993).

[19] Admittedly, many of the mediators in Gelpi's sample are great powers who may be using carrots and sticks to influence their alliance partners, a different mechanism from that considered here.

of one of the parties in the negotiations will be credible in this context. This finding supports the arguments of mediation scholars such as Touval and Zartman who have long held that neutrality is not a necessary attribute of a successful mediator. However, the model develops the idea further in two important ways. First, it shows that a certain degree of bias is not only acceptable but is actually necessary for some roles that mediators play. Second, it demonstrates this result for a role, information provision, that some scholars have argued properly belongs to "neutral" weak mediators rather than powerful, and potentially biased, mediators. This result should lead us to reevaluate the theoretical underpinnings of this conventional distinction in the literature.

Appendix

The mediator's belief after receiving the signal that player 2 has high costs is, from Bayes rule, $h_H = \frac{(1-\epsilon)h}{(1-\epsilon)h+\epsilon(1-h)}$, substituting in $h = \frac{p+k-s}{k}$ this becomes $h_H = \frac{(1-\epsilon)(p+k-s)}{(1-\epsilon)(p+k-s)+\epsilon(s-p)}$. After receiving the signal that player 2 has low costs, the mediator's beliefs are $h_L = \frac{\epsilon h}{\epsilon h+(1-\epsilon)(1-h)}$ or $h_L = \frac{\epsilon(p+k-s)}{\epsilon(p+k-s)+(1-\epsilon)(s-p)}$.

The solution concept I employ is perfect Bayesian equilibrium.

I restrict attention to offers $x \in [p, s]$. Player 1 can buy off all types of player 2 by offering $x = p$, so all offers $x < p$ are dominated by $x = p$. Similarly, any offer $x > s$ would be rejected in a trivially more complicated game in which player 2 could reject an offer but not attack, since player 1 has no credible threat to fight. So interesting offers are located between p and s.

High-cost types of player 2 will accept any offer $x \leq s$. Low-cost types will accept the offer if $x \leq p + c_2$, and fight if $x > p + c_2$.

First consider the case where player 1 operates on the basis of prior beliefs. The payoff for offering $x \in [p, s]$ is the likelihood of acceptance times the value of the offer, plus the likelihood of rejection times the value of war, $(h + (s - x)\frac{1-h}{s-p})x + (1 - (h + (s - x)\frac{1-h}{s-p}))(p - c_1)$. For the maximum concession, $x = p$, this reduces to just p, because all types of player 2 will accept the offer. For the minimum concession, $x = s$, this reduces to $hs + (1 - h)(p - c_1)$ which could be greater or less than p depending on the prior beliefs.

The payoff could have a maximum at $x = p$, $x = s$, or some interior point. To find possible interior maxima, we can set the derivative equal to zero, $h + \frac{1-h}{s-p}s - \frac{1-h}{s-p}2x + \frac{1-h}{s-p}(p - c_1) = 0$ and solve for x which gives $x^0 = \frac{\frac{h}{1-h}(s-p)+s+p-c_1}{2}$ or $x^0 = p + \frac{k-c_1}{2}$.

By comparing this with the payoffs for no concession and a full concession, we can determine two cutoff values for k. For $k < A$, player 1 makes the full concession, $x = p$. For $k \in (A, B)$, player 1 makes the interior offer, and for $k > B$ player 1 offers the status quo where, $A \equiv c_1$ and $B \equiv 2(s - p) + c_1$.

The likelihood of war for a maximal concession, $x = p$, is zero. As k increases and the offer decreases, the likelihood of war as a function of the offer is $(1 - (h + (s - x)\frac{1-h}{s-p}))$. Substituting in the equilibrium offer, we can find the likelihood of war as a function of player 2's maximum cost $(1 - (h + (s - (p + \frac{k-c_1}{2}))\frac{1-h}{s-p}))$ which reduces to $\frac{1}{2} - \frac{c_1}{2}\frac{1}{k}$. Once the offer declines to s, the likelihood of war is just $1 - h$ or $\frac{s-p}{k}$.

Now consider the case where player 1 has received and believed the H signal. The payoff for an offer x is $(h_H + (s - x)\frac{1-h_H}{s-p})x + (1 - (h_H + (s - x)\frac{1-h_H}{s-p}))(p - c_1)$. Once again with the maximum concession, $x = p$, all types of player 2 will accept so this reduces to p. For no concession, $x = s$, this reduces to $h_H s + (1 - h_H)(p - c_1)$ which, given that $h_H > h$, is better than in the previous case.

The interior solution is found by taking the derivative and setting it equal to zero, $h_H + \frac{1-h_H}{s-p}s - \frac{1-h_H}{s-p}2x + \frac{1-h_H}{s-p}(p - c_1) = 0$ and solving for x, which gives $x(H) = \frac{\frac{h_H}{1-h_H}(s-p)+s+p-c_1}{2}$ or $x(H) = \frac{\frac{1-\epsilon}{\epsilon}(p+k-s)+s+p-c_1}{2}$. Note that if $\epsilon = 0.5$, that is the signal is uninformative, this reduces to the previous case based on the priors. As ϵ decreases from 0.5, the signal grows more informative, the equilibrium concession decreases.

The boundaries for the interior offer in this case are $A_H \equiv s - p + \frac{\epsilon}{1-\epsilon}(p - s + c_1)$, and $B_H \equiv s - p + \frac{\epsilon}{1-\epsilon}(s - p + c_1)$. These also reduce to the former case when the signal is uninformative, and decrease as the signal accuracy improves.

The likelihood of war for interior offers is $(1 - (h_H + (s - x(H))\frac{1-h_H}{s-p}))$ which reduces to $\frac{1}{2} - \frac{c_1}{2}\frac{\epsilon}{(1-\epsilon)(p+k-s)+\epsilon(s-p)}$. If the signal is inaccurate this is equivalent to the previous case, and as the signal gets more accurate, the likelihood of war increases, which is a result of the smaller offer because player 1 grows more convinced that player 2 has high costs and will not fight. If the mediator lies, and sends the L signal instead, the probability of war is $(1 - (h_H + (s - x(L))\frac{1-h_H}{s-p}))$ or $\frac{1}{2}\frac{\frac{\epsilon^2}{1-\epsilon}(p+k-s)+\epsilon(s-p-c_1)}{(1-\epsilon)(p+k-s)+\epsilon(s-p)}$. These two probabilities of war are illustrated in Figure 4.

Now consider the case where player 1 has received and believed the L signal. Here the payoff for making offer x is $(h_L + (s - x)\frac{1-h_L}{s-p})x + (1 - (h_L + (s - x)\frac{1-h_L}{s-p}))(p - c_1)$. Once again, the maximum concession, $x = p$,

will be accepted for sure, yeilding p, while the minimum concession, $x = s$, will give $h_L s + (1 - h_L)(p - c_1)$ which is worse than in the previous two cases. The interior solution is found by differentiating and setting equal to zero, $h_L + \frac{1-h_L}{s-p}s - \frac{1-h_L}{s-p}2x + \frac{1-h_L}{s-p}(p - c_1) = 0$, and solving for x, producing $x(L) = \frac{\frac{h_L}{1-h_L}(s-p)+s+p-c_1}{2}$ or $x(L) = \frac{\frac{\epsilon}{1-\epsilon}(p+k-s)+s+p-c_1}{2}$. The boundaries for the interior offer in this case are $A_L \equiv s - p + \frac{1-\epsilon}{\epsilon}(p - s + c_1)$, and $B_L \equiv s - p + \frac{1-\epsilon}{\epsilon}(s - p + c_1)$.

Now consider the mediator's decision about what to say.

For $k < A_H$ even if player 1 has beliefs h_H, she will make the maximum offer, p, which will be accepted for sure. Therefore the mediator's communication can have no impact on player 1's behavior, and hence, on the mediator's payoff, which will be βp regardless of her communication. Since the mediator is indifferent between telling the truth and lying, both truth-telling and babbling equilibria will exist.

For $k \in [A_H, \min\{A_L, B_H\}]$, if player 1 receives the L signal and believes it, she will make the maximal offer, p, which will be accepted for sure, while if she receives the H signal and believes it, she will make an interior offer, $x(H)$, which may be rejected or accepted.

First consider the case where the mediator has received the H signal. If the mediator credibly communicates this to player 1, player 1 will offer $x(H)$ which will yield a payoff for the mediator of $(h_H + (s - x(H))\frac{1-h_H}{s-p})\beta x(H) + (1 - (h_H + (s - x(H))\frac{1-h_H}{s-p}))(\beta p - c_m)$ while lying will convince player 1 to offer only p, leading to a payoff of βp for the mediator. Telling the truth beats lying if $\beta > \beta^\star \equiv \frac{c_m}{\frac{1}{2}\frac{1-\epsilon}{\epsilon}(p+k-s)+\frac{1}{2}(s-p+c_1)}$.

Now consider the case in which the mediator receives the L signal. To convey this to player 1 and be believed would yield a payoff of βp, since war is avoided for sure. To lie and send the H message, would cause player 1 to offer $x(H)$, giving a payoff for the mediator of $(h_L + (s - x(H))\frac{1-h_L}{s-p})\beta x(H) + (1 - (h_L + (s - x(H))\frac{1-h_L}{s-p}))(\beta p - c_m)$. Thus telling the truth beats lying if $\beta < \beta^{\star\star} \equiv \frac{c_m}{(\frac{\epsilon}{1-\epsilon} - \frac{1}{2}\frac{1-\epsilon}{\epsilon})(p+k-s)+\frac{1}{2}(s-p+c_1)}$. It can be shown that for $\epsilon \in \{0, 0.5\}$, that is for informative signals, $0 < \beta^\star < \beta^{\star\star}$. For a truth-telling equilibrium to exist, the mediator must be biased in favor of player 1, $\beta \in [\beta^\star, \beta^{\star\star}]$.

For $k \in [A_L, B_H]$, if player 1 receives the L signal and believes it she will offer $x(L)$, whereas if she receives the H signal and believes it she will shift to $x(H)$. If the mediator receives the H signal, and conveys this to player 1 and is believed, the mediator's payoff is $(h_H + (s - x(H))\frac{1-h_H}{s-p})\beta x(H) + (1 - (h_H + (s - x(H))\frac{1-h_H}{s-p}))(\beta p - c_m)$. If the mediator lies and sends the L signal, the payoff will be $(h_H + (s - x(L))\frac{1-h_H}{s-p})\beta x(L) + (1 - (h_H + (s - x(L))\frac{1-h_H}{s-p}))(\beta p - c_m)$. Telling the truth beats lying if β is greater than $\beta^\star \equiv \frac{c_m}{\frac{1}{2}(\frac{1-\epsilon}{\epsilon} - \frac{\epsilon}{1-\epsilon})(p+k-s)+c_1}$. If the mediator receives the L signal, and conveys this to player 1, the payoff to the mediator is $(h_L + (s - x(L))\frac{1-h_L}{s-p})\beta x(L) + (1 - (h_L + (s - x(L))\frac{1-h_L}{s-p}))(\beta p - c_m)$. Lying by sending the H signal instead produces a payoff for the mediator equal to $(h_L + (s - x(H))\frac{1-h_L}{s-p})\beta x(H) + (1 - (h_L + (s - x(H))\frac{1-h_L}{s-p}))(\beta p - c_m)$. Telling the truth beats lying if $\beta < \beta^{\star\star} \equiv \frac{c_m}{\frac{1}{2}(\frac{\epsilon}{1-\epsilon} - \frac{1-\epsilon}{\epsilon})(p+k-s)+c_1}$.

For $k \in [B_H, A_L]$, if player 1 believes the H signal, she will offer the status quo, s, and if she receives the L signal she will only offer p. If the mediator receives the H signal and conveys this to player 1, the payoff is $h_H\beta s + (1 - h_H)(\beta p - c_m)$. If the mediator lies instead, the payoff is just βp because war is avoided. Telling the truth beats lying if $\beta > \beta^\star \equiv \frac{c_m}{\frac{1-\epsilon}{\epsilon}(p+k-s)}$. If the mediator receives the L signal and conveys it to player 1, the payoff is βp. If the mediator lies instead, the payoff is $h_L\beta s + (1 - h_L)(\beta p - c_m)$. Telling the truth beats lying if $\beta < \beta^{\star\star} \equiv \frac{c_m}{\frac{\epsilon}{1-\epsilon}(p+k-s)}$.

For $k \in [\max\{A_L, B_H\}, B_L]$, if player 1 believes the L signal, she will offer $x(L)$, and if she believes the H signal, she will make the minimal offer, s. If the mediator receives the H signal and credibly communicates it to player 1 her payoff is $h_H\beta s + (1 - h_H)(\beta p - c_m)$. If she lies and sends the L signal instead, the payoff will be $(h_H + (s - x(L))\frac{1-h_H}{s-p})\beta x(L) + (1 - (h_H + (s - x(L))\frac{1-h_H}{s-p}))(\beta p - c_m)$. Telling the truth beats lying if β is greater than $\beta^\star \equiv \frac{c_m}{(\frac{1-\epsilon}{\epsilon} - \frac{1}{2}\frac{\epsilon}{1-\epsilon})(p+k-s)+\frac{1}{2}(p-s+c_1)}$. If the mediator receives the L signal and communicates this to player 1, her payoff is $(h_L + (s - x(L))\frac{1-h_L}{s-p})\beta x(L) + (1 - (h_L + (s - x(L))\frac{1-h_L}{s-p}))(\beta p - c_m)$ whereas lying by sending the H signal would yield $h_L\beta s + (1 - h_L)(\beta p - c_m)$. Telling the truth beats lying if $\beta < \beta^{\star\star} \equiv \frac{c_m}{\frac{1}{2}\frac{\epsilon}{1-\epsilon}(p+k-s)+\frac{1}{2}(p-s+c_1)}$.

For $k > B_L$, regardless of player 1's beliefs, she will make the minimal offer, s, which will produce a payoff for the mediator of $h\beta(s) + (1 - h)(p\beta(1) - c_m)$. Since the mediator's communication cannot affect her payoff, there will be truth-telling and babbling equilibria.

References

Austen-Smith, David, and Jeffrey S. Banks. 2000. "Cheap Talk and Burned Money." *Journal of Economic Theory* 91(1):1–16.

Baliga, Sandeep, Luis C. Corchon, and Tomas Sjöström. 1997. "The Theory of Implementation When the Planner is a Player." *Journal of Economic Theory* 77(1):15–33.

Banks, Jeffrey S., and Randall L. Calvert. 1992. "A Battle of the Sexes Game with Incomplete Information." *Games and Economic Behavior* 4(3):347–72.

Bercovitch, Jacob. 1996. *Resolving International Conflicts: The Theory and Practice of Mediation.* Boulder, CO: Lynne Rienner.

Bercovitch, Jacob, and Allison Houston. 1993. "Influence of Mediator Characteristics and Behavior on the Success of Mediation in International Relations." *International Journal of Conflict Management* 4(4):297–321.

Bercovitch, Jacob, and Allison Houston. 1996. "The Study of International Mediation: Theoretical Issues and Empirical Evidence." In *Resolving International Conflicts: The Theory and Practice of Mediation*, ed. Jacob Bercovitch. Boulder, CO: Lynne Rienner, pp. 11–35.

Bercovitch, Jacob, and Allison Houston. 2000. "Why Do They Do It Like This? An Analysis of the Factors Influencing Mediation Behavior in International Conflcits." *Journal of Conflict Resolution* 44(2):170–202.

Bercovitch, Jacob, and Jeffrey Langley. 1993. "The Nature of the Dispute and the Effectiveness of International Mediation." *Journal of Conflict Resolution* 37(4):670–91.

Bercovitch, Jacob, and Jeffrey Z. Rubin. 1992. *Mediation in International Relations: Multiple Approaches to Conflict Management*. New York: St. Martins Press.

Blume, Andreas, and Joel Sobel. 1995. "Communication-Proof Equilibria in Cheap-Talk Games." *Journal of Economic Theory* 65(2):359–82.

Calvert, Randall. 1985. "The Value of Biased Information: A Rational Choice Model of Political Advice." *Journal of Politics* 47(2):530–55.

Carnevale, PeterJ., and Sharon Arad. 1996. "Bias and Impartiality in International Mediation." In *Resolving International Conflicts: The Theory and Practice of Mediation*, ed. Jacob Bercovitch. Boulder, CO: Lynne Rienner, pp. 39–53.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50(6):1431–51.

Crocker, Chester A., Fen Osler Hampson, and Pamela Aall. 1999. *Herding Cats: Multiparty Mediation in a Complex World*. Washington, D.C.: United States Institute of Peace Press.

Daalder, Ivo H., and Michael E. O'Hanlon. 2000. *Winning Ugly: NATO's War to Save Kosovo*. Washington, D.C.: Brookings.

Dixon, William J. 1996. "Third Party Techniques for Preventing Conflict Escalation and Promoting Peaceful Settlement." *International Organization* 50(4):653–81.

Farrell, Joseph. 1993. "Meaning and Credibility in Cheap-Talk Games." *Games and Economic Behavior* 5(4):514–31.

Farrell, Joseph, and Robert Gibbons. 1989a. "Cheap Talk with Two Audiences." *American Economic Review* 79(5):1214–23.

Farrell, Joseph, and Robert Gibbons. 1989b. "Cheap Talk Can Matter in Bargaining." *Journal of Economic Theory* 48(1):221–37.

Farrell, Joseph, and Matthew Rabin. 1996. "Cheap Talk." *Journal of Economic Perspectives* 10(3):103–18.

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

Fisher, Ronald J. 1995. "Pacific, Impartial Third-Party Intervention in International Conflict: A Review and Analysis." In *Beyond Confrontation: Learning Conflict Resolution in the Post-Cold War Era*, ed. John A. Vasquez, James Turner Johnson, Sanford Jaffe, and Linda Stamato. Ann Arbor: University of Michigan Press, pp. 39–59.

Gelpi, Christopher. 1999. "Alliances as Instruments of Intra-Allied Control." In *Imperfect Unions: Security Institutions over Time and Space*, ed. Helga Haftendorn, Robert O. Keohane, and Celeste Wallander. Oxford: Oxford University Press, pp. 107–39.

Haig, Alexander. 1984. *Caveat: Realism, Reagan and Foreign Policy*. New York: Macmillan.

Keohane, Robert. 1984. *After Hegemony: Cooperation and Discort in the World Political Economy*. Princeton: Princeton University Press.

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

Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. "The Rational Design of International Institutions." *International Organization* 55(4):761–800.

Kressel, Kenneth, and Dean G. Pruitt, eds. 1989. *Mediation Research: The Process and Effectiveness of Third-Party Intervention.* San Francisco: Jossey Bass Publishers.

Maoz, Zeev, and Lesley G. Terris. 2002. "The Strategy of Mediation: A Rational Model of Mediator's Choices." Unpublished manuscript, Tel-Aviv University.

Matthews, Stephen A., Masahiro Okuno-Fujiwara, and Andrew Postlewaite. 1991. "Refining Cheap Talk Equilibria." *Journal of Economic Theory* 55(2):247–73.

Mitchell, C. R., and K. Webb. 1988. *New Approaches to International Mediation.* New York: Greenwood Press.

Moravcsik, Andrew. 1999. "A New Statecraft? Supranational Entrepreneurs and International Cooperation." *International Organization* 53(2):267–306.

Morrow, James D. 1994. "Modeling the Forms of International Cooperation: Distribution versus Information." *International Organization* 48(3):387–423.

Myers, Marissa. 1998. "When Biased Advice is a Good Thing: Information and Foreign Policy Decision Making." *International Interaction* 24(4):379–403.

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

Myerson, Roger B. 1991. "Analysis of Incentives in Bargaining and Mediation." In *Negotiation Analysis*, ed. H. Peyton Young. Ann Arbor: University of Michigan Press, pp. 67–85.

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

Powell, Robert. 2002. Bargaining Theory and International Conflict. *Annual Review of Political Science* 5:1–30.

Princen, Thomas. 1992. *Intermediaries in International Conflict*. Princeton: Princeton University Press.

Quandt, Willam B. 1993. *Peace Process: American Diplomacy and the Arab-Israeli Conflict since 1967*. Berkeley: University of California Press.

Rabin, Matthew. 1990. "Communication Between Rational Agents." *Journal of Economic Theory* 51(1):144–70.

Regan, Patrick M. 2002. Third-Party Interventions and the Duration of Intrastate Conflicts. *Journal of Conflict Resolution* 46(1):55–73.

Smith, William P. 1985. "Effectiveness of the Biased Mediator." *Negotiation Journal* 1(4):363–72.

Stigler, Andrew L. 2002/03. "A Clear Victor for Airpower: NATO's Empty Threat to Invade Kosovo." *International Security* 27(3):124–57.

Touval, Saadia. 1975. "Biased Intermediaries: Theoretical and Historical Considerations." *Jerusalem Journal of International Relations* 1(1):51–69.

Touval, Saadia. 1982. *The Peace Brokers: Mediators in the Arab-Israeli Conflict, 1948–79*. Princeton: Princeton University Press.

Touval, Saadia. 1985. "The Context of Mediation." *Negotiation Journal* 1(4):373–78.

Touval, Saadia, and I. William Zartman, eds. 1985. *International Mediation in Theory and Practice*. Boulder, CO: Westview Press.

Touval, Saadia, and I. William Zartman. 1989. "Mediation in International Conflicts." In *Mediation Research: The Process and Effectiveness of Third-Party Intervention*, ed. K. Kressel and D. G. Pruitt. Hoboken: Jossey-Bass, pp. 115–37.

Vasquez, John A., James Turner Johnson, Sanford Jaffe, and Linda Stamato, eds. 1995. *Beyond Confrontation: Learning Conflict Resolution in the Post-Cold War Era*. Ann Arbor: University of Michigan Press.

Wall, James A. 1981. "Mediation: An Analysis, Review and Proposed Research." *Journal of Conflict Resolution* 25(1):157–83.

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

Wall, James A., John B. Stark, and Rhetta L. Standifer. 2001. "Mediation: A Current Review and Theory Development." *Journal of Conflict Resolution* 45(3):370–91.

Walter, Barbara F. 2002. *Committing to Peace: The Successful Settlement of Civil Wars*. Princeton: Princeton University Press.

Young, O. R. 1967. *The Intermediaries: Third Parties in International Crises*. Princeton: Princeton University Press.

Zartman, I. William, and Saadia Touval. 1996. "International Mediation in the Post-Cold War Era." In *Managing Global Chaos: Sources of and Responses to International Conflict*, ed. Chester A. Crocker, Fen O. Hampson, and Pamela Aall. Washington, D.C.: United States Institute of Peace Press, pp. 445–61.