Which Side Are You On?:

Bias, Credibility and Mediation

Andrew Kydd

January 2, 2002

Abstract

Mediators are often thought to be more effective if they are unbiased, or have no preferences over the issue in dispute. This paper 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 not to make a large demand because their opponent has high resolve and will fight. An unbiased mediator who is simply interested in minimizing the probability of war 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 a mediator who is effectively "on your side" will be believed if they counsel restraint. Parties will accept mediators who are biased against them if there are social norms which imply that rejecting mediation is a sign of weakness and something to hide, rather than of strength. There is a social benefit to promulgating such norms, because of the loss involved in letting conflicts go unmediated.

In 1982, Argentina invaded the Falkland/Malvinas Islands and Great Britain vowed to retake them. Faced with a potential war between a key NATO ally and an important regional partner, the United States attempted to mediate the dispute. Secretary of State Alexander Haig recounted his perceptions of his role:

> *While my sympathy was with the British, I believed that the most practical expression of that sympathy would be impartial United States mediation in the dispute. The honest broker must, above all, be neutral. (Haig, 1984:266)*

In order to resolve the conflict, Haig attempted to persuade the Argentine Generals that if they did not back down, the British would fight:

> *Even after all I had said to the contrary, I feared that Galtieri and his colleagues were unable to believe that the British would fight. 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 beliefs about British resolve were correct, and therefore his statements to Argentina were truthful, which makes it especially tragic that Argentina did not believe these statements until it was too late.

In early 1999, the conflict between the Serbian government and the Kosovo Albanians spun out of control. In April, Serbia definitively rejected a deal brokered at Rambouillet France by NATO, which had gradually taken on the mission of protecting the Kosovars from

Milosevic's regime. 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 persuaded Milosevic to back down and agree to a deal only slightly better than that which was offered at Rambouillet. The Russian Foreign Minister was quoted as saying, "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]

These two mediation attempts were similar in many respects. In each case, the mediator attempted to persuade one side to moderate their demands because the other side would not accept them without a fight. Haig tried to persuade the Argentines to back down because the British would fight to retain the Falklands, Chernomyrdin tried to persuade Milosevic to accept the deal on offer because NATO would invade if necessary to reverse the ethnic cleansing of Kosovo. Both NATO and the British had attempted to signal their resolve, Britain by assembling and dispatching a fleet to the South Atlantic, NATO by bombing Serbia and moving troops to Albania. However, considerable uncertainty remained about whether either Britain or NATO would fight. Prime Minister Margaret Thatcher faced domestic opposition to the war, and President Clinton had famously, many argued unwisely, ruled out a ground invasion at the start of the Kosovo bombing campaign.

Though the situations were similar, mediation succeeded in the Kosovo case and failed in the Argentine case. Haig was unable to persuade the Argentines to give in and they continued to doubt British resolve. In contrast, Chernomyrdin was able to persuade Milosevic that he had got the best deal he could. Why was mediation successful in one case and not the other?

[1]Newsweek 7/26/99. See also Daalder and O'Hanlon 2000 pp. 203-205.

The answer has to do with the incentives for a mediator to be truthful in the bargaining context.

In bargaining situations, conflict should be avoided if both sides have complete information over their relative power and the preferences of the other side (Fearon 1995). A possibility of conflict arises if there is uncertainty about the resolve of the parties. If one side thinks the other is likely to have low resolve, it will make a larger demand that it would if it thought the other was resolute, thereby raising the chance of war. A mediator can reduce the likelihood of conflict by providing information on this score, bringing the game closer to a situation of complete information. For instance, in the Falklands case, Argentina presumably would not have attacked if they knew that the British would fight rather than acquiesce in a military takeover of the islands. Their uncertainty about British resolve led them to take a chance on invading. In his efforts at mediation, Haig attempted to provide Argentina with information about British resolve, indicating that they would fight rather than acquiesce, yet the Argentine Junta apparently did not listen.

The reason for Argentina's skepticism about Haig's message can be found in Haig's earlier admission that he was at best neutral, and in fact somewhat biased in favor of the British. For a mediator to be effective, the parties must believe that the mediator is telling the truth if the mediator counsels one side not to make a large demand because their opponent has high resolve and will fight. However, a mediator who is biased in favor of one's opponent will have a strong incentive to make such statements whether they are true or not. After all, if the US preferred a solution involving an Argentine withdrawl from the islands, and believed that Argentina would withdraw if it thought Britain would fight, the US would have every incentive to tell Argentina that Britian would fight, even if the US thought Britain was

unlikely to do so. Thus a mediator who is biased against you will not be believed when they say that your adversary has high resolve, and that you should therefore lower your demands.

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. A 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 their demands are lowered. If the mediator is believed, each side will lower their demands, 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 accept a large demand. 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, therefore Russia was a credible mediator to Milosevic. If Russia said NATO would invade, Serbia could believe it. To be credible, at least when making statements about resolve, a mediator must be biased.[2]

While a mediator biased in favor of Serbia is obviously acceptible to Serbia, it might seem that such a mediator would be rejected by NATO. Why accept a mediator biased against one? In fact, as the case indicates, the reverse is true. NATO sought a mediator biased against itself precisely because only such a mediator could credibly communicate NATO's

[2]See Calvert 1985 and Myers 1998 for related models of biased advice.

resolve to Serbia. Even irresolute parties will accept mediators biased against them if there is a social norm in place that stigmatizes those who reject mediators as being weak actors who are attempting to cover up their lack of resolve. The pacifying effects of mediation provide a rationale for fostering such a social norm.

In this article I develop these arguments with a game theoretic model of mediation. The article has five 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 and acceptable. Third, I consider the social benefits of mediation, that is, under what conditions mediation is beneficial to society, Fourth, I draw out the empirical implications of the model. Finally I consider how the analysis might be extended.

1 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 et al. 2001, and for collections of essays, Touval and Zartman 1985, Mitchell and Webb 1988, Kressel and Pruitt 1989, Bercovitch and Rubin 1992, Vasquez, et al. 1995 and Bercovitch 1996). 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). Oran Young argued 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, Secretary of State Haig's statement above on the proper role of the mediator is by no means unique. Some, however, have questioned the importance of impartiality (Touval 1982; Bercovitch and Houston 1996:26; Carnevale and Arad 1996). Saadia Touval, for instance, argues of mediators that, "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).[3]

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). Neutral mediators are often thought

[3]Thomas Princen addresses this split in the literature by positing two distinct types of mediator, the "neutral" (weak and impartial), and the "principal" (strong and interested or biased) (Princen 1992:18). See Smith (1985) for a similar distinction between "traditional" and "international" mediation, and Touval (1985) for a response.

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 grasp on what exactly mediators are doing in the bargaining process. This problem can best be remedied by incorporating mediators directly into bargaining models of conflict.

This article focuses on the informational and communication role of mediators and treats mediation as an instance of what economists call cheap talk[4]. 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 mobilizing 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 would be a mediator telling one side that a certain demand it is making is so high that the other side is likely to reject it. In itself, the communication is not costly to the mediator, but if it has the effect of causing the party to lower its demands, it may 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.

[4]The active bargaining role of powerful mediators also deserves attention. However, I would argue that information provision is mediation in its purest and least understood form, and hence is most in need of study.

[5]For a good introduction see Farell 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, Farell 1993, Blume and Sobel 1995, Austen-Smith and Banks 2000.

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 truthtelling is an equilibrium.

In bargaining situations, where the interests of the parties are more directly opposed, cheap talk is more problematic because of credibility issues. 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.[6] One party cannot credibly just tell another that it has high resolve and will fight if its demand is not met, since they face an incentive to say this even if it is not true (Fearon 1995).

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 if its demand is not met, regardless of whether this is true. The mediator will then convey this to the other side, the other side will then

[6] An exception to this logic is analysed by Farell and Gibbons 1989b.

believe that it faces a high resolve type and will acquiesce in the demand. 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 not solve the credibility problem of 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

[7]For related issues, see Myerson 1979, 1991, Banks and Calvert 1992, Baliga, Corchon and Sjöström 1997).

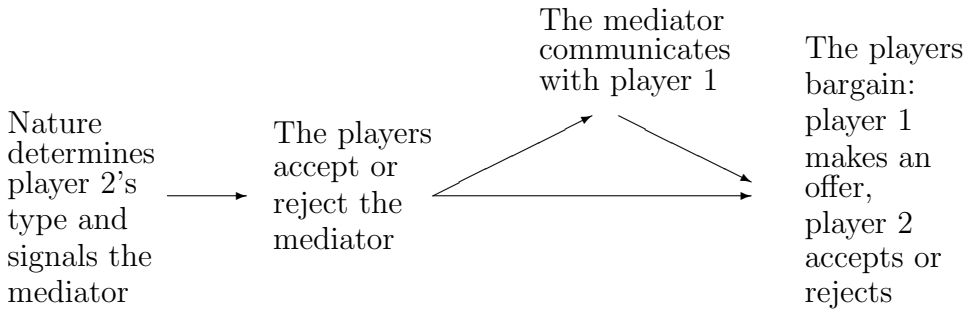
aligned with both of them at once.[8] However if the mediator were completely indifferent to the issue in dispute, the mediator would seem to be credible to both sides. This would seem to justify the focus on impartiality in the mediation literature, an impartial mediator can be trusted by both sides because she has no interest in favoring or harming either side. Intuition further suggests that each party would reject a mediator who was known to be biased against them, so that the selection process would tend to produce unbiased mediators. Why accept a mediator who is biased against one? In fact, as outlined in the introduction, these intuitons do not turn out to be correct. To see why, it is necessary to consider the problem formally.

2 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 war, and dissatisfied means prefering war 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 war, the satisfied power may wish to make some concession to the other side, hoping to buy

[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. I leave this for future work.

Figure 1: The Model in Timeline Form

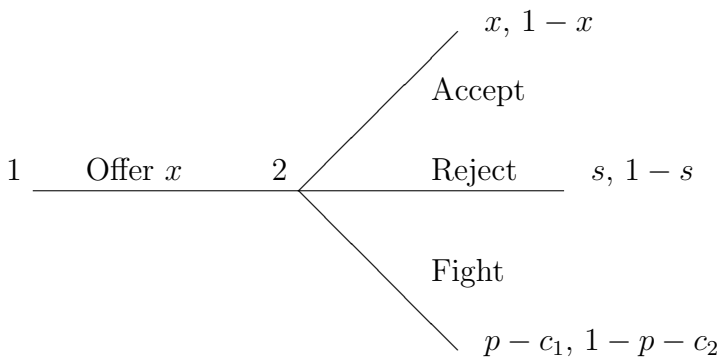


them off. The danger is that the concession will not be big enough, and that the other side will decide to attack rather than accept it or live with the status quo.

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 model is illustrated in timeline form in Figure 1. First, Nature determines player 2's type and the information that the mediator has about player 2's level of resolve. Next, the players decide whether to accept the mediator. I assume that either player can veto a mediation effort. The players make this decision with full knowledge of the mediator's preferences, but do not know what information the mediator has. If the mediator is accepted, she can then communicate with player 1 about player 2's level of resolve. Finally, player 1 and player 2 bargain together.

The bargaining game is illustrated in Figure 2. Player 1 makes an offer $x \in [0 \dots 1]$. Player 2 can accept the offer, reject it but do nothing, or attack. If player 2 accepts the offer, player 1 gets x and player 2 gets $1 - x$. If player 2 rejects the deal but does not attack,

Figure 2: The Bargaining Game



the status quo payoffs prevail, s and $1-s$. If player 2 decides to fight, player 1's war payoff is $p-c_1$, while player 2 receives $1-p-c_2$ where p is player 1's chance of winning the war and c_i is player i's cost of fighting. The mediator also has a preference function over the issue space mx, and pays a cost if there is a war c_m. The mediator's payoff for war is therefore $mp-c_m$. If $m>0$ 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 $m<0$, the mediator is biased in favor of player 2, and if $m=0$, the mediator is unbiased. Note if the mediator is unbiased, her payoff for any accepted offer is 0 and her payoff for war is $-c_m$, so the mediator is indifferent over the possible settlements and just wants to secure a deal.[9]

[9]Thus the unbiased mediator here corresponds to Princen's "neutral" mediator, Princen 1992:22.

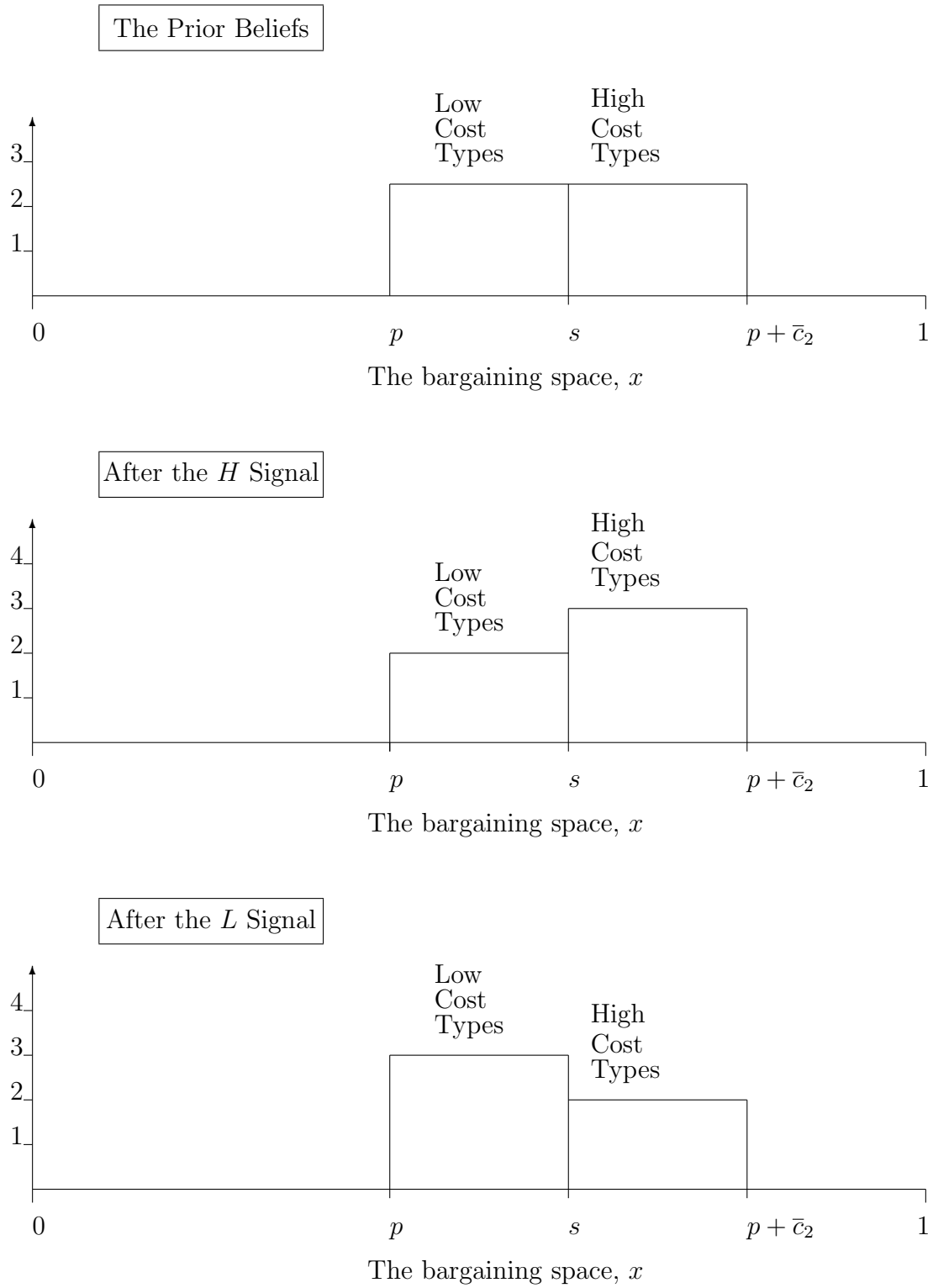
2.1 Beliefs

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 $c_2 \in [0 \ldots \bar{c}_2]$. Thus player 2's reservation value in the bargining game, or the minimum offer it will accept rather than fight, ranges from p to $p + \bar{c}_2$ as ilustrated in Figure 3. The horizontal axis is 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 war, p, is less than the status quo, s, hence player 1 prefers the status quo to war or is satisfied, while player 2 may not be, depending on player 2's cost of fighting. Types with $c_2 \in [0 \ldots s - p)$ are dissatisfied or have "low" costs for fighting, types with $c_2 \in [s - p \ldots \bar{c}_2]$ are satisfied or have "high" costs for fighting. The prior belief that player 2 has high costs is denoted h^0.

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

Figure 3: The Bargaining Space and Distribution of Reservation Values



that she received the message H are illustrated in the middle of Figure 3. Her confidence that player 2 has high costs has increased, and her belief that player 2 has low costs has declined. Her new belief that player 2 has high costs is denoted $h(H)$.

Finally, if the mediator receives the L signal, her beliefs are as shown in the bottom of Figure 3. 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^0, which is in turn greater than the posterior belief after receiving the low signal, $h(L)$, that is, $h(H) > h^0 > h(L)$. In Figure 3, the parameter values are $p = 0.4$, $s = 0.6$, $\overline{c}_2 = 0.4$, and $\epsilon = 0.4$ so that $h^0 = 0.5$, $h(H) = 0.6$, and $h(L) = 0.4$.

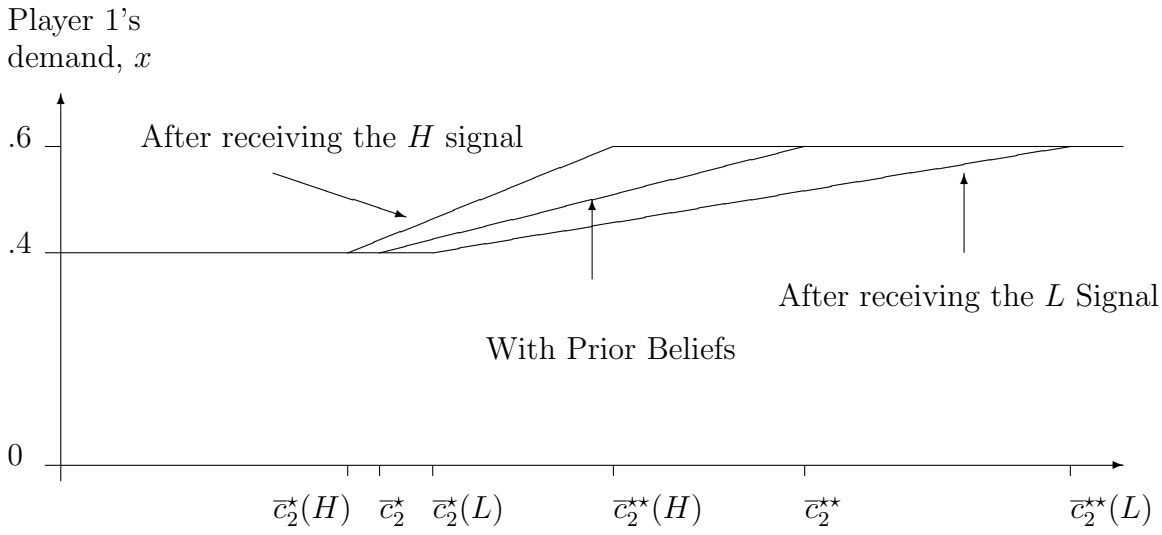
I assume that the parties are not informed about what signal the mediator received, and hence do not know for certain what the mediator's beliefs are. Thus after Nature moves, player 2 is informed of her type, player 1 is not, the mediator has some additional information that she can share with player 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 war. The likelihood of war will in turn influence the mediator's incentive to be truthfull.

2.2 The Equilibrium Offer

If player 1 knew player 2's type, she would make no concession to a high cost type (offer $x = s$), and just buy off a low cost type, (offer $x = p + c_2$). In neither case would war occur.

Figure 4: The Equilibrium Offer



The prior likelihood that player 2 has high costs, $\overline{c}_2$

Thus with complete information, bargaining is successful, war 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 4. The horizontal axis is the upper bound on the distribution of player 2's cost for fighting, $\overline{c}_2$. Thus this dimension can be thought of as a measure of how likely player 2 is to have high costs, ex-ante. The greater $\overline{c}_2$ is, the more likely player 2 is to have high costs, and be unwilling to fight. The smaller $\overline{c}_2$ 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, x.

For very low values of $\overline{c}_2$, 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 $x = p$ or 0.4 in the example used so far. For very large values of $\overline{c}_2$, 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, demanding the status quo remain in place, $x = s$ or 0.6. In between is a region in which the demand depends on player 1's beliefs. In this region, the demand increases steadily from p to s as $\overline{c}_2$ increases. The more convinced player 1 is that player 2 has high costs, the larger demand she makes.

The boundaries of this middle region depend on player 1's beliefs as well. If player 1 is operating with her prior beliefs, because the mediator has been rejected or is untrustworthy, the boundaries are denoted $\overline{c}_2^{\star}$, and $\overline{c}_2^{\star\star}$. With the parameter values used above, the lower bound is $\overline{c}_2^{\star} = 0.3$ and the upper bound is $\overline{c}_2^{\star\star} = 0.7$. Below $\overline{c}_2^{\star}$, player 1 only demands $x = 0.4$, between $\overline{c}_2^{\star}$ and $\overline{c}_2^{\star\star}$ the demand increases, and above $\overline{c}_2^{\star\star}$ player 1 makes no concession at all, demanding the status quo, $x = 0.6$.

Now consider the case where player 1 has received the H signal from the mediator and believed it. Since player 1 has become more convinced that player 2 has high costs, and will not fight, she is willing to make a larger demand in equilibrium. This has the effect of shifting the curve to the left. Player 1 stops making a full concession earlier ($\overline{c}_2^{\star}(H) < \overline{c}_2^{\star}$), and her demand increases more quickly, until player 1 is so confident that she is facing a weak type that she makes no concession at all, when $\overline{c}_2 > \overline{c}_2^{\star\star}(H)$. With the parameter values used above, the lower bound is $\overline{c}_2^{\star}(H) = 0.26$ and the upper bound is $\overline{c}_2^{\star\star}(H) = 0.53$.

Finally, on the right is the case where player 1 has received the L signal and believed it. Player 1 is now more convinced that player 2 has low costs and will fight. This makes player 1 more reluctant to make a big demand, 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 complete concession, and must be very convinced before demanding the status quo. With the param-

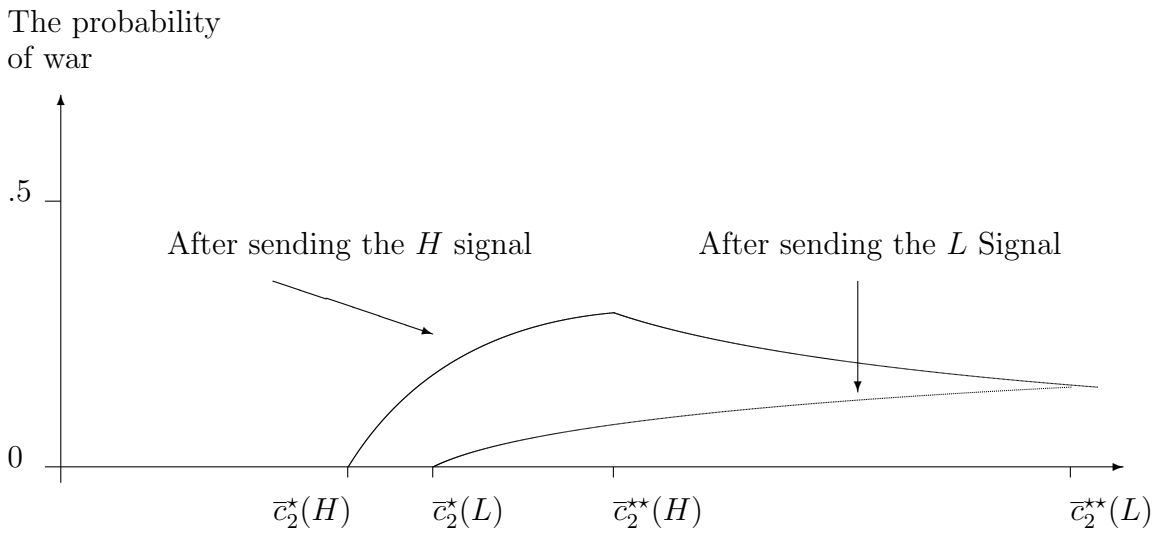
eter values used above, the lower bound in this case is $\overline{c}_2^{\star}(L) = 0.35$ and the upper bound is $\overline{c}_2^{\star\star}(L) = 0.95$.

The bottom line is that if player 1 receives the H message from the mediator and believes it, she will make a bigger demand of player 2 than she would if she received the L message and believed it, if it affects her behavior at all. It is possible that player 1 has such strong prior beliefs that no information from the mediator could change her course of action. She could be so convinced that player 2 is tough that she will make a full concession even if the mediator tells her otherwise. Or she could be so convinced that player 2 is weak that she will make no concession even if the mediator tells her he is resolved. For all middling beliefs, however, where player 1 cares about the mediator's information, if the mediator tells her that player 2 is likely to have high costs, she will make a higher demand than she would have had the mediator told her the opposite. To tell player 1 that player 2 is likely to have high costs, therefore, is to encourage her to raise her demands. To tell her that player 2 has low costs is to encourage her to lower her demands. This fact has ramifications for the likelihood of war.

2.3 The Likelihood of War

If the mediator is credible, player 1 will condition her demand on what the mediator says. The size of her demand, in turn, will influence the probability that negotiations will fail and that war will break out. Essentially, the bigger the demand, the higher the probability of war. The bigger the demand, the more low cost types of player 2 that will reject it and decide to attack. For the mediator, therefore, to tell player 1 that player 2 has high costs is

Figure 5: The Likelihood of War (Mediator's perspective)



The prior likelihood that player 2 has high costs, $\overline{c}_2$

to increase the likelihood of war.

This relationship is illustrated in Figure 5. The horizontal axis is the same as in Figure 4, the upper bound on the prior distribution of player 2's costs, $\overline{c}_2$. The vertical axis is the probability of war. 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 higher demand, or lie and say she got the L signal, causing her to make a lower one.

The top curve, starting at $\overline{c}_2^{\star}(H)$, is the probability of war if the mediator truthfully sends the H signal. It rises with player 1's demand, until it reaches a maximum at $\overline{c}_2^{\star\star}(H)$ where player 1's demand also maximizes at s. From then on it declines as $\overline{c}_2$ increases, reflecting that player 2 is growing increasingly unlikely to have low costs and be willing to fight.

The bottom curve, starting at $\overline{c}_2^{\star}(L)$, is the probability of war 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 $\overline{c}_2^{\star\star}(L)$.

The obvious fact revealed in the figure is that if the mediator thinks player 2 has high costs, being truthfull will increase the likelihood of war, whereas lying will reduce it. For a peace loving mediator, this presents a troubling incentive to lie. It can easily be shown in a figure similar to Figure 5 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 war. In the end, regardless of what the mediator actually believes, telling player 1 that player 2 has high costs increases the chance of war, telling her that player 2 has low costs will reduce it. This has crucial implications for the mediator's incentives to be truthfull.

2.4 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. That is, if she receives the H signal, she has an incentive to report this to player 1 rather than say she got the L signal, and if she got the L signal, she prefers to pass that on rather than pretend she got the H signal. These two conditions lead to constraints on the degree of bias the mediator can have if she is to be credible. Recall that the mediator's bias is measured by m, if m is positive, the mediator is biased in favor of player 1, if m is negative, the mediator is biased in favor of player 2, and if m is zero, the mediator is unbiased.

If the mediator receives the L signal, she must prefer to tell player 1 this, rather than lie and send the H signal instead. Being truthful in this case causes player 1 to make the smaller demand, lying will increase the demand that player 1 makes, increasing the likelihood

of war. Given that the mediator suffers a cost if war occurs, this is something the mediator would prefer to avoid. This condition basically puts an upper bound on how biased in favor of player 1 the mediator can be, an upper bound on m denoted $m^{\star\star}$. An unbiased mediator would be happy to be truthful in this case, since being truthful leads to a lower likelihood of war. A mediator would only wish to lie in this case if she really wanted to see player 1 make a large demand despite the increased chance of war, that is if she were very biased in favor of player 1. In fact, the mediator would have to be so biased that she would prefer that player 1 make the big demand even if player 1, with the same information, would prefer to make a smaller 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 biased in favor of player 1 the mediator can be.

If the mediator receives the high costs signal, she faces more of a dilemma. Telling the truth will encourage player 1 to make a bigger demand, which will precipitate a higher chance of war. Lying will convince player 1 to make a smaller demand which is more likely to be accepted. Thus if the mediator is unbiased and simply wants to minimize the chance of war, she will face an insuperable incentive to lie. Lying will make peace more likely, telling the truth will make war more likely. Thus the unbiased mediator faces a serious credibility problem. She has no incentive to tell player 1 that she thinks player 2 is more likely to accept the bigger demand, because this will raise the chance of war. 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 the issue resolution associated with the larger demand to that associated with the smaller one, then she has an incentive to encourage player 1 to make the larger demand if she thinks player 2 is likely to accept it. This policy preference

outweighs the downside of making the bigger demand, the increased chance of war. Thus this constraint places a lower bound on how biased in favor of player 1 the mediator can be, a positive lower bound on m denoted $m^\star$.

The upper and lower bounds on the mediator's bias are illustrated in Figure 6. The horizontal axis is once again the upper bound on the prior distribution of player 2's costs, $\overline{c}_2$. The vertical axis is m, the mediator's bias. The lower curve is the lower bound, $m^\star$, and the upper curve is the upper bound, $m^{\star\star}$.[10] Between the lines, a truthtelling equilibrium is possible, above and below, only babbling equilibria exist. Immediately apparent is the fact that the mediator must be somewhat, but not too, biased in favor of player 1 to have the proper incentives to tell the truth regardless of the signal she has received. A neutral mediator will not be believed, nor will one that is fanatically committed to getting player 1 to make large demands.

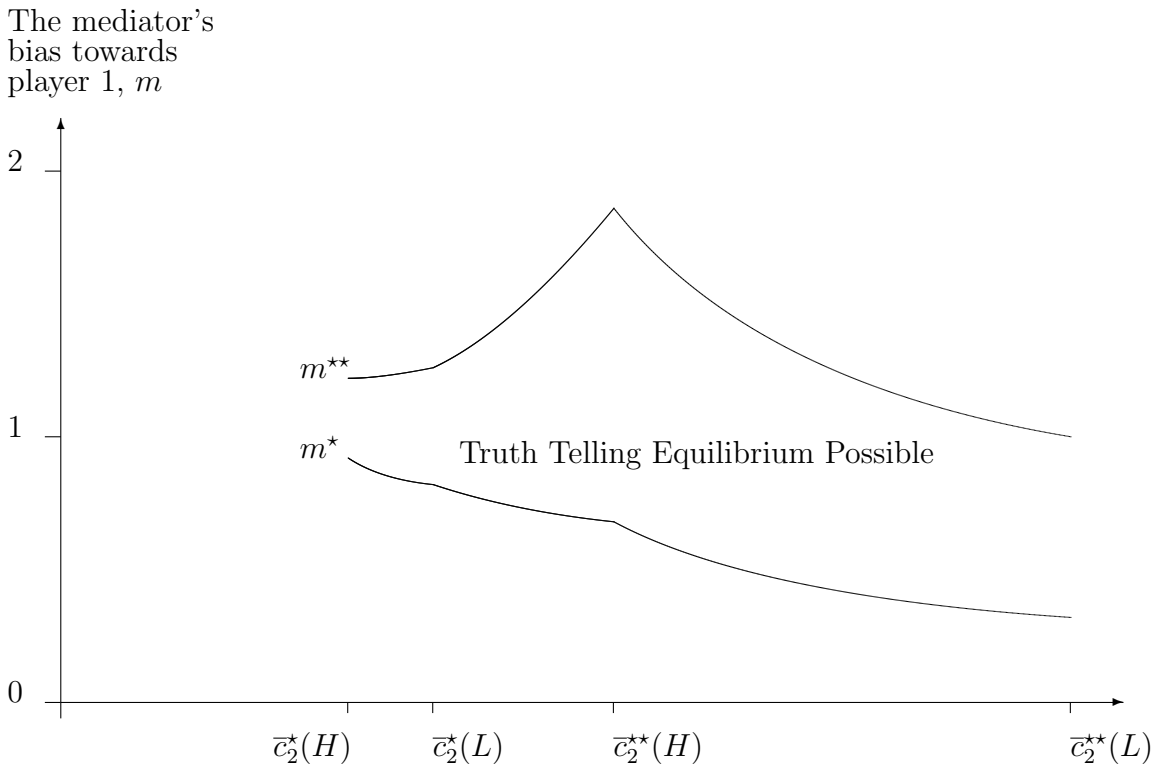
2.5 The Acceptability of the Mediator

The final issue in the solution of the model to consider is when the players will accept the mediator. Player 1 faces no tradeoff here, since the mediator is providing information player 1 cannot possibly lose out by hearing what the mediator has to say. For player 2, the question is more interesting.

Player 2 is in the position of deciding whether to accept a mediator biased against him. A credible mediator will communicate to player 1, giving her an opinion on whether player 2 has high or low costs for fighting. A natural intuition would be that if player 2 has low costs he will welcome such a mediator while if he has high costs he will not. A low cost or

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

Figure 6: The Credibility of the Mediator



The prior likelihood that player 2 has high costs, $\overline{c}_2$

dissatisfied player 2 will think the mediator is likely to tell player 1 that he has low costs, which will lead player 1 to make a lower demand. A high cost player 2 will worry that the mediator will tell player 1 that he has high costs, leading player 1 to make a bigger demand than she would absent mediation. Thus the two categories of player 2 would separate, low cost types accepting and high cost types rejecting mediation.

In fact such a separating equilibrium is not possible for the obvious reason that if high cost types reject mediation while low cost types accept it, the high cost types have revealed their lack of resolve by rejecting the mediator, which will cause player 1 to raise her demand. Since this is what the high cost types were trying to avoid by rejecting the mediator in the first place, they are in a catch 22.

Consider instead a pooling equilibrium in which all types of player 2 accept mediation. If all types of player 2 accept mediation, player 1 learns nothing from the acceptance decision. Her beliefs will change only in reaction to the mediator's communication. Player 2 hopes that the mediator tells player 1 that he has low costs, but realizes that she may tell player 1 that his costs are high. When is such an equilibrium sustainable? The answer depends on what player 1's beliefs will be in the zero likelihood event that player 2 rejects mediation. Since in equilibrium all types of player 2 accept mediation, this move would take the game off the equilibrium path, making it impossible to use Bayes rule to update beliefs. Instead we must make assumptions about what beliefs would be in such a circumstance, assumptions which can be informed by prevailing social expectations or norms about behavior.

Two such possible expectations seem salient. One is that rejecting mediation is a sign of weakness, because it signals the player has something to hide, and the other is that rejecting mediation is a sign of strength, because it signals the player has little interest in conflict

resolution because her war payoff is so high. Which expectation prevails off the equilibrium path is crucial in determining what equilibrium is possible. If the second expectation prevails, so that rejecting mediation when player 2 is expected to accept it convinces player 1 that player 2 has low costs, it destroys the equilibrium in which mediation is accepted. If rejecting mediation convinces the other guy that you are tough, why accept mediation?

Conversely, if the first expectation is prevalent, that unexpectedly rejecting mediation is a sign of weakness and an attempt to prevent the mediator from revealing this weakness, then an equilibrium is sustainable in which all types of player 2 accept mediation. The reason is that while rejecting mediation will cause player 1 to think player 2 has high costs and therefore make a large demand, if mediation is accepted there is at least some chance that the mediator will tell player 1 that player 2 has low costs, leading player 1 to make a smaller demand. This chance, however small, beats the alternative, because rejection leads to a larger demand for sure.

Thus the mediator will be accepted by all types even if it is biased against player 2 provided that the understanding is that anyone who rejects mediation must have high costs, and be attempting to cover that up. This understanding could arise as a general expectation in a social setting, as a social norm. Social norms are often thought to provide expectations that may guide equilibrium behavior in strategic settings (Kreps 1990:143). They are particularly salient in the generation of off equilibrium path beliefs, given the inability of Bayesian updating to provide guidance in this area.

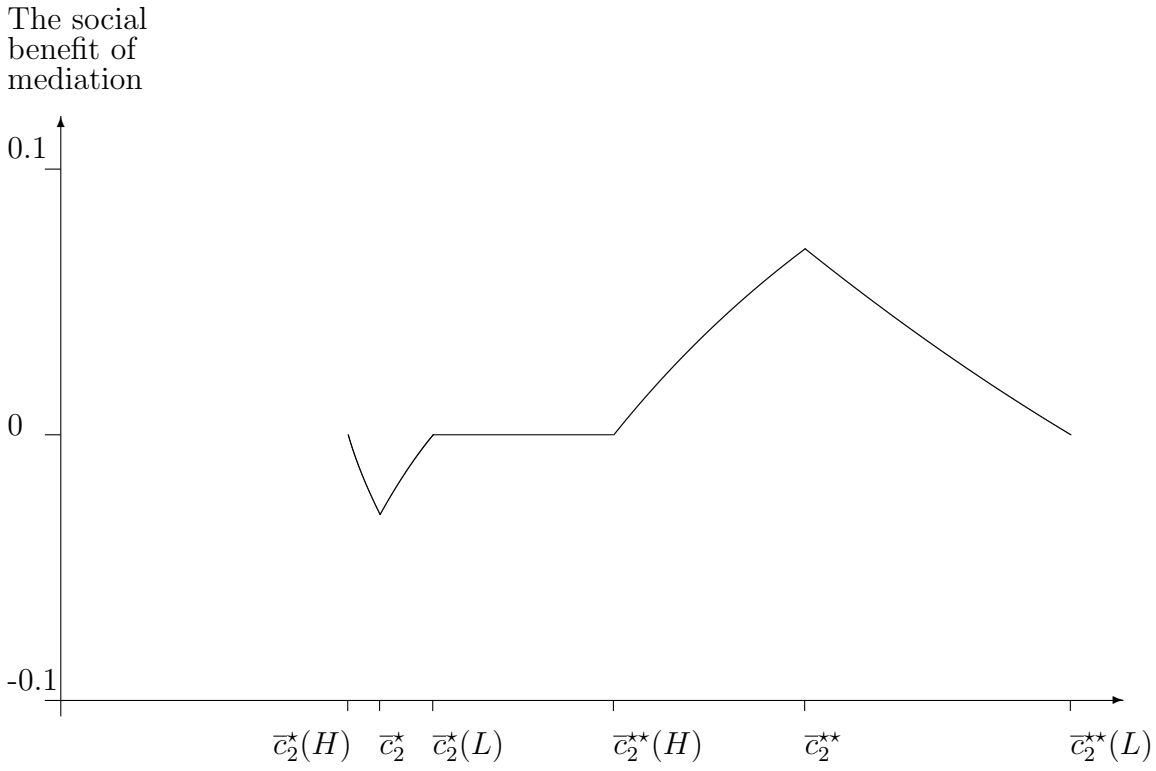
3 The Social Benefit of Mediation

A social norm designed to support mediation only makes sense if mediation produces social benefits. The model can shed some light on when this is the case. As a mesure of the social benefits of mediation we can subtract the likelihood of war with mediation from the likelihood of war without mediation. If war is more likely without mediation, this will be positive, indicating that society benefits from mediation. If war is more likely with mediation, this will be negative, indicating that mediation is socially harmful. I assume that "society" has no interest in the issue at stake, so that social welfare is just a function of the likelihoods of war, not the specific issue resolution attained.

The social benefit of mediation is shown in Figure 7. The horizontal axis is once again the upper bound of the prior distribution of player 2's costs, $\overline{c}_2$. The vertical axis is the difference between the likelihood of war with and without mediation, positive values indicate that mediation is beneficial, negative values indicate that mediation increases the likelihood of war and is harmful.

Mediation is socially harmful for low values of $\overline{c}_2$, between $\overline{c}_2^\star(H)$ and $\overline{c}_2^\star(L)$. In this region, without mediation player 1 will make a large concession, in some cases the largest possible concession, p. With mediation, player 1 will sometimes make a concession this large, but will also sometimes make a larger demand, $x(H)$, having been egged on by an H signal from the mediator. This result indicates that mediation is not harmless and should not be recommended in all cases. The very credibility of the mediator, which assures that if the mediator counsels restraint player 1 will moderate her demands, also assures that if the mediator recommends that player 1 hang tough, player 1 will do so, raising the likelihood of

Figure 7: The Social Benefit of Mediation



The prior likelihood that player 2 has high costs, $\overline{c}_2$

war.

For middling values of $\overline{c}_2$, between $\overline{c}_2^{\star}(L)$ and $\overline{c}_2^{\star\star}(H)$, mediation is socially neutral. In this zone, without mediation player 1 makes a middling demand, with mediation she will make a larger one if she receives the H signal, and a smaller one if she receives the L signal. The effect is to keep the likelihood of war the same, with or without mediation. Once again, mediation is not necessarily socially beneficial.

Finally, for higher values of $\overline{c}_2$, between $\overline{c}_2^{\star\star}(H)$ and $\overline{c}_2^{\star\star}(L)$ mediation is socially beneficial. Here, without mediation, player 1 makes a large and possibly maximal demand, at or near s. With mediation, if player 1 gets the H signal, she will still make a big demand, but if she gets the L signal, she will make a lower demand. The net result is to make the likelihood of war lower with mediation that without it. Here mediation is socially useful, in that it encourages moderation in demands and reduces the likelihood of war.

Given that the result differs in the three different regions, can anything general be said about the social benefit of mediation? A case can be made that mediation is more likely to arise in the second and third zones, where it is neutral or beneficial, than in the first zone where it is harmful. The first zone is where player 1 is so pessimistic about her opponent's toughness that she is prepared to make a large concession even without mediation. In these circumstances, conflict resolution is not as much of a problem and international mediators may not even have a chance to get involved. The second and third zones correspond to more intractable disputes, where player 1 is more convinced of the weakness of her opponent, and hence is making little or no concessions. Here, in the face of player 1's intransigence, is where conflicts will last a long time and become subject to international concern and mediation attempts. In these cases, mediation will be neutral or positively beneficial in reducing the

likelihood of war. Thus on balance I would argue that mediation will be beneficial where it is most likely to be applied.

4 Empirical Implications

The model shows that the effect of mediation is to reduce the likelihood of war and encourage moderation in demands. 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.[11]

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

Hypothesis 1 *Mediators who attempt to persuade one side to lower its demands because the other side will reject them must be biased in favor of the side they are communicating with in order to be successful.*

In addition, the model implies that biased mediators succeed by getting the side towards which they are biased to make concessions, or not make large demands. This suggests the following hypothesis.

Hypothesis 2 *Within the sample of successful mediation efforts, in the cases in which the mediator is biased towards one of the parties, that party will moderate its demands in the negotiation in comparison with what the average party does in cases in which the mediator is unbiased.*

[11]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, ie. Bercovitch and Langley 1996, Bercovitch and Houston 1993.

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 limit their demands. Alliances, that is, act as instruments of control as well as signaling devices towards 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).

An 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 towards 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 judgement 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 aquire new information, yet not so much as to come to share their preferences, and hence lose their usefulness.

5 Extending the Analysis

One possible way to extend the framework would be to consider more complicated bargaining games, involving two sided incomplete information, alternating offers, or indefinitely long

bargaining sessions, for instance. While these factors might seem important, I conjecture that the same basic result would hold in more complicated models so long as the uncertainty takes the form of private information about resolve. For instance, assume in a more complicated bargaining model there is a set of messages m_i and for each message an associated probability p_i of war if that message is sent to the parties and believed. If the mediator's payoffs were simply equal to the likelihood of peace, the mediator would again face the temptation to select that message m_i which is associated with the lowest probability of war.

One rationale that might support the truthfulness of unbiased mediators is reputational considerations. If the mediator's statements were subject to some kind of verification later on and mediators had an incentive to preserve a reputation for truthtelling then perhaps they could be induced to tell the truth even if it cut against their short term interests in a particular case by raising the likelihood of war (see Kim 1996 for a related model). A model in which mediators receive a monetary payoff for their services and face a market in which they can aquire a reputation for truthfulness or the reverse might exibit this property, for then mediators could be punished for lying by simply not being hired again in future contingencies. This kind of model would have certain interesting complexities to deal with, however. For instance, it might be difficult to verify the statements of the mediators in the repeated game situation. For instance, if war is avoided because a low demand is made, it is not clear if it would be possible to say whether the target really had low costs for war or not, since the definitive test–making a higher demand and seeing if they accept it–will not have been carried out. This logic might limit the ability of the parties to verify the performance of the mediator.

There may also be a role for unbiased mediators where the information problem takes

the form of shared uncertainty about environmental factors, rather than private information about resolve. In the literature on congressional signaling, for instance, committees acquire expertise in specific issue areas and are rewarded by grants of agenda setting power (Krehbiel 1991). The resulting system benefits the body as a whole because of the reduction in uncertainty about the effects of policies. In the mediation context, an "expert" mediator could provide information about the likely effects of agreements that could be of use to both parties. In this context, unbiasedness might not prevent truthful communication, and might be selected for by the parties.

6 Conclusion

Mediation can indeed be useful in bargaining situations. By providing information about the resolve of the parties, mediators can reduce the likelihood of war. 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 war 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 war. If she cannot be trusted to send messages that increase the likelihood of war, 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 of one of the parties in the negotiations will be credible in this context. Furthermore biased mediators can be acceptable to the parties against whom they are biased, particularly if there is a social norm that those who reject mediation are attempting to cover up weakness, rather than acting from strength. The attractiveness of fostering such a norm stems from

the social benefit produced by a practice of mediation in the form of a reduced likelihood of war.

A Appendix

The mediator's belief after receiving the signal that player 2 has high costs is $h(H) = \frac{(1-\epsilon)h^0}{(1-\epsilon)h^0+\epsilon(1-h^0)}$, substituting in $h^0 = \frac{p+\bar{c}_2-s}{\bar{c}_2}$ this becomes $h(H) = \frac{(1-\epsilon)(p+\bar{c}_2-s)}{(1-\epsilon)(p+\bar{c}_2-s)+\epsilon(s-p)}$. After receiving the signal that player 2 has low costs, the mediator's beliefs are $h(L) = \frac{\epsilon h^0}{\epsilon h^0+(1-\epsilon)(1-h^0)}$ or $h(L) = \frac{\epsilon(p+\bar{c}_2-s)}{\epsilon(p+\bar{c}_2-s)+(1-\epsilon)(s-p)}$.

The solution concept I employ is perfect Bayesian equilibrium. Off the equilibrium path I assume that rejecting mediation is taken as evidence that player 2 has high costs. I solve the game from the back forwards.

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

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, the payoff for all offers $x > s$ is the same, because they will be rejected by all types of player 2, leading to war or the status quo depending on whether player 2 has low or high costs. Thus the interesting offers are between p and s.

First consider the case where player 1 operates on the basis of prior beliefs. The payoff for offering $x \in [p \dots s]$ is the likelihood of acceptance times the value of the offer, plus the likelihood of rejection times the value of war, $\left(h^0 + (s-x)\frac{1-h^0}{s-p}\right)x + \left(1 - \left(h^0 + (s-x)\frac{1-h^0}{s-p}\right)\right)(p - c_1)$. 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 $h^0 s + (1-h^0)(p-c_1)$ 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, $h^0 + \frac{1-h^0}{s-p}s - \frac{1-h^0}{s-p}2x + \frac{1-h^0}{s-p}(p-c_1) = 0$

and solve for x $x^0 = \frac{\frac{h^0}{1-h^0}(s-p)+s+p-c_1}{2}$ or $x^0 = p + \frac{\bar{c}_2 - c_1}{2}$.

By comparing this with the payoffs for no concession and a full concession, we can determine two cutoff values for $\bar{c}_2$. For $\bar{c}_2 < \bar{c}_2^\star$ player 1 makes the full concession, $x = p$. For $\bar{c}_2 \in (\bar{c}_2^\star \ldots \bar{c}_2^{\star\star})$, player 1 makes the interior demand, and for $\bar{c}_2 > \bar{c}_2^{\star\star}$ player 1 demands the status quo where, $\bar{c}_2^\star \equiv c_1$ and $\bar{c}_2^{\star\star} \equiv 2(s-p) + c_1$.

The likelihood of war for a maximal concession, $x = p$, is zero. As $\bar{c}_2$ increases and the demand increases, the likelihood of war as a function of the demand is $\left(1 - \left(h^0 + (s-x)\frac{1-h^0}{s-p}\right)\right)$. Substituting in the equilibrium demand, we can find the likelihood of war as a function of player 2's maximum cost $\left(1 - \left(h^0 + (s - (p + \frac{\bar{c}_2 - c_1}{2}))\frac{1-h^0}{s-p}\right)\right)$ which reduces to $\frac{1}{2} - \frac{c_1}{2}\frac{1}{\bar{c}_2}$. Once the demand maxes out at s, the likelihood of war is just $1 - h^0$ or $\frac{s-p}{\bar{c}_2}$.

Now consider the case where player 1 has received and believed the H signal. The payoff for an offer x is $\left(h(H) + (s-x)\frac{1-h(H)}{s-p}\right)x + \left(1 - \left(h(H) + (s-x)\frac{1-h(H)}{s-p}\right)\right)(p - c_1)$. 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 $h(H) > h^0$, is better than in the previous case.

The interior solution is found by taking the derivative and setting 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, $x(H) = \frac{\frac{h(H)}{1-h(H)}(s-p)+s+p-c_1}{2}$ or $x(H) = \frac{\frac{1-\epsilon}{\epsilon}(p+\bar{c}_2-s)+s+p-c_1}{2}$. 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 demand increases).

The boundaries for the interior demand in this case are $\bar{c}_2^\star(H) \equiv s - p + \frac{\epsilon}{1-\epsilon}(p - s + c_1)$, and $\bar{c}_2^{\star\star}(H) \equiv s - p + \frac{\epsilon}{1-\epsilon}(s - p + c_1)$. 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 demands is $\left(1-\left(h(H)+(s-x(H))\frac{1-h(H)}{s-p}\right)\right)$ or $\frac{1}{2}-\frac{c_1}{2}\frac{\epsilon}{(1-\epsilon)(p+\bar{c}_2-s)+\epsilon(s-p)}$. If the signal is inaccurate this reduces to the previous case, and as the signal gets more accurate, the likelihood of war increases, which is a result of the greater demand 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 $\left(1-\left(h(H)+(s-x(L))\frac{1-h(H)}{s-p}\right)\right)$ or $\frac{1}{2}\frac{\frac{\epsilon^2}{1-\epsilon}(p+\bar{c}_2-s)+\epsilon(s-p-c_1)}{(1-\epsilon)(p+\bar{c}_2-s)+\epsilon(s-p)}$. These two probabilities of war are illustrated in Figure 5.

Now consider the case where player 1 has received and believed the L signal. Here the payoff for making offer x is $\left(h(L)+(s-x)\frac{1-h(L)}{s-p}\right)x+\left(1-\left(h(L)+(s-x)\frac{1-h(L)}{s-p}\right)\right)(p-c_1)$. 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, $x(L)=\frac{\frac{h(L)}{1-h(L)}(s-p)+s+p-c_1}{2}$ or $x(L)=\frac{\frac{\epsilon}{1-\epsilon}(p+\bar{c}_2-s)+s+p-c_1}{2}$. The boundaries for the interior demand in this case are $\overline{c}_2^{\star}(L)\equiv s-p+\frac{1-\epsilon}{\epsilon}(p-s+c_1)$, and $\overline{c}_2^{\star\star}(L)\equiv s-p+\frac{1-\epsilon}{\epsilon}(s-p+c_1)$.

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

For $\bar{c}_2<\overline{c}_2^{\star}(H)$ even if player 1 has beliefs $h(H)$, she will make the minimum demand, 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 mp regardless of her communication. Since the mediator is indifferent between telling the truth and lying, both truth telling and babbling equilibria will exist.

For $\bar{c}_2\in[\overline{c}_2^{\star}(H)\ldots\min\{\overline{c}_2^{\star}(L),\overline{c}_2^{\star\star}(H)\}]$, if player 1 receives the L signal and believes it,

she will make the minimal demand, p, which will be accepted for sure, while if she receives the H signal and believes it, she will make an interior demand, $x(H)$ 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 demand $x(H)$ which will yeild a payoff for the mediator of $\left(h(H)+(s-x(H))\frac{1-h(H)}{s-p}\right)mx(H)+\left(1-\left(h(H)+(s-x(H))\frac{1-h(H)}{s-p}\right)\right)(mp-c_m)$ while lying will convince player 1 to demand only p, leading to a payoff of mp for the mediator. Telling the truth beats lying if $m > m^\star \equiv \frac{c_m}{\frac{1}{2}\frac{1-\epsilon}{\epsilon}(p+\overline{c}_2-s)+\frac{1}{2}(s-p+c_1)}$

Now consider the case in which the mediator receives the L signal. To convey this to player 1 and be believed would yeild a payoff of mp, since war is avoided for sure. To lie and send the H message, would cause player 1 to demand $x(H)$, giving a payoff for the mediator of $\left(h(L)+(s-x(H))\frac{1-h(L)}{s-p}\right)mx(H)+\left(1-\left(h(L)+(s-x(H))\frac{1-h(L)}{s-p}\right)\right)(mp-c_m)$. Thus telling the truth beats lying if $m < m^{\star\star} \equiv \frac{c_m}{(\frac{\epsilon}{1-\epsilon}-\frac{1}{2}\frac{1-\epsilon}{\epsilon})(p+\overline{c}_2-s)+\frac{1}{2}(s-p+c_1)}$. It can be shown that for $\epsilon \in \{0, 0.5\}$, that is for informative signals, $0 < m^\star < m^{\star\star}$. For a truthtelling equilibrium to exist, the mediator must be biased in favor of player 1, but not too biased, $m \in [m^\star \ldots m^{\star\star}]$.

For $\overline{c}_2 \in [\overline{c}_2^\star(L) \ldots \overline{c}_2^{\star\star}(H)]$, if player 1 receives the L signal and believes it she will demand $x(L)$, 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 $\left(h(H)+(s-x(H))\frac{1-h(H)}{s-p}\right)mx(H)+\left(1-\left(h(H)+(s-x(H))\frac{1-h(H)}{s-p}\right)\right)(mp-c_m)$. If the mediator lies and sends the L signal, the payoff will be $\left(h(H)+(s-x(L))\frac{1-h(H)}{s-p}\right)mx(L)+\left(1-\left(h(H)+(s-x(L))\frac{1-h(H)}{s-p}\right)\right)(mp-c_m)$. Telling the truth beats lying if $m > m^\star \equiv$

$\frac{c_m}{\frac{1}{2}(\frac{1-\epsilon}{\epsilon}-\frac{\epsilon}{1-\epsilon})(p+\overline{c}_2-s)+c_1}$. If the mediator receives the L signal, and conveys this to player 1, the payoff to the mediator is $\left(h(L)+(s-x(L))\frac{1-h(L)}{s-p}\right)mx(L)+\left(1-\left(h(L)+(s-x(L))\frac{1-h(L)}{s-p}\right)\right)(mp-c_m)$. Lying by sending the H signal instead produces a payoff of: $\left(h(L)+(s-x(H))\frac{1-h(L)}{s-p}\right)mx(H)+\left(1-\left(h(L)+(s-x(H))\frac{1-h(L)}{s-p}\right)\right)(mp-c_m)$. Telling the truth beats lying if $m < m^{\star\star} \equiv \frac{c_m}{\frac{1}{2}(\frac{\epsilon}{1-\epsilon}-\frac{1-\epsilon}{\epsilon})(p+\overline{c}_2-s)+c_1}$.

For $\overline{c}_2 \in [\overline{c}_2^{\star\star}(H)\ldots\overline{c}_2^{\star}(L)]$, if player 1 believes the H signal, she will demand the status quo, s, and if she receives the L signal she will only demand p. If the mediator receives the H signal and conveys this to player 1, the payoff is $h(H)ms+(1-h(H))(mp-c_m)$. If the mediator lies instead, the payoff is just mp because war is avoided. Telling the truth beats lying if $m > m^{\star} \equiv \frac{c_m}{\frac{1-\epsilon}{\epsilon}(p+\overline{c}_2-s)}$. If the mediator receives the L signal and conveys it to player 1, the payoff is mp. If the mediator lies instead, the payoff is $h(L)ms+(1-h(L))(mp-c_m)$. Telling the truth beats lying if $m < m^{\star\star} \equiv \frac{c_m}{\frac{\epsilon}{1-\epsilon}(p+\overline{c}_2-s)}$.

For $\overline{c}_2 \in [\max\{\overline{c}_2^{\star}(L),\overline{c}_2^{\star\star}(H)\}\ldots\overline{c}_2^{\star\star}(L)]$, if player 1 believes the L signal, she will demand $x(L)$, and if she believes the H signal, she will make the maximal demand, s. If the mediator receives the H signal and credibly communicates it to player 1 her payoff is $h(H)ms+(1-h(H))(mp-c_m)$. If she lies and sends the L signal instead, the payoff will be $\left(h(H)+(s-x(L))\frac{1-h(H)}{s-p}\right)mx(L)+\left(1-\left(h(H)+(s-x(L))\frac{1-h(H)}{s-p}\right)\right)(mp-c_m)$. Telling the truth beats lying if $m > m^{\star} \equiv \frac{c_m}{(\frac{1-\epsilon}{\epsilon}-\frac{1}{2}\frac{\epsilon}{1-\epsilon})(p+\overline{c}_2-s)+\frac{1}{2}(p-s+c_1)}$. If the mediator receives the L signal and credibly communicates this to player 1, the mediator's payoff is $\left(h(L)+(s-x(L))\frac{1-h(L)}{s-p}\right)mx(L)+\left(1-\left(h(L)+(s-x(L))\frac{1-h(L)}{s-p}\right)\right)(mp-c_m)$ whereas lying by sending the H signal would yield $h(L)ms+(1-h(L))(mp-c_m)$. Telling the truth beats lying if $m < m^{\star\star} \equiv \frac{c_m}{\frac{1}{2}\frac{\epsilon}{1-\epsilon}(p+\overline{c}_2-s)+\frac{1}{2}(p-s+c_1)}$.

For $\overline{c}_2 > \overline{c}_2^{\star\star}(L)$, regardless of player 1's beliefs, she will make the maximal demand, s,

which will produce a payoff for the mediator of $hm(s) + (1-h)(pm(1) - c_m)$. Since the mediator's communication cannot affect her payoff, there will be truthtelling and babbling equilibria.

Finally, the social benefit of mediation is conceived of as the likelihood of war without mediation minus the likelihood of war with mediation. For $\bar{c}_2 \in [\bar{c}_2^{\star}(H) \ldots \bar{c}_2^{\star}]$ without mediation a full concession is made so there is no chance of war. The social benefit of mediation is therefore $0 - \frac{x(H)-p}{\bar{c}_2}\epsilon$ or $\frac{(1-2\epsilon)(s-p)-(1-\epsilon)\bar{c}_2+\epsilon c_1}{2\bar{c}_2}$. Evaluated at $\bar{c}_2 = \bar{c}_2^{\star}(H)$ this equals zero, and at $\bar{c}_2 = \bar{c}_2^{\star}$ it equals -0.03. For $\bar{c}_2 \in [\bar{c}_2^{\star} \ldots \bar{c}_2^{\star}(L)]$ the social benefit is $\frac{x^0-p}{\bar{c}_2} - \frac{x(H)-p}{\bar{c}_2}\epsilon$ or $\frac{(1-2\epsilon)(s-p)+\epsilon\bar{c}_2-(1-\epsilon)c_1}{2\bar{c}_2}$. Evaluated at $\bar{c}_2 = \bar{c}_2^{\star}$ this is -0.03, and at $\bar{c}_2 = \bar{c}_2^{\star}(L)$ it equals zero. For $\bar{c}_2 \in [\bar{c}_2^{\star}(L) \ldots \bar{c}_2^{\star\star}(H)]$ the social benefit is $\frac{x^0-p}{\bar{c}_2} - \frac{x(L)-p}{\bar{c}_2} - \frac{x(H)-x(L)}{\bar{c}_2}\epsilon$ which reduces to zero in the entire interval. For $\bar{c}_2 \in [\bar{c}_2^{\star\star}(H) \ldots \bar{c}_2^{\star\star}]$ the social benefit is $\frac{x^0-p}{\bar{c}_2} - \frac{x(L)-p}{\bar{c}_2} - \frac{s-x(L)}{\bar{c}_2}\epsilon$ or $\frac{p-s+(1-\epsilon)\bar{c}_2-\epsilon c_1}{2\bar{c}_2}$ which is zero at $\bar{c}_2 = \bar{c}_2^{\star\star}(H)$ and at $\bar{c}_2 = \bar{c}_2^{\star\star}$ equals 0.07. Finally, for $\bar{c}_2 \in [\bar{c}_2^{\star\star} \ldots \bar{c}_2^{\star\star}(L)]$ the social benefit is $1 - h^0 - \frac{x(L)-p}{\bar{c}_2} - \frac{s-x(L)}{\bar{c}_2}\epsilon$ or $\frac{s-p-\epsilon\bar{c}_2+(1-\epsilon)c_1}{2\bar{c}_2}$ which is 0.07 at $\bar{c}_2 = \bar{c}_2^{\star\star}$ and zero at $\bar{c}_2 = \bar{c}_2^{\star\star}(L)$.

References

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

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

Banks, Jeffrey S. and Randall L. Calvert. 1992. "A Battle of the Sexes Game with Incomplete Information." *Games and Economic Behavior* 4:347-372.

Bercovitch, Jacob. 1996. *Resolivng 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." *The International Journal of Conflict Management* 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. Bercovitch, Jacob. Boulder: Lynne-Rienner.

Bercovitch, Jacob and Jeffrey Langley. 1993. "The Nature of the Dispute and the Effectiveness of International Mediation." *Journal of Conflict Resolution* 37:670-691.

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.* 65:359-382.

Calvert, Randall. 1985. "The Value of Biased Information: A Rational Choice Model of Political Advice." *Journal of Politics* 47:530-555.

Carnevale, Peter J. and Sharon Arad. 1996. "Bias and Impartiality in International Mediation." *In Resolving International Conflicts: The Theory and Practice of Mediation,* ed. Bercovitch, Jacob. Boulder: Lynne-Rienner.

Crawford, Vincent P. and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica.* 50: 1431-1451.

Daalder, I. H. and M. 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* 50:653-681.

Farell, Joseph. 1993. "Meaning and Credibility in Cheap-Talk Games." *Games and Economic Behavior.* 5:514-531.

Farell, Joseph and Robert Gibbons. 1989a. "Cheap Talk with Two Audiences." *American Economic Review.* 79:1214-1223.

Farell, Joseph and Robert Gibbons. 1989b. "Cheap Talk Can Matter in Bargaining." *Journal of Economic Theory.* 48:221-237.

Farell, Joseph and Matthew Rabin. 1996. "Cheap Talk." *Journal of Economic Perspectives.* 10:103-118.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49:379-414.

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. Vasquez, John A., James Turner Johnson, Sanford Jaffe and Linda Stamato. Ann Arbor: University of Michigan Press.

Gelpi, Christopher. 1999. "Alliances as Instruments of Intra-Allied Control." in *Imperfect Unions: Security Institutions over Time and Space* edited by Helga Haftendorn, Robert O. Keohane, and Celeste Wallander (Oxford: Oxford University Press) pp. 107-139.

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).

Kim, Jeong-Yoo. 1996. "Cheap talk and reputation in repeated pretrial negotiation." *Rand Journal of Economics.* 27:787-802.

Kleiboer, Marieke. 1996. 'Understanding Success and Failure of International Mediation." *Journal of Conflict Resolution* 40:360-389.

Krehbiel, Keith. 1991. *Information and Legislative Organization.* (Ann Arbor: University of Michigan Press).

Kreps, David M. 1990. *Game Theory and Economic Modelling* (Oxford: Oxford University Press).

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

Matthews, Steven A., Masahiro Okuno-Fujiwara, and Andrew Postlewaite. 1991. "Refining Cheap Talk Equilibria." *Journal of Economic Theory.* 55:247-273.

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* 53:267-306.

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

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

Myerson, Roger B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47:61-73.

Myerson, Roger B. 1991. "Analysis of Incentives in Bargaining and Mediation." in *Negotiation Analysis* edited by H. Peyton Young (Ann Arbor: University of Michigan Press).

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

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

Rabin, Matthew. 1990. "Communication between rational agents." *Journal of Economic Theory.* 51:144-70.

Smith, William P. 1985. "Effectiveness of the Biased Mediator" *Negotiation Journal* October.

Touval, S. 1975. "Biased Intermediaries: Theoretical and Historical Considerations." *Jerusalem Journal of International Relations* 1:51-69.

Touval, S. 1982. *The Peace Brokers: Mediators in the Arab-Israeli Conflict, 1948-1979.* Princeton: Princeton University Press.

Touval, Saadia. 1985. "The Context of Mediation" *Negotiation Journal* October.

Touval, S. and I. William Zartman, editors. 1985a. *International Mediation in Theory and Practice.* (Colorado, Westview Press).

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

Wall, James A. 1981. "Mediation: An Analysis, Review and Proposed Research." *Journal of Conflict Resolution* 25:157-80.

Wall, James A., and Ann Lynn. 1993. "Mediation: A Current Review." *Journal of Conflict Resolution* Vol. 37, No. 1 (March) pp. 160-194.

Wall, J. A., J. B. Stark, et al. (2001). "Mediation: A Current Review and Theory Development." *Journal of Conflict Resolution* 45(3): 370-391.

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

Andrew Kydd

Department of Government

Harvard University

1737 Cambridge Ave. Room 612

Cambridge, MA 02138

akydd@cfia.harvard.edu

An earlier version of this paper was presented at the Research Frontiers in International Relations Conference, April 14-16, 2000 La Jolla, CA, at the PIPES Workshop, University of Chicago, April 27, 2000 at the 2000 meetings of the Midwest Political Science Association, April 27-30, Chicago, at the University of California, Davis, May 12, 2000, and at the DVPW in Halle, Germany, Ocober 5, 2000. I would like to thank participants at these occasions, and especially Jasen Castillo, Jörg Faust, Page Fortna, Erik Gartzke, Lisa Martin, Robert Powell, Naunihal Singh, Benjamin Valentino and Barbara Walter for helpful comments.