The Ball is in Your Court: Mediation and Blamecasting

Andrew Kydd*

March 18, 2006

Abstract

A common tactic of mediators in international and civil conflict is to urge each side to make a concession and threaten that if they do not, the mediator will end the negotiations and blame the side that does not make a concession for the failure to reach an agreement. If the mediator is powerless, there must be some other powerful audience before whom the parties wish to appear accommodating, which in turn trusts the mediator's word about which side is being intransigent. I present a model of this tactic and derive conditions under which it can reduce the likelihood of conflict.

*Harvard University, Department of Government. Please send comments to akydd@wcfia.harvard.edu.

When Secretary of State James Baker was attempting to put together a Middle East peace conference in the wake of the first Gulf War, he engaged in a round of shuttle diplomacy to the relevant capitals in the region where he attempted gain every government's agreement to participate. After a making concessions to the Israeli's, he urged them to "give me enough procedural flexibility to make this work," or "put us in a position where together we can leave this dead cat on the Arab doorstep." He noted that "it wasn't the first or last time I'd use this analogy, whose origin escapes me. From the beginning it was the main leverage I had". Later on he threatened Syrian President Assad with isolation if Syria did not participate, writing that he "wanted to make sure he understood the ball was in his court"(Baker III and Defrank 1995, 450,57).

In a more recent example of this kind of tactic, consider the six party talks on the North Korean nuclear program. China mediated between the United States and North Korea. In the endgame, China made a take it or leave it offer and threatened to blame each party for failure. A senior administration official remarked that President Bush agreed "only after China turned over a draft of an agreement and told the Americans they had hours to decide whether to take it or leave it." The *New York Times* reported that over the final weekend of negotiations, "the Chinese increased pressure on the United States to sign - or take responsibility for a breakdown in the talks. 'At one point they told us that we were totally isolated on this and that they would go to the press,' and explain that the United States sank the accord, the senior administration official said."[1]

The threat to leave dead cats on doorsteps or balls in courts seems to be an important

[1]Joseph Kahn and David E. Sanger, "U.S.-Korean Deal on Arms Leaves Key Points Open," *New York Times*, A1, 9/20/05.

tactic in mediation. Baker is not alone in thinking that such threats convey leverage to mediators who can use it to coerce concessions out of the parties. But why should it? What does it mean for one side to be blamed for the failure of negotiations? Why should any side fear such an outcome? Why should a mediator be able to allocate such blame in a way that the parties care about?

I address these questions with a model of third party mediation. There are four actors in the model: two parties engaged in a negotiation, a mediator who makes a proposal and can publicly blame any party who rejects it, and an audience which can then sanction the players. I assume that the mediator is personally powerless, but that the audience is powerful so its actions are of interest to the parties. While the tactic of blame casting is also used by powerful mediators, such as Baker, blame casting is different from the simple direct exertion of leverage. Baker could simply offer rewards or threaten punishments directly, but to threaten parties with blame is to invoke a broader audience who will care about which side is to blame for the failure to agree and will take steps based on that knowledge. In the case of the Middle East negotiations, some audiences of interest are public opinion in America, Europe and the Arab world, and the governments of these regions. If a negotiating party is seen as too intransigent, it may lose popular support, which can lead to a drop of of material support in the form of aid from governments or charitable contributions from ordinary citizens.

The model yields several interesting results. To be effective as a blame caster, the mediator must either not care at all about the issue in dispute, or have similar preferences to the audience, favoring some intermediate solution. The audience must trust the mediator, and it can do so only if the mediator's preferences are similar, or the mediator does not care at all about the outcome. If the mediator is biased in favor of one of the parties, in that it shares the same preference ordering, it will be useless in a blame casting role because it would always want its favored side to be rewarded and the other side to be punished by the audience. Conversely, a biased audience is not necessarily fatal, since the costs of rewarding and punishing players may encourage even a biased audience to punish intransigence from its favored side. Finally, the presence of an audience willing to punish and a mediator with the right preferences to be honest can lower the likelihood of conflict by raising the payoff to agreeing to the mediator's proposal. Thus, blamecasting can indeed be a factor in facilitating conflict resolution.

In what follows, I discuss the literature relevant to the question, describe the model and main results, and conclude with some historical examples of blamecasting in action.

1 Mediation and Blame

International and civil wars are frequently resolved through negotiations that are mediated by third parties. Prominent successes include the Israeli-Egyptian peace agreement negotiated at Camp David in 1979 and the Dayton accord that ended the war in Bosnia in 1995; the Camp David negotiation between Israel and the Palestinians in 2000 was a notable failure (Quandt 1993; Holbrooke 1998; Ross 2005). A large literature by and for practitioners testifies to the importance of mediation in the diplomatic world, and discusses blame casting as one diplomatic tactic (Crocker, Hampson, and Aall 1999; Crocker, Hampson, and Aall 2004, 124). While mediators like Kissinger and Baker note their use of the blame casting threat, they provide little systematic reflection on when it is calculated to work and why.

Mediation has also attracted a great deal of scholarly attention.[2] Unfortunately, little consensus has emerged on how mediation works and under what conditions it will be successful. A number of statistical studies of mediation address the question of what strategies mediators employ and what makes for success (Bercovitch and Houston 1993; Bercovitch and Houston 2000). However, these fail to shed light on the issue of blame casting because the categorization of meditator strategies is too coarse. Three categories of mediator behavior are often employed, communication-facilitation, procedural and directive. Blame casting would seem to fall under the last category, but so would making a proposal, or criticizing a demand as too extreme. A much finer grained analysis is necessary to illuminate the conditions that foster blame casting as a strategy.

Theoretical insight can be gained from the literature on cheap talk (Crawford and Sobel 1982; Farrell and Rabin 1996). In cheap talk models, the sender, the mediator in our case, can send a message to the receiver, the audience. The audience then takes an action that affects both the receiver's and the sender's payoffs. The main results from the cheap talk literature are that for honest communication to be feasible the sender's payoffs must be similar to the receiver's, and the more similar they are, the more information can be transmitted. Similar results will be shown to hold here, but with some differences. For instance, even if the mediator and audience share identical preferences over the issue space the parties are bargaining over, if the mediator has monotonically increasing or decreasing preferences, there will be no possibility for honest communication. This kind of result would not be obvious from simply thinking of the problem non-formally using the cheap talk literature as a guide.

There is a small but growing game theoretic literature on mediation proper. In one set of

[2] For a review see (Wall, Stark, and Standifer 2001).

models, the mediator attempts to provide information about the resolve of the negotiating parties, in order to improve their information about each other and facilitate a peaceful deal (Kydd 2003; Smith and Stam 2003; Rauchhaus 2005; Crescenzi, Kadera, Mitchell, and Thyne 2005). The central result of this literature is that if the mediator does not care about the issue over which the parties are negotiating, but just wants to make an agreement more likely, it will be ineffective. Essentially, a mediator who just wants to promote peace will say whatever maximizes the likelihood of peace, which may not necessarily be the same as what is true. To be honest, the mediator must have preferences over the possible issue resolutions, and in particular, to be honest to one side about the other, must share the first side's preference ordering. We will see below that this result does not extend to the role of blamecasting, mediators with no preferences over the issue space make very good blamecasters.[3]

The most directly relevant paper, however, comes from the literature on bargaining between the Congress and President in the United States. Groseclose and McCarty present a model in which Congress passes a bill which the President must sign or veto (Groseclose and McCarty 2001). The voters then register their approval or disapproval of the President. Under some conditions the President wishes to appear moderate before the voters and so may accept a bill that he would prefer to veto. My model differs from the Groseclose and McCarty set up by including a fourth actor, the mediator, who both makes a proposal and then allocates blame for rejecting it. The relationship between the mediator and the audience

[3]Other models of mediation include (Jarque, Ponsati, and Sakovics 2003; Kydd 2005; Schmidt 2004; Favretto 2005; O'Neill 2003). Foundational approaches to mediated bargaining in economics include (Myerson 1986; Forges 1986).

therefore is central here but absent from the Groseclose and McCarty paper.

Also, Katja Favretto paper, basically like mine without the audience.

A final related literature concerns third party intervention in civil wars. Barbara Walter develops a model of civil war termination in which the parties fear to disarm because of suspicion that the other side will renege on the deal and take advantage of their weakness (Walter 2002). To overcome this problem, third party guarantees that any party that reneges on the deal will be punished sufficiently to make reneging not worth while are helpful. Like the blame casting mediator, the third party intervenor punishes intransigence in an effort to encourage cooperation. A key difference, however, is that the mediator does not itself impose the punishments, it simply identifies the intransigent party and leaves it to the relevant audiences to impose the punishments. As a result, the ability of the mediator to communicate with the audience about the behavior of the players is central to the analysis, but missing from the third party intervention models.[4]

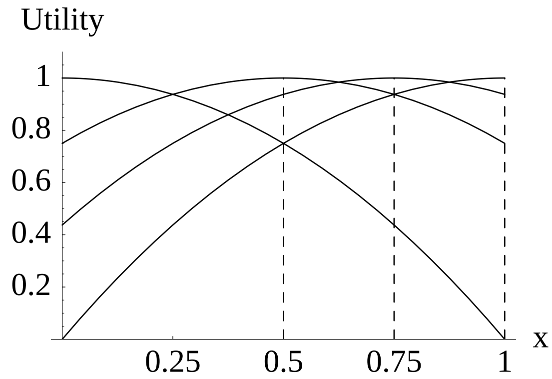
Each of these literatures provides some insight into how a blamecasting mediator could work, and identifies issues to focus on. A key issue will be how the mediator's preferences, in relation to those of the audience and the negotiating parties, dispose it towards honesty or dishonesty in a blame casting role. To fully analyze the problem, a formal model is needed.

2 The Model

There are four players: the two bargaining parties, denoted player 1 and player 2, the mediator and the audience. Players 1 and 2 are bargaining over an issue represented by the

[4]For related models, see (Carment and Rowlands 2003; Werner and Yuen 2005).

Figure 1: The Bargaining Space



unit interval, $[0,1]$, illustrated in Figure 1.[5] Player i's utility for an issue resolution x is $u_i(x)$ and the players are risk neutral or risk averse. I normalize the utility functions such that for player 1, $u_1(0) = 0$ and $u_1(1) = 1$ while for player 2, $u_2(0) = 1$ and $u_2(1) = 0$, and I assume their preferences are monotonic. The mediator and audience have utility functions $u_m(x)$ and $u_a(x)$, and $I_m \in [0,1]$ and $I_a \in [0,1]$ are the mediator's and audience's ideal points. In the figure, the mediator's ideal point is 0.5 and the audience's is 0.75. Notation is summarized in Table 1 in the Appendix.

The mediator, player 1 and player 2 play a bargaining game in which the mediator announces an issue resolution x and the players then simultaneously either accept it, action A_i, or reject it, action R_i. A player's strategy is a map from $[0,1]$ to $\{A_i, R_i\}$, but given that their payoff is monotonic in x, the problem reduces to choosing a reservation value x_i above (below) which player 1 (2) will accept and below (above) which player 1 (2) will reject the deal. If both players accept the proposal then the game ends with the deal implemented with payoffs $u_1(x)$, $u_2(x)$, $u_m(x)$ and $u_a(x)$. There are three ways a deal can

[5] The utility functions illustrated are $u_i(x) = 1 - (x - I_i)^2$, where I_i is player i's ideal point.

be rejected, corresponding to the set $\{\{A_1, R_2\}, \{R_1, R_2\}, \{R_1, A_2\}\}$, which can be ordered $\{A_1, R_2\} \prec \{R_1, R_2\} \prec \{R_1, A_2\}$ so that player 1 becomes more intransigent and player 2 more accomodating. If either or both players reject it, the mediator makes a statement to the audience blaming player 1, player 2, or both parties for the failure to agree, where statement a_i implies that player i accepted, and statement r_i implies that player i rejected the deal. Statement $\{r_1, a_2\}$, for instance, means that player 1 rejected the deal and player 2 accepted. We can denote the mediator's blaming choice $\lambda \in \{\{a_1, r_2\}, \{r_1, r_2\}, \{r_1, a_2\}\}$ where the message space is assumed to have the same ordering as the action space. This means that higher messages are increasingly bad for player 1. The mediator's communication strategy is a map from $\{\{A_1, R_2\}, \{R_1, R_2\}, \{R_1, A_2\}\}$ to $\{\{a_1, r_2\}, \{r_1, r_2\}, \{r_1, a_2\}\}$.

The audience is aware that no agreement was reached, but does not know which player rejected it (independently of what the mediator says). I assume the audience has two options open to it with respect to each player. Posit that each side has a stock of military resources, m_i, which can be devoted to conflict. The audience can *subsidize* a player which will increase its military forces from m_i to $m_i + s_i$, at a cost of k_s to the audience. Or, the audience can *embargo* a player, which will reduce its military power from m_i to $m_i - e_i$, at a cost of k_e to the audience. If we let S_i stand for subsidize, N_i stand for do nothing, and E_i stand for embargo, the audience's punishment strategy can be denoted $P \in \{S_1, N_1, E_1\} \times \{S_2, N_2, E_2\}$ and its strategy is a map from $\{\{a_1, r_2\}, \{r_1, r_2\}, \{r_1, a_2\}\}$ to $\{S_1, N_1, E_1\} \times \{S_2, N_2, E_2\}$, or a function $P(\lambda)$. We can think of the actions having the order $S_i \prec N_i \prec E_i$, so that they increase in punishment for a player, and the combined actions being ordered such that one action is higher than another if, relatively speaking, it hurts player 1 either directly or by helping player 2. For instance, $\{N_1, N_2\} \prec \{E_1, N_2\}$ and $\{S_1, E_2\} \prec \{N_1, N_2\}$ but $\{N_1, N_2\}$

and $\{E_1, E_2\}$ are unranked.[6]

Thus the punishment a player may face could take the form of the loss of an anticipated subsidy, or the imposition of an embargo. Not subsidizing a player saves the audience money, embargoing a player costs money.[7] The audience's actions will shift the expected outcome of the conflict, which is something the audience and mediator may both care about.

After the audience moves, the parties engage in conflict and receive their conflict payoffs. I model conflict as a costly process that leads to some outcome in the bargaining space. We can think of the ratio of military forces, $\frac{m_i}{m_i+m_j}$ as player i's relative power, the higher this is, the more likely the outcome is to be better for player i. Normally, war is modeled as a one time flip of the coin or Bernoulli trial in which with the above probability, player i wins and imposes its most preferred outcome. In reality, however, many different outcomes are possible from war, including negotiated settlements or stalemates that produce intermediate outcomes. A way to model this that retains the simplicity of a one time event is to work with a probability distribution defined over the $[0, 1]$ interval that depends on the military capabilities of both sides. A convenient candidate is the β distribution,

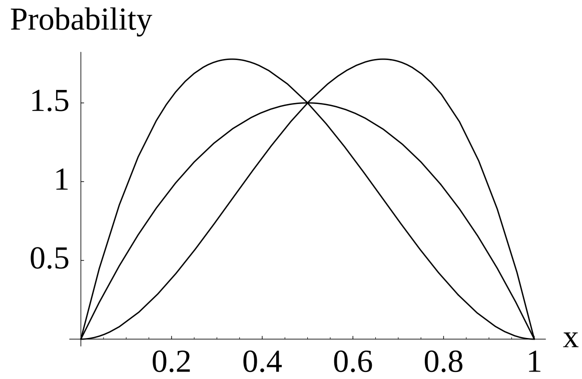
$$\beta(x, m_1, m_2) = \frac{x^{m_1-1}(1-x)^{m_2-1}}{\int_0^1 u^{m_1-1}(1-u)^{m_2-1}du}.$$

The β distribution has several appropriate properties for such a function. First, the mean is equal to the ratio of military forces from player 1's perspective, $\frac{m_1}{m_1+m_2}$. This means that

[6]The audience's action space is a lattice (Topkis 1998, 13).

[7]One could make the strategy set continuous by allowing the audience to provide military power at a certain cost per unit or attempt to interdict military power at a certain cost per unit. Subsidies are probably fairly continuous in this way, while embargoes have a more discrete character, though the amount spent on enforcement can make it more continuous. I start with the discrete case for simplicity.

Figure 2: The Likelihood of Conflict Outcomes



as player 1's forces increase, the mean shifts up, giving greater probability to outcomes that favor player 1, while if player 2's forces increase, the mean decreases, putting more weight on outcomes that player 2 likes. If the forces are balanced, $m_1 = m_2$, then the mean is equal to 0.5. In general, the distribution reacts symmetrically to the military forces, $\beta(x, m_1, m_2) = \beta(1 - x, m_2, m_1)$.

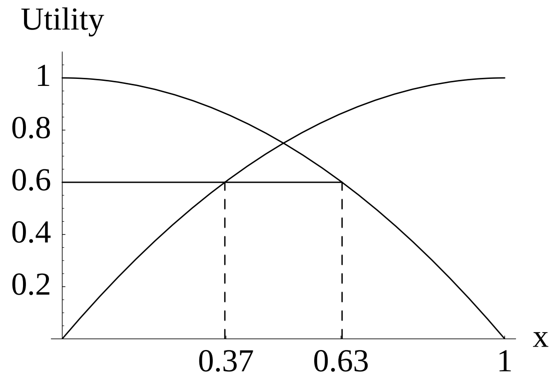
The β distribution is illustrated in Figure 2. The symmetrical curve with a maximum at 0.5 is the case when each side has equal power at $m_1 = m_2 = 2$. The curve with its maximum to the left is the case where m_2 has increased to 3 and m_1 remains at 2. The curve with a maximum to the right is the opposite case where m_2 is equal to 2 and m_1 has increased to 3, so higher outcomes favoring player 1 are more likely.

Player i's cost of conflict is c_i. Player i's war payoff, if the balance of power is known, is therefore

$$w_i(m_1, m_2, c_i) = \int_0^1 u_i(x)\beta(x, m_1, m_2)dx - c_i$$

The war payoff responds to the underlying variables in a natural fashion as described in the following lemma.

Figure 3: Complete Information Bargaining Range



Lemma 1 *The war payoff* $w_i(m_1, m_2, c_i)$ *is increasing in a player's own power and decreasing in its adversary's power and its own costs of conflict.*

Proof: See Appendix. ■

If the balance of power were known, there would exist deals that both sides prefer to war. For instance, Figure 3 illustrates the case where $m_1 = m_2 = 2$ and $c_1 = c_2 = 0.1$. The war payoff for each side is 0.6, which means that player 1 will accept any deal higher than 0.37 and player 2 will accept any deal lower than 0.63, so any deal in the range $[0.37, 0.63]$ will be preferred by both parties to war. If the game featured complete information, therefore, a mediator who wanted to foster an agreement could pick any $x \in [0.37, 0.63]$ and the parties would agree to it without need for punishment by the audience.

In reality, however, there is uncertainty. Specifically, conflict has often been argued to be a function of incomplete information about two factors: the relative power and cost tolerances of the states (Blainey 1988; Fearon 1995). I model uncertainty about both parameters. I assume that a player's military power and costs of conflict are jointly distributed according to the PDF $f_i(m_i, c_i)$ with CDF $F_i(m_i, c_i)$. Costs are assumed to be greater than zero and military power is assumed to be greater than $1+e_i$, where e_i is the amount of weapons that can be denied to player i by an embargo imposed by the audience.[8] Each state's power and costs are private information, unknown both to the other bargainer and to the mediator and audience.

It will be convenient to derive the distribution over outcomes from the audience's and mediator's perspective, taking into account the uncertainty over the parties' relative power. The marginal distribution over player i's power (summing over the costs) is $f_i(m_i) = \int_0^\infty f_i(m_i, c_i)dc_i$. The probability distribution function over outcomes is therefore

$$g(x) = \int_1^\infty \int_1^\infty \beta(x, m_1, m_2) f_1(m_1) f_2(m_2) dm_1 dm_2$$

with associated CDF

$$G(x) = \int_0^x \int_1^\infty \int_1^\infty \beta(x, m_1, m_2) f_1(m_1) f_2(m_2) dm_1 dm_2 dx.$$

3 Equilibrium Behavior

I solve for perfect Bayesian equilibria in pure strategies. In cheap talk games, there are two classes of equilibria, babbling equilibria and truthtelling equlibria (Farrell and Rabin 1996). In babbling equilibria, the mediator's communication is unrelated to the players actual behavior, so the audience does not learn from it. Such equilibria are always possible regardless of the preferences of the actors. The question of interest is when truthtelling equilibria are possible, where in truthtelling equilibria the mediator honestly relays which side(s) rejected the proposal. I further distinguish between trivial and non-trivial truthtelling

[8] This is just to ensure that each side has non-negative military forces even if embargoed.

equilibria. In trivial truthtelling equilibria, the mediator tells the truth but the audience does not condition its behavior on what the mediator says, because the audience has a dominant strategy and pursues it regardless of what the mediator says. If the audience has a dominant strategy, the mediator can tell the truth in equilibrium regardless of its preferences, since its communication has no impact on the game. In a non-trivial truthtelling equilibrium, the mediator tells the truth and the audience conditions its behavior on what the mediator says. The mediator's communication therefore has importance because it affects the outcome. The key question will be under what circumstances there can be a truthtelling equilibrium in which the mediator correctly relays to the audience who rejected the deal, and this induces the audience to punish recalcitrant players, which in turn encourages acceptance of the deal in the first place.

3.1 The Audience's Behavior

If the parties reach agreement the audience does not have a move and its payoff is just $u_a(x)$. If the parties fail to reach agreement the audience can subsidize or embargo either or both parties.

In a truthtelling equilibrium, the audience will know what deal was proposed and who rejected it. Let the audience's posterior beliefs be denoted $f_i^p(m_i, c_i)$, which yields a distribution over outcomes of $g^p(x)$. The audience's conflict payoff is

$$\int_0^1 u_a(x) g^p(x) dx - c_a$$

The first result to note is that if the audience is indifferent between the various issue resolutions, valuing them all at $u_a(x) = 1$, say, then the above payoff reduces to $1 - c_a$. In this case, the audience does not care about the power of the players, only about avoiding war, and conflict is inevitable when the audience has to move. It will therefore offer no subsidies and impose no blockades, because these policies are costly and do not improve the audience's payoff. The audience might wish to threaten punishment if the parties do not agree, in order to encourage a settlement, but such threats would be non-credible because the audience would not have an incentive to carry them out if put to the test. The other players will therefore behave as if there was no audience, hence the following result.

Theorem 1 *If $u_a(x)$ is constant, the audience's equilibrium action is $\{N_1, N_2\}$, regardless of the actions of the players and mediator. The players and mediator will therefore play the game as if the audience did not exist. A trivial truthtelling equilibrium is possible.*

A second result is that if the audience shares one party's orientation towards the issue, it will not embargo that party or subsidize the other party. To do so would involve spending money to make less favorable outcomes more likely. However, such an audience could still play a role in encouraging peace by threatening to not subsidize its favored side or threatening to embargo the unfavored side. Not subsidizing the favored side saves money, though it moves the outcome in the wrong direction, so might be worth while on balance. Similarly, embargoing the opposing side might be worthwhile despite its cost.

Theorem 2 *If $u_a(x)$ is increasing (decreasing) then the audience will not embargo player 1 (2) or subsidize player 2 (1). However, it may fail to subsidize 1 (2) and may embargo 2 (1). The audience's potential equilibrium action space therefore reduces to $\{S_1, N_1\} \times \{N_2, E_2\}$ ($\{N_1, E_1\} \times \{S_2, N_2\}$).*

Proof: The audience's payoff can be written in terms of the war payoff as

$$\int_{1+e_2}^{\infty} \int_{1+e_1}^{\infty} w_a(m_1, m_2, c_a) dm_1, dm_2.$$

By the same logic as Lemma 1, if $u_a(x)$ is increasing, then $w_a(m_1, m_2, c_a)$ is increasing in m_1 and decreasing in m_2. Embargoing player 1 would replace m_1 with $m_1 - e_1$, while subsidizing player 2 would replace m_2 with $m_2 + s_2$, both of which would lower w_a at a cost to the audience. By symmetry, the reverse holds for player 2. ■

If the audience's payoff is neither constant nor monotonic, the audience may wish to subsidize or embargo one or both parties. It is possible that the audience could wish to subsidize both parties or embargo both parties, depending on the magnitudes of the impacts on the power of the two players and the costs of the options.[9] If the costs of the policies are not negligible, however, it is likely that the audience will at most be willing to consider subsidizing one party and embargoing the other, in order to shift the distribution over outcomes in a favorable direction.

To assess the costs and benefits of subsidies and embargoes in this case, we need more information about the audience's posterior beliefs. Since the players' payoffs for war increase in their power and decrease in their cost, one can posit positively sloped functions $c_i^*(m_i)$ such that if $c_i < c_i^*(m_i)$, the player will reject the deal, and if $c_i \geq c_i^*(m_i)$ then the player will

[9]Counterintuitively, the incentive to subsidize both may be greater than the incentive to embargo both because the variance of the β distribution is decreasing in the levels of military power. Because of this a big mutual subsidy could enable an audience to significantly raise the likelihood of its preferred outcome, while a mutual embargo would increase the dispersion, though it would also allow the audience to adjust the mean. Because this result relies on the least intuitively compelling aspect of the β distribution, I do not emphasize it here.

accept it. Given such functions, the audience's posterior beliefs, in a truthtelling equilibrium conditional on being informed that player i rejected the deal, are as follows.

$$f_i^p(m_i, c_i|R_i) = \begin{cases} 0 & \text{if } c_i(m_i) \geq c_i^*(m_i) \\ \frac{f_i(m_i,c_i)}{\int_{1+e_i}^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_i dm_i} & \text{if } c_i(m_i) < c_i^*(m_i) \end{cases}$$

The marginal distribution over the relative power is

$$f_i^p(m_i|R_i) = \frac{\int_0^{c_i^*(m_i)} f_i(m_i, c_i)dc_i}{\int_0^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i, c_i)dc_i dm_i}$$

with a cumulative distribution of

$$F_i^p(m_i|R_i) = \frac{\int_0^{m_i}\int_0^{c_i^*(m_i)} f_i(m_i, c_i)dc_i dm_i}{\int_0^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i, c_i)dc_i dm_i}.$$

The corresponding posterior CDF if player i accepts the deal is

$$F_i^p(m_i|A_i) = \frac{\int_0^{m_i}\int_{c_i^*(m_i)}^{\infty} f_i(m_i, c_i)dc_i dm_i}{\int_0^{\infty}\int_{c_i^*(m_i)}^{\infty} f_i(m_i, c_i)dc_i dm_i}.$$

The key question is, if the audience knows a player rejected the deal, what happens to the audience's beliefs about that player's power, in comparison to what they would have been had the player accepted it? Intuitively, it would seem that the audience should become more convinced that the player is more powerful, given that more powerful types have a greater incentive to reject the deal and go to war. The next lemma indicates that this is correct provided that power and costs are not positively correlated.

Lemma 2 *In a truthtelling equilibrium, if m_i and c_i are not positively correlated in the prior distribution, then if player i rejects the deal it is a sign that it is likely to be more powerful,* $F_i^p(m_i|R_i) < F_i(m_i) < F_i^p(m_i|A_i)$.

Proof: See Appendix. ■

Therefore, in a truthtelling equilibrium, if a side rejects the deal, the audience's posterior beliefs that it is stronger will rise. This should affect the audience's beliefs about how likely the various outcomes of a conflict are. If stronger players reject the deal and weaker players accept it, then if player 1 rejects the deal, player 1 looks stronger and so higher outcomes become more likely than if player 1 had accepted the deal.

Lemma 3 *In a truthtelling equilibrium, if m_i and c_i are not positively correlated in the prior distribution, the posterior CDF over the outcomes is declining in the mediator's message λ, that is, $G^p(x|\{a_1, r_2\}) > G^p(x|\{r_1, r_2\}) > G^p(x|\{r_1, a_2\})$.*

Proof: See Appendix. ■

For a great enough change, this may lower the payoff of an audience with non-monotonic preferences. For instance, consider an audience with strong preferences for a middling outcome, such that $I_a = 0.5$. Posit that the distribution over outcomes $g(x)$ is a spike concentrated around 0.25 if player 1 accepts and player 2 rejects, around 0.5 if both reject, and around 0.75 if player 1 rejects and player 2 accepts. If the deal is to be rejected, the audience would be happiest if both sides rejected it, and would be less happy if only one side rejected it, because it makes that side look stronger which increases the likelihood of extreme outcomes.

Therefore, it could be worthwhile for audience to lower the power of a player who has rejected the deal, or boost power of opposite side if it accepted. This will come at a cost, but will have the effect of shifting $G(x)$ back in in a preferred direction.

Theorem 3 *If the audience believes extreme outcomes favoring one side have become more likely because of a message blaming that side for rejecting the deal, it may wish to punish that side by embargoing it or subsidizing its rival. In particular, equilibria may exist in which* $P(\lambda)$ *is increasing in* λ*.*

Proof: See Appendix. ■

The audience's preferences and beliefs are therefore crucial to its strategy. If the audience does not care about the issue in dispute between the parties, it has no credibly threat to embargo or subsidize either side and so will have no impact on the likelihood of negotiation success. If the audience is biased towards one side it's range of strategies is reduced, since it will not embargo its favored side or subsidize the opponent. However, it could credibly threaten to reduce the subsidy to its favored side or embargo the opponent, and so can still play a role in an equilibrium involving conditional punishment for rejecting the deal. If a side rejects the deal, the audience becomes convinced that side is stronger than expected, and therefore that outcomes favoring it are more likely. It may therefore wish to punish that side, reduce its relative power, and therefore increase the likelihood of more central outcomes.

3.2 The Mediator's Communication Choice

The mediator's payoff, in the case of no agreement, is similar to the audience's in form:

$$\int_0^1 u_m(x) g^p(x) dx - c_m$$

The mediator must consider how its communication will affect this payoff, through the mechanism of the audience's decision to subsidize or embargo the parties. In a babbling equilibrium, the audience will not condition its behavior on the mediator's behavior, so the mediator can randomize over its messages, or say one thing all the time, so long as it's messages are uncorrelated with the parties' behavior. The mediator's preferences are unimportant in a babbling equilibrium because the audience will pursue the same strategy regardless of the mediator's communication.

In a truthtelling equilibrium, the mediator's strategy may depend on its preferences. Truthtelling equilibria will be easy to sustain if the mediator does not care about the issue resolution, that is, if $u_m(x) = 1$ or some other constant. In this case, the mediator's payoff is $1 - c_m$ regardless of its communication to the audience, so the mediator is willing to tell the truth.

Theorem 4 *If $u_m(x)$ is constant, a truthtelling equilibrium is always possible.*

This result contrasts with that obtained in several studies of mediation, in which the mediator is not credible if it does not care about the issue in dispute but only wants to get a settlement (Kydd 2003; Smith and Stam 2003; Rauchhaus 2005). The difference is that in these models, the mediator is attempting to prevent a conflict by offering advice, so the mediator with no policy preferences is tempted to offer whatever advice minimizes the likelihood of conflict, rather than whatever advice is actually true. In this model, by contrast, conflict is inevitable when the mediator has to make its communication, so there is no longer any possibility of making it more or less likely. Given that conflict is inevitable, the only question facing the mediator is what the outcome of the conflict will be. A mediator who does not care what that outcome is can easily tell the truth and let the chips fall where they may.

In contrast, if the mediator's preferences are monotonic, a truthtelling equilibrium is not possible.

Theorem 5 *If $u_m(x)$ is monotonic then a non-trivial truthtelling equilibrium is impossible.*

Proof: Assume $u_m(x)$ is increasing. The mediator wants $G'(x)$ to be as low as possible, so higher values of x are more likely. If the audience would punish player 1 for rejecting the deal, the mediator can only lower its payoff by blaming player 1. If the audience would punish player 2 for rejecting the deal, the mediator can only raise its payoff by blaming player 2. Therefore the mediator will not be honest about either party. ■

One might think that if the mediator and the audience were aligned towards the same player, so that both $u_m(x)$ and $u_a(x)$ were increasing say, a truthtelling equilibrium would be possible. This does not hold because whereas the audience has to spend money to embargo or gets to save money by not subsidizing, for the mediator these policies are costless and revenueless, so the mediator is free to concentrate solely on what maximizes the likelihood of preferred issue resolutions. While an audience might wish to economize by not subsidizing a player 1 who has rejected a deal and therefore shown itself to be more powerful, the mediator would prefer that subsidies be maintained, since the mediator does not pay the cost of the subsidy.

If the mediator's preferences are not constant or monotonic, they must be sufficiently aligned with the audience's for a truthtelling equilibrium to be possible. Essentially, the mediator must want the parties to be dealt with in the way the audience will deal with them on receipt of truthful information.

Theorem 6 *If $u_m(x)$ is close enough to $u_a(x)$ and k_s and k_e are not too high, a truthtelling* *equilibrium will exist.*

The constraint on the costs of the subsidies and embargoes is required because these costs are what differentiate the mediator from the audience. They cannot be too great or the mediator will prefer subsidies and embargoes that the audience finds too costly given the information about the parties power revealed by their actions.

3.3 The Bargainers' Behavior

Rejecting the proposal results in conflict regardless of the other player's decision to accept or reject. However, the audience may sanction rejection, especially if the other side accepts, so the payoff for rejecting must take into account that the audience may intervene conditional on what the other side does. We can write the payoff to player 1 for rejecting as follows

$$\int_1^{\infty}\left(\int_0^{c_2^*(m_2)} w_1(m_1,m_2,c_1)f_2(c_2)dc_2 + \int_{c_2^*(m_2)}^{\infty} w_1(m_1,m_2,c_1)f_2(c_2)dc_2\right) g_2(m_2)dm_2$$

or

$$\int_1^{\infty}\left\{F_2(c_2^*(m_2))w_1(m_1,m_2,c_1) + (1-F_2(c_2^*(m_2)))w_1(m_1,m_2,c_1)\right\} g_2(m_2)dm_2$$

where it is understood that the war payoffs in the first and second term may differ. For instance, if the audience does nothing if both reject, but subsidizes 2 if player 1 rejects and player 2 accepts, then in the second term m_2 is replaced by $m_2 + s_2$, which lowers player 1's payoff. Player 2's payoff is defined analogously.

What is the payoff for accepting the deal? Accepting the deal will only lead to peace if the other side accepts it as well. The payoff for player 1 for accepting the deal can then be written

$$\int_1^{\infty}\left\{F_2(c_2^*(m_2))w_1(m_1,m_2,c_1) + (1-F_2(c_2^*(m_2)))u_1(x)\right\} g_2(m_2)dm_2.$$

Accepting the deal means that war can only result if player 2 rejects the deal. The audience may choose to reward player 1 in this case by subsidizing it (replacing m_1 with $m_1 + s_1$) or embargoing player 2 (replacing m_2 with $m_2 - e_2$) both of which would raise the payoff to player 1 for accepting.

In order to recover the $c_i^*(m_i)$ functions, we equate the payoff for rejection with the payoff for acceptance. I consider two cases. First assume the audience stays out by playing $\{N_1, N_2\}$ regardless of the parties behavior. Equating the payoffs for accepting and rejecting the deal, we get for the two players:

$$\int_1^\infty \{1 - F_2(c_2^*(m_2))\} \{u_1(x) - w_1(m_1, m_2, c_1)\} g_2(m_2) dm_2 = 0 \quad (1)$$

$$\int_1^\infty \{1 - F_1(c_1^*(m_1))\} \{u_2(x) - w_2(m_1, m_2, c_2)\} g_1(m_1) dm_1 = 0. \quad (2)$$

From equation 1, one can recover, for any deal x, $c_1^*(m_1)$, or the cost of conflict for which player 1 is indifferent between accepting the deal and rejecting it, given its relative power and given $c_2^*(m_2)$. This cost could be negative, if the deal is good and player 1's power is low, but I will make the common assumption that war is costly, so that if the solution to equation 1 is negative, then $c_i^*(m_i) = 0$. From Lemma 1, we know that $w_i(m_1, m_2, c_i)$ is increasing in player i's military power, m_i, and decreasing in player i's costs, c_i, which implies that as military power increases, the cost that will keep the player indifferent between accepting and rejecting must also increase, so $c_i^*(m_i)$ will have a positive slope.

An equilibrium in the game consists of a pair of functions, $c_1^*(m_1)$ and $c_2^*(m_2)$, that constitute a fixed point of the system of equations 1 and 2. There may be multiple equilibria. One equilibrium of the game, in fact, has both sides rejecting the deal with certainty, $c_i^*(m_i) = \infty$, which makes each side indifferent between accepting and rejecting, sustaining the equilibrium. However, if we assume that any player which has costs of conflict greater than the greatest possible gains, $c_i > 1$, accepts any deal, which it at least weakly prefers to do, then the above equilibrium is ruled out. The "best" equilibrium in which the likelihood of a deal being accepted is maximized can be obtained by starting with $c_i^*(m_i) = 0$ (assuming that all types accept the deal) and then applying equations 1 and 2 until the first fixed point is reached.

Now consider a symmetrical case in which the audience embargoes any player who unilaterally rejects the deal, pursuing the strategy

$$\begin{aligned} \{a_1, r_2\} &\rightarrow \{N_1, E_2\} \\ \{r_1, r_2\} &\rightarrow \{N_1, N_2\} \\ \{r_1, a_2\} &\rightarrow \{E_1, N_2\}. \end{aligned}$$

The payoff for player 1 for rejecting and accepting are, respectively,

$$\int_1^\infty \{F_2(c_2^*(m_2))w_1(m_1, m_2, c_1) + (1 - F_2(c_2^*(m_2)))w_1(m_1 - e_1, m_2, c_1)\} \, g_2(m_2)dm_2$$

and

$$\int_1^\infty \{F_2(c_2^*(m_2))w_1(m_1, m_2 - e_2, c_1) + (1 - F_2(c_2^*(m_2)))u_1(x)\} \, g_2(m_2)dm_2.$$

The payoff for rejecting has been reduced, and the payoff for accepting has been increased. Therefore we can deduce the following result.

Theorem 7 *In equilibria in which the audience punishes unilateral rejection, so $P(\lambda)$ is strictly increasing, $c_i^*(m_i)$ is less than in equilibria in which it does not, where $P(\lambda)$ is flat. The probability of conflict is therefore lower as well.*

3.3.1 A Numerical Example

For instance, consider the following numerical examle. Assume that military power is distributed according to the Poisson distribution

$$g_i(m_i) = \frac{M_i^{m_i - 2} e^{-M_i}}{(m_i - 2)!}$$

with means $M_1 = 27$ and $M_2 = 22$, so player 1 is more likely to have higher military power than player 2.[10] Assume the costs of conflict are distributed according to the exponential distribution

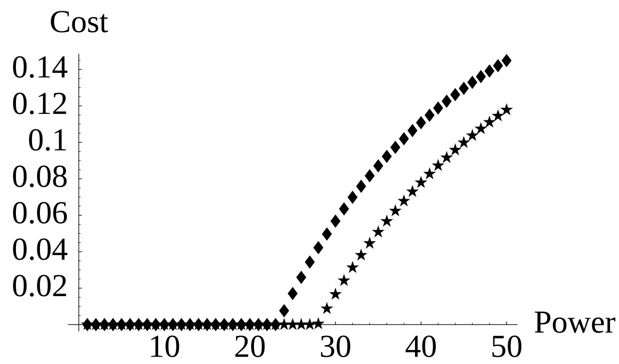
$$f_i(c_i) = \frac{1}{C_i} e^{-\frac{c_i}{C_i}}$$

with means $C_1 = \frac{1}{2}$ and $C_2 = \frac{2}{3}$ so that player 2 is more likely to have higher costs. In this case, the equilibrium values of $c_i^*(m_i)$ are illustrated in Figure 4, where $c_1^*(m_1)$ is depicted with diamonds and $c_2^*(m_2)$ with stars and $x = \frac{1}{2}$. Note, player 1 is willing to reject the deal for lower levels of power and higher costs than player 2 is, because player 2 is expected to be weaker than player 1. For low enough levels of power, neither side will reject the deal regardless of their costs, because they are too weak in comparison to the other side's expected level of power.

There is positive feedback between the likelihood of acceptance on each side. If, for some exogenous reason, player i becomes more likely to accept the deal, $c_i^*(m_i)$ decreases, this means that the expected strength of player i, conditional on player i accepting, has gone up. This means that starting a war against a deal accepting player i becomes less attractive, so player j will be more likely to accept the deal as well, $c_j^*(m_j)$ will go down. This will in

[10] I use the discrete Poisson distribution to simplify computation, and subtract 2 from the military power to reflect the fact that I am assuming the lowest level of power is 2 but the Poisson starts at 0.

Figure 4: Equilibrium values of $c_i^*(m_i)$



turn further reduce $c_i^*(m_i)$ in a positive feedback loop, until a new equilibrium is reached. Thus any exogenous factors that make the deal more attractive or war less attractive for one player will increase the likelihood that both accept the deal (Milgrom and Roberts 1994). The audience punishments potentially imposed are one such factor. If the audience pursues a strategy in which $P(\lambda)$ is increasing, then this will shift the $c_i^*(m_i)$ curves to the right, which further encourages cooperation and further shifts to the right.

The effect of the location of the deal is more complex. As the deal gets better for a player, accepting will become more attractive compared with rejecting it, holding the other side's behavior constant, making the player more likely to accept. Similarly, at a first cut, for the other side, the deal has gotten worse, increasing the incentive to reject the deal. However, the feedback effects are self canceling in this case. For instance, if x increases, Player 1 becomes happier with the deal and more likely to accept, while player 2 becomes unhappy and less likely to accept. However, the fact that player 1 is more likely to accept makes war less attractive for player 2, increasing Player 2's likelihood of accepting, whereas player 2's reluctance to accept makes the deal less attractive for player 1, reducing player 1's likelihood of accepting. Where the process ends up depends on the magnitude of the effects. However, I would conjecture that in real life cases, the direct effects of the attractiveness of the deal would outweigh the feedback effects caused by the other side's behavior.

3.4 The Mediator's Proposal

Any proposal the mediator can make generates some chance of acceptance and some chance of rejection. For a given distribution of power, m_1, m_2, the likelihood a deal is accepted is $(1 - F_1(c_1^*(m_1|x)))(1 - F_2(c_2^*(m_2|x)))$ and the likelihood that it is rejected is one minus this quantity, where the dependence of the c_i^*'s on x is now of interest. If the deal is accepted, the mediator gets $u_m(x)$, if the deal is rejected, the mediator receives its conflict payoff. The mediator's payoff for proposing any deal x is therefore

$$\int_1^\infty \int_1^\infty \{[1 - (1 - F_1(c_1^*(m_1|x)))(1 - F_2(c_2^*(m_2|x)))]w_m(m_1, m_2, c_m) +$$
$$(1 - F_1(c_1^*(m_1|x)))(1 - F_2(c_2^*(m_2|x)))u_m(x)\}\, g_2(m_2)g_1(m_1)dm_2dm_1.$$

The mediator will select the proposal that maximizes this quantity. The task is simplified if the mediator does not care about the issue, in that case it can just select the deal that maximizes the likelihood of acceptance.

Theorem 8 *If the mediator is indifferent over the issue space, $u_m(x) = 1$, then the mediator will select the proposal that maximizes the likelihood of acceptance, x^*. If the mediator is not indifferent and if $I_m \neq x^*$, then the mediator will select some other proposal and the likelihood of conflict will increase.*

This result reinforces the desirabilty of a mediator who does not care about the issue at stake. Such a mediator can always tell the truth in the communication phase, and will always select the issue resolution which maximizes the likelihood of success in the proposal stage. Thus in marked contrast to mediators who are offering advice on the acceptability of proposals (Kydd 2003; Smith and Stam 2003), or trying to build trust (Kydd 2005), here the mediator who does not have a stake in the issue is the best.

4 Discussion

The main results of the model can be summed up as follows. For the audience to have any impact on the game, it must care about the issue resolution reached by the parties. An audience that simply wants to make peace more likely will be ineffective because by the time it moves, a renewed period of conflict is inevitable. The mediator, in contrast, can be indifferent to the issue resolutions and only care about maximizing the likelihood of peace. Such a mediator will select whichever proposal the parties are most likely to agree to, and will then be happy to honestly report to the audience which side rejected the deal. If the mediator is biased towards one side, in that it shares the same preference ordering over the issue as one of the bargainers, it will be useless as a blame casting mediator, even if the audience shares the exact same utility function. The reason is that the mediator does not bear any cost from punishing or rewarding the bargainers, so the mediator will always want to reward its favored side and punish the other side, regardless of who rejects the deal. The audience, however, even if it favors one side, will not hand out subsidies frivolously, but will only want to reward or punish if it is worth the costs. A mediator with a moderate ideal point, somewhere between those of the bargaining parties, can be credible, provided that its preferences are similar enough to the audience's.

The relationship of the mediator to the audience is of central importance. In some cases the audience can be domestic public opinion in the mediator's own country. For instance, in a case almost exactly parallel to Baker's efforts to arrange the Madrid conference after the first Gulf War, in 1973 after the Arab-Israeli war, Secretary of State Henry Kissinger visited the Middle East to arrange a conference in Geneva to negotiate the disengagement of forces. Having secured Sadat's acquiescence, he then turned to the Israelis. During a long haggling session about the wording of a memorandum of understanding, Kissinger burst out:

> I'm trying to bring a sense of reality to this discussion. The mood in America is such that if Israel is increasingly seen as the obstacle to the negotiations and the cause of the oil pressure, you'll have tremendous difficulty. Memorandum or no memorandum (Kissinger 1982, 790).

Kissinger is implicitly threatening to blame Israel before American public opinion for the failure to hold the conference.[11] Even Presidents are not above this tactic. Nixon had earlier urged Israeli participation in a letter to the Prime Minister:

> I want to say to you in all solemnity that if Israel now fails to take a favorable decision to participate in the conference on the basis of the letter that we have worked out, this will not be understood either in the United States or in the world, and I will not be able to justify the support which I have consistently

[11] Later, in negotiation over a disengagement with Syria, Kissinger told the Israelis, "If this negotiation fails, I think we have to conclude that the dominant American role in the peace effort is at an end. . . The nature of the situation is, can the United States effectively back up an Israeli position over an extended period of time? I don't believe we can back up a negotiation that breaks down on issues that the American public doesn't understand . . ." (Kissinger 1982, 1056-7).

> rendered in our mutual interests to your government (Kissinger 1982, 759).

Presidents obviously have direct sources of leverage and a reference to public opinion can be a polite cloak for direct threats, but a democratic leader's dependence on public support makes such references not entirely rhetorical.

Intriguingly, at the very end of the shuttle negotiating the Israeli-Syrian disengagement of forces, Kissinger explicitly disavowed blame casting, saying to the Israelis,

> You've gone a long way. If you decide against it, no one on the American side will feel you were unreasonable, so that shouldn't enter into it. What should enter into it is the basic merits of the consequences of an agreement against the consequences of no agreement . . .(Kissinger 1982, 1102)

While this seems to throw away one of his main sources of leverage, Kissinger made this statement only after being sure that Golda Meir, the Israeli Prime Minister, was committed to the deal, and would fight for it in the cabinet. It also followed many uses of precisely this kind of leverage. It therefore seems that he was trying to make the Israelis feel better about the deal by not making it seem like a product of coercion.

One of the most prominent recent cases of blame being cast by a mediator after a failed negotiation is the Clinton administration decision to blame Yasser Arafat for the failure of the Israeli-Palestinian negotiations in 2000 and 2001. The U.S. played a prominent role in mediating these negotiations, most notably at the Camp David summit in July 2000 and in the Clinton proposal of December 2000. When these negotiations failed, and again after the final negotiations at Taba ended, U.S. officials praised Israeli concessions and condemned Arafat's intransigence (Pressman 2003; Ross 2005). As the second intifada got underway and the newly elected government of Ariel Sharon ended the negotiations and responded harshly, the U.S. supported Israel and launched an initiative to isolate Arafat and remove him from leadership. This blamecasting has been rejected by Palestinians and authors more sympathetic to their side, but the general public perception appears to reflect the Clinton administration line (Swisher 2004).

5 Conclusion

Blamecasting can indeed be an important tool in the international mediator's toolkit. An interesting aspect of it is that it can be done by weak mediators with no special gifts of insight or information. As long as the mediator is believable to the audience, the mediator itself does not have to have powerful capabilities or access to private information. However blame casting is only as strong as the audience before whom blame is allocated. If the available audiences are weak or uninterested or so strongly committed to one side that they will subsidize them in any event, blame for failed negotiations will have no ability to moderate the negotiating positions of the parties. This highlights the dependence of successful mediation on the context of mediation beyond the two parties and the immediate mediation setting. For a mediator to use blame casting effectively, there must be a powerful audience which cares about the issue, and trusts the mediator's judgment about who has rejected the deal.

Table 1: Notation in the Game

$u_i(x)$	Player i's utility over the issue space
I_a, I_m	The audience's and mediator's ideal points
m_i	Player i's military forces
c_i	Player i's cost of conflict
f_i,F_i	Beliefs about player i's military power and costs of conflict
$\beta(x)$,$B(x)$	Beliefs about likelihood of outcomes
$g(x)$,$G(x)$	Beliefs about likelihood of outcomes, with uncertainty over power
w_i	Player i's war payoff
s_i	Increase in i's military forces resulting from subsidy
k_s	Cost to audience of subsidy
e_i	Decrease in i's military forces resulting from embargo
k_e	Cost to audience of embargo
λ	Mediator's communication
P	Audience's punishment

Appendix

Notation in the game is summarized in Table 1.

Proof of Lemma 1

The war payoff $w_i(m_1, m_2, c_i)$ is increasing in a player's own power and decreasing in its adversary's power and its own costs of conflict.

For the players costs, it is obvious that

$$\frac{\partial w_i(m_1, m_2, c_i)}{\partial c_1} = -1$$

so as the costs of war increase it becomes less attractive. For the effect of military power, consider player 1 and how the war payoff reacts to an increase in m_1.

First, consider how the distribution over outcomes reacts to the increase in power. Let $m_1' > m_1$ and consider for what outcome x the likelihoods are equal under the two distributions.

$$\begin{aligned}
\frac{x^{m_1'-1}(1-x)^{m_2-1}}{\int_0^1 u^{m_1'-1}(1-u)^{m_2-1}du} &= \frac{x^{m_1-1}(1-x)^{m_2-1}}{\int_0^1 u^{m_1-1}(1-u)^{m_2-1}du} \\
x^* &= \sqrt[m_1'-m_1]{\frac{\int_0^1 u^{m_1'-1}(1-u)^{m_2-1}du}{\int_0^1 u^{m_1-1}(1-u)^{m_2-1}du}}
\end{aligned}$$

This is a unique solution between zero and 1, below which $\beta(m_1', m_2, c_1) < \beta(m_1, m_2, c_1)$, and above which the reverse is true. Now consider the difference in war payoffs,

$$\begin{aligned}
w_1(m_1', m_2, c_1) &- w_1(m_1, m_2, c_1) \\
\int_0^1 \beta(x|m_1', m_2)u_1(x)dx - c_1 &- \int_0^1 \beta(x|m_1, m_2)u_1(x)dx - c_1
\end{aligned}$$

or

$$\int_0^1 (\beta(x|m_1', m_2) - \beta(x|m_1, m_2))u_1(x)dx$$

or, making use of the value of x where the likelihoods are equal,

$$\int_0^{x^*} (\beta(x|m_1', m_2) - \beta(x|m_1, m_2))u_1(x)dx + \int_{x^*}^1 (\beta(x|m_1', m_2) - \beta(x|m_1', m_2))u_1(x)dx. \quad (3)$$

The first term will be negative and the second positive. Given that β is a probability distribution we know that

$$\begin{aligned}\int_0^1 \beta(x|m_1', m_2)dx &= \int_0^1 \beta(x|m_1, m_2)dx \\ \int_0^{x^*} \beta(x|m_1', m_2)dx + \int_{x^*}^1 \beta(x|m_1', m_2)dx &= \int_0^{x^*} \beta(x|m_1, m_2)dx + \int_{x^*}^1 \beta(x|m_1, m_2)dx\end{aligned}$$

so that

$$\int_0^{x^*} (\beta(x|m_1', m_2) - \beta(x|m_1, m_2))dx + \int_{x^*}^1 (\beta(x|m_1', m_2) - \beta(x|m_1', m_2))dx = 0.$$

If $u_1(x)$ were a constant, therefore, equation 3 would be zero as well. However, we know that every $u_1(x)$ in the second (positive) term is greater than every $u_1(x)$ in the first (negative) term, so we know the overall expression is positive.

Player 1's war payoff therefore increases in m_1. The proof that it decreases in m_2 is parallel, as are those for player 2's payoff.

Proof of Lemma 2

In a truthtelling equilibrium, if m_i and c_i are not positively correlated in the prior distribution, then $F_i^p(m_i|R_i) < F_i(m_i) < F_i^p(m_i|A_i)$.

The posterior CDF in case of rejection is

$$F_i^p(m_i|R_i) = \frac{\int_0^{m_i} \int_0^{c_i^*(m_i)} f_i(m_i, c_i)dc_i dm_i}{\int_0^{\infty} \int_0^{c_i^*(m_i)} f_i(m_i, c_i)dc_i dm_i},$$

the prior CDF is

$$F_i(m_i) = \int_0^{m_i} \int_0^{\infty} f(m_i, c_i)dc_i dm_i,$$

and the posterior CDF in case of acceptance is

$$F_i^p(m_i|A_i) = \frac{\int_0^{m_i} \int_{c_i^*(m_i)}^{\infty} f_i(m_i, c_i)dc_i dm_i}{\int_0^{\infty} \int_{c_i^*(m_i)}^{\infty} f_i(m_i, c_i)dc_i dm_i}.$$

We wish to prove the posterior in case of rejection is smaller than the prior.

$$\int_0^{m_i}\int_0^{\infty} f(m_i,c_i)dc_idm_i \geq \frac{\int_0^{m_i}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i}{\int_0^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i}$$

$$\int_0^{m_i}\int_0^{c_i^*(m_i)} f(m_i,c_i)dc_idm_i + \int_0^{m_i}\int_{c_i^*(m_i)}^{\infty} f(m_i,c_i)dc_idm_i \geq \frac{\int_0^{m_i}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i}{\int_0^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i}$$

$$\int_0^{m_i}\int_{c_i^*(m_i)}^{\infty} f(m_i,c_i)dc_idm_i \geq \frac{1-\int_0^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i}{\int_0^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i}\int_0^{m_i}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i$$

$$\int_0^{m_i}\int_{c_i^*(m_i)}^{\infty} f(m_i,c_i)dc_idm_i \geq \frac{\int_0^{\infty}\int_{c_i^*(m_i)}^{\infty} f_i(m_i,c_i)dc_idm_i}{\int_0^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i}\int_0^{m_i}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i$$

$$\frac{\int_0^{m_i}\int_{c_i^*(m_i)}^{\infty} f(m_i,c_i)dc_idm_i}{\int_0^{m_i}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i} \geq \frac{\int_0^{\infty}\int_{c_i^*(m_i)}^{\infty} f_i(m_i,c_i)dc_idm_i}{\int_0^{\infty}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i} \tag{4}$$

On the right, we have the ratio of the mass above $c_i^*(m_i)$ to the mass below it. On the left we have the same ratio up to the point m_i. We know that the left hand side approaches the right as $m_i \to 1$. As we go from m_i to $m_i + \Delta m$, we add $\Delta m \times \int_{c_i^*(m_i)}^{\infty} f_i(m_i,c_i)dc_i$ to the numerator, and $\Delta m \times \int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_i$ to the denominator. We know that $c_i^*(m_i)$ is increasing with m_i. Therefore, if m_i and c_i are uncorrelated, and the distribution of c_i is fixed for all values of m_i, we know that we are adding more weight to the denominator than to the numerator as we increase m_i.

As for the belief after acceptance, we wish to show

$$\int_0^{m_i}\int_0^{\infty} f(m_i,c_i)dc_idm_i \leq \frac{\int_0^{m_i}\int_{c_i^*(m_i)}^{\infty} f_i(m_i,c_i)dc_idm_i}{\int_0^{\infty}\int_{c_i^*(m_i)}^{\infty} f_i(m_i,c_i)dc_idm_i}$$

$$\int_0^{m_i}\int_0^{c_i^*(m_i)} f(m_i,c_i)dc_idm_i + \int_0^{m_i}\int_{c_i^*(m_i)}^{\infty} f(m_i,c_i)dc_idm_i \leq \frac{\int_0^{m_i}\int_{c_i^*(m_i)}^{\infty} f_i(m_i,c_i)dc_idm_i}{\int_0^{\infty}\int_{c_i^*(m_i)}^{\infty} f_i(m_i,c_i)dc_idm_i}$$

$$\int_0^{m_i}\int_0^{c_i^*(m_i)} f(m_i,c_i)dc_idm_i \le \frac{1-\int_0^\infty\int_{c_i^*(m_i)}^\infty f_i(m_i,c_i)dc_idm_i}{\int_0^\infty\int_{c_i^*(m_i)}^\infty f_i(m_i,c_i)dc_idm_i}\int_0^{m_i}\int_{c_i^*(m_i)}^\infty f_i(m_i,c_i)dc_idm_i$$

$$\int_0^{m_i}\int_0^{c_i^*(m_i)} f(m_i,c_i)dc_idm_i \le \frac{\int_0^\infty\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i}{\int_0^\infty\int_{c_i^*(m_i)}^\infty f_i(m_i,c_i)dc_idm_i}\int_0^{m_i}\int_{c_i^*(m_i)}^\infty f_i(m_i,c_i)dc_idm_i$$

$$\frac{\int_0^\infty\int_{c_i^*(m_i)}^\infty f_i(m_i,c_i)dc_idm_i}{\int_0^\infty\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i} \le \frac{\int_0^{m_i}\int_{c_i^*(m_i)}^\infty f(m_i,c_i)dc_idm_i}{\int_0^{m_i}\int_0^{c_i^*(m_i)} f_i(m_i,c_i)dc_idm_i} \tag{5}$$

which is the same as equation 4.

The effect is stronger if there is negative correlation, since the distribution is shifting down as well. Therefore the left hand side is declining, as it approaches the right. It therefore must be bigger for all m_i. Some amount of positive correlation could be accommodated if the condition that more weight is being added to the denominator is maintained. In fact, more weight does not need to be added for every m_i, so long as more has been added by every m_i. But if there is a strong enough positive correlation or bimodal distribution with one lump for small m_i and below the line and the other for a larger amount and above, then the posterior after rejection could be larger than prior, and one's belief that power is below a certain threshold would increase. For instance, if there is a clump of weaker states with low costs and another clump of stronger states with higher costs, then it could be that the weaker states reject the deal and the stronger ones do not. Then rejection is a sign of weakness. It is possible in this case that the audience would not punish rejection, but might even encourage it.

Proof of Lemma 3

In a truthtelling equilibrium, if m_i and c_i are not positively correlated in the prior distribution, the posterior CDF over the outcomes is declining in the mediator's message λ, that is, $G^p(x|\{a_1, r_2\}) > G^p(x|\{r_1, r_2\}) > G^p(x|\{r_1, a_2\})$.

Consider the first two, (the second two follow by symmetry). We have

$$G^p(x|\{a_1, r_2\}) = \int_0^x \int_0^\infty \int_0^\infty \beta(x, m_1, m_2) f_1^p(m_1|a_1) f_2^p(m_2|r_2) dm_1 dm_2 dx$$

and

$$G^p(x|\{r_1, r_2\}) = \int_0^x \int_0^\infty \int_0^\infty \beta(x, m_1, m_2) f_1^p(m_1|r_1) f_2^p(m_2|r_2) dm_1 dm_2 dx.$$

We know $F_1^p(m_1|a_1) > F_1^p(m_1|r_1)$ and we wish to show that $G^p(x|\{a_1, r_2\}) > G^p(x|\{r_1, r_2\})$. Note, we hold player 2's behavior fixed in the comparison.

This will be true if

$$\int_0^x \int_0^\infty \int_0^\infty \beta(x, m_1, m_2) f_1^p(m_1|a_1) f_2^p(m_2|r_2) dm_1 dm_2 dx >$$
$$\int_0^x \int_0^\infty \int_0^\infty \beta(x, m_1, m_2) f_1^p(m_1|r_1) f_2^p(m_2|r_2) dm_1 dm_2 dx$$

or

$$\int_0^\infty \int_0^\infty B(x, m_1, m_2) f_2^p(m_2|r_2) dm_2 f_1^p(m_1|a_1) dm_1 >$$
$$\int_0^\infty \int_0^\infty B(x, m_1, m_2) f_2^p(m_2|r_2) dm_2 f_1^p(m_1|r_1) dm_1$$

First, let's establish that $B(x, m_1, m_2)$ is decreasing in m_1. The cumulative distribution over outcomes is

$$\begin{aligned} B(x, m_1, m_2) &= \frac{\int_0^x u^{m_1-1}(1-u)^{m_2-1} du}{\int_0^1 u^{m_1-1}(1-u)^{m_2-1} du} \\ &= \frac{1}{1 + \frac{\int_x^1 u^{m_1-1}(1-u)^{m_2-1} du}{\int_0^x u^{m_1-1}(1-u)^{m_2-1} du}} \end{aligned}$$

so if $m_1' > m_1$, then we want to show

$$\frac{1}{1+\frac{\int_x^1 u^{m_1-1}(1-u)^{m_2-1}du}{\int_0^x u^{m_1-1}(1-u)^{m_2-1}du}} > \frac{1}{1+\frac{\int_x^1 u^{m_1'-1}(1-u)^{m_2-1}du}{\int_0^x u^{m_1'-1}(1-u)^{m_2-1}du}}$$
$$\frac{\int_0^x u^{m_1-1}(1-u)^{m_2-1}du}{\int_x^1 u^{m_1-1}(1-u)^{m_2-1}du} > \frac{\int_0^x u^{m_1'-1}(1-u)^{m_2-1}du}{\int_x^1 u^{m_1'-1}(1-u)^{m_2-1}du}$$

which can be rewritten as

$$\int_0^x u^{m_1-1}(1-u)^{m_2-1}du - \frac{\int_x^1 u^{m_1-1}(1-u)^{m_2-1}du}{\int_x^1 u^{m_1'-1}(1-u)^{m_2-1}du}\int_0^x u^{m_1'-1}(1-u)^{m_2-1}du > 0$$

or, combining the two and substituting v for u to keep the integrands straight

$$\int_0^x v^{m_1-1}(1-v)^{m_2-1}\left(1 - v^{m_1'-m_1}\frac{\int_0^1 u^{m_1-1}(1-u)^{m_2-1}du}{\int_0^1 u^{m_1'-1}(1-u)^{m_2-1}du}\right)dv > 0$$
$$\int_0^x v^{m_1-1}(1-v)^{m_2-1}\left(1 - \frac{\int_x^1 v^{m_1'}u^{m_1-1}(1-u)^{m_2-1}du}{\int_x^1 v^{m_1}u^{m_1'-1}(1-u)^{m_2-1}du}\right)dv > 0$$

Since v goes from zero to x and u goes from x to 1, we know that $v < u < 1$. We also know that $m_1' > m_1$, therefore we know that $v^{m_1'}u^{m_1-1} < v^{m_1}u^{m_1'-1}$.[12] This implies that the numerator is smaller than the denominator in the quantity subtracted from 1 in the formula so it is less than 1, so the entire term is positive.

So we know that $B(x, m_1, m_2)$ is declining in m_1. This implies that

$$\int_0^\infty B(x, m_1, m_2) f_2^p(m_2|r_2)dm_2$$

is as well. So the question becomes, does the expectation of a decreasing function decline when the distribution shifts to the right, such that $F_1^p(m_1|a_1) > F_1^p(m_1|r_1)$? There must be an official proof of this, but in the meantime, consider the following logic.

[12] $v < u < 1 \Rightarrow v^{m_1'-m_1} < u^{m_1'-1-(m_1-1)} \Rightarrow \frac{v^{m_1'}}{v^{m_1}} < \frac{u^{m_1'-1}}{u^{m_1-1}} \Rightarrow v^{m_1'}u^{m_1-1} < v^{m_1}u^{m_1'-1}$.

Divide the positive real line up into segments of width x. For brevity, let $h(m_1)$ be the declining function identified above. The expectation is approximated by summing the value of $h(m_1)$ in the segment times the "length" of the segment given by $f_1^p(m_1|a_1)$:

$$\sum_{i=1}^{\infty} h(ix) f_1^p(ix|a_1)$$

Consider a scheme by which we take $f_1^p(m_1|a_1)$ as the norm and "borrow" lengths of probability from $f_1^p(m_1|r_1)$ from higher levels of m_1 to fill up the gap at each segment to match $f_1^p(m_1|a_1)$, but where each bit of $f_1^p(m_1|r_1)$ is "tagged" with its (lower) level of $h(m_1)$. We know that $F_1^p(m_1|a_1) > F_1^p(m_1|r_1)$, so that

$$\sum_{i=1}^{\frac{m_1}{x}} f_1^p(ix|a_1) > \sum_{i=1}^{\frac{m_1}{x}} f_1^p(ix|r_1).$$

This implies that we will always be borrowing from higher levels of m_1 to make up the gap. The total length is the same, for small enough x, since they are probability distributions. As a result, in each segment, we will have a length of $f_1^p(m_1|a_1)$ multiplied by $h(m_1)$ compared to assorted lengths of $f_1^p(m_1|r_1)$, some or all of which are borrowed from higher levels of m_1 and which therefore have lower levels of $h(m_1)$ attached to them. Therefore, the first sum will be bigger than the second. As the interval size shrinks to zero, the quantities converge on the integrals of interest.

Proof of Theorem 3

First, establish that the cumulative distribution over outcomes is increasing in the audience's punishment, that is that $G^p(x|P)$ is increasing in P. Consider two actions, $P \prec P'$. By the definition of the order over the P, $P_1 \prec P_1'$ or $P_2 \succ P_2'$ or both. This implies that player 1's military strength is decreased in the primed case, $m_1 > m_1'$, player 2's is increased, $m_2 < m_2'$, or both. The cumulative distribution over outcomes is

$$G^p(x|P) = \int_0^\infty \int_0^\infty B(x, m_1, m_2) f_1^p(m_1) f_2^p(m_2) dm_1 dm_2.$$

We know that $B(x, m_1, m_2)$ is decreasing in m_1 and increasing in m_2. Therefore $G^p(x|P)$ is also. As a result, $G^p(x|P)$ increases in P.

Also, from Lemma 3, we know that $G^p(x|\lambda)$ is declining in λ. Therefore, increasing the punishment counteracts the effect of an increased λ, so if the audience preferred the situation with the previous $G^p(x)$, then increasing the punishment as λ increases can be an equilibrium.

Would like to prove more clear cut theorem in which in any equilibrium, $P(\lambda)$ is increasing in λ. The audience's payoff is

$$\int_0^1 u_a(x) g^p(x|\lambda, P) dx - c_a$$

First we need to show increasing differences. Assume that $P \prec P'$ and $\lambda \prec \lambda'$, we want to show that

$$\int_0^1 u_a(x) g^p(x|\lambda, P') dx - \int_0^1 u_a(x) g^p(x|\lambda, P) dx < \int_0^1 u_a(x) g^p(x|\lambda', P') dx - \int_0^1 u_a(x) g^p(x|\lambda', P) dx.$$

We know that $u_a(x)$ is convex, which may help establish this.

Then need to show supermodularity.

Then invoke Topkis' theorem.

Leave this for next draft.

References

Baker III, J. A. and T. M. Defrank (1995). *The Politics of Diplomacy*. New York: G. P. Putnam's Sons.

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

Bercovitch, J. and A. Houston (2000). Why do they do it like this? An analysis of the factors influencing mediation behavior in international conflicts. *Journal of Conflict Resolution 44*(2), 170–202.

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

Carment, D. and D. Rowlands (2003). Twisting One Arm: The Effects of Biased Intervenors. *International Peacekeeping 10*(3), 1–24.

Crawford, V. P. and J. Sobel (1982). Strategic Information Transmission. *Econometrica 50*(6), 1431–1451.

Crescenzi, M. J. C., K. M. Kadera, S. M. Mitchell, and C. L. Thyne (2005). A Supply Side Theory of Third Party Conflict Management. *Manuscript*.

Crocker, C. A., F. O. Hampson, and P. Aall (1999). *Herding Cats: Multiparty Mediation in a Complex World*. Washington DC: United States Institute of Peace Press.

Crocker, C. A., F. O. Hampson, and P. Aall (2004). *Taming Intractable Conflicts*. Washington D.C.: United States Institute of Peace.

Farrell, J. and M. Rabin (1996). Cheap Talk. *Journal of Economic Perspectives 10*(3), 103–118.

Favretto, K. (2005). Does Love Make War Fair? Mediator Bias and Coercion in International Mediation. *manuscript UCLA*.

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

Forges, F. (1986). An Approach to Communication Equilibria. *Econometrica 54*(6), 1375–1385.

Groseclose, T. and N. McCarty (2001). The Politics of Blame: Bargaining before an Audience. *American Journal of Political Science 45*(1), 100–119.

Holbrooke, R. C. (1998). *To End a War*. New York: Random House.

Jarque, X., C. Ponsati, and J. Sakovics (2003). Mediation: Incomplete Information Bargaining with Filtered Communication. *Journal of Mathematical Economics 39*(7), 803–830.

Kissinger, H. A. (1982). *Years of Upheaval*. Boston: Little, Brown and Company.

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

Kydd, A. (2005). The Honest Broker: Mistrust and Mediation. *manuscript*.

Milgrom, P. and J. Roberts (1994). Comparing Equilibria. *American Economic Review 84*(3), 441–459.

Myerson, R. (1986). Multistage Games with Communication. *Econometrica 54*(2), 323–358.

O'Neill, B. (2003). Mediating National Honor: Lessons from the Era of Duelling. *Journal of Institutional And Theoretical Economics 159*(1), 229–247.

Pressman, J. (2003). Visions in Collision: What Happened at Camp David and Taba? *International Security 28*(2), 5–43.

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

Rauchhaus, R. W. (2005). Asymmetric Information, Mediation and Conflict Management.

Ross, D. (2005). *The Missing Peace: The Inside Story of the Fight for Middle East Peace*. New York: Farrar, Straus and Giroux.

Schmidt, H. (2004). When (and Why) Do Brokers Have to be Honest? Impartiality and Third Party Support for Peace Implementation after Civil Wars 1945-1999. *manuscript*.

Smith, A. and A. Stam (2003). Mediation and Peacekeeping in a Random Walk Model of Civil and Interstate War. *International Studies Review 5*(4), 115–135.

Swisher, C. E. (2004). *The Truth about Camp David: The Untold Story about the Collapse of the Middle East Peace Process*. New York: Nation Books.

Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton: Princeton University Press.

Wall, J. A., J. B. Stark, and R. L. Standifer (2001). Mediation: A Current Review and Theory Development. *Journal of Conflict Resolution 45*(3), 370–391.

Walter, B. F. (2002). *Committing to Peace: the Successful Settlement of Civil Wars*.

Princeton: Princeton University Press.

Werner, S. and A. Yuen (2005). Making and Keeping Peace. *International Organization 59*(2), 261–292.