When Can Mediators Build Trust?

ANDREW H. KYDD *University of Pennsylvania*

Mediation is one of the most widespread techniques for preventing conflict and promoting cooperation. I examine the conditions under which mediators can facilitate cooperation by building trust between two parties. Assuming mediators have no intrinsic commitment to honesty, they can be credible trust builders in one-round interactions only if they care about the issue at stake, have a moderate ideal point, and do not find conflict to be too costly. A mediator that is solely interested in promoting cooperation, or one that is biased toward one side, will be ineffective. A desire to preserve a reputation for honesty can overcome a limited amount of bias or issue indifference.

Mediation is one of the most important tools for promoting conflict resolution. It is applied in fields as diverse as international and civil war, economic exchange, labor negotiations, and divorce proceedings. Mediation can make the difference between a war that continues to take thousands of lives and a successfully negotiated peace deal, or between a strike that costs thousands of hours of lost production and a contract that puts people back to work. Given the importance of mediation, it is not surprising that scholars have devoted considerable attention to studying what makes mediation work.[1]

Any understanding of how mediation can facilitate cooperation must be built on a diagnosis of what causes conflict in the first place and what mediators can do about it. One widely discussed cause of conflict that seems amenable to mediation is uncertainty or asymmetric information. Blainey (1988, 122) famously argued that wars begin when states disagree about their relative power and end when they agree again. More generally, in models of bargaining, disputants that share complete information can typically reach agreement without delay or conflict because the cost of conflict provides an incentive to agree sooner rather than later (Rubinstein 1982). However, when each side has private information about its own resolve or relative power, it may have an incentive to pretend to be stronger than it is in order to get a better deal, which can prevent a bargain being struck (Fearon 1995). These insights have led to a large literature in economics and international relations on the relationship between bargaining, private information, and conflict.[2] Studies of mediation have begun to be grounded in this theory as well, analyzing how mediators can facilitate cooperation by providing information about the parties' resolve or power (Kydd 2003; Rauchhaus 2005; Smith and Stam 2003).[3]

Another form of uncertainty that can cause conflict is mistrust. Mistrust arises when actors are uncertain about whether the other side prefers to reciprocate cooperation or exploit it (Gambetta 1988; Hardin 2002). Mistrust problems often become salient when actors contemplate the implementation of a deal that is being negotiated. Unlike in the bargaining stage, in which actors are uncertain how far they can push one another to get a better deal, in the implementation stage actors are uncertain whether the other side will abide by the deal, or whether it will exploit any vulnerability that may arise in the implementation of the deal. If two sides mistrust each other they can engage in conflict because they believe that the other side will take advantage of them if they try to cooperate. Mistrust has been held responsible for negative outcomes in a variety of settings, from general economic performance to the Cold War (Fukuyama 1995; Larson 1997).

Mediators may be able to help with mistrust problems. For instance, if two disputants are engaged in an ongoing civil or international conflict but wish to conclude a peace treaty, a mediator can reassure each side that the other is genuinely interested in peace and not attempting to deceive and exploit them. Kelman argues in the context of the Oslo negotiations that the Israelis and Palestinians "had to be persuaded that there was a genuine readiness on the other side to make the necessary concessions" and that unofficial mediation efforts "contributed to the gradual development within the two political communities of . . . a degree of working trust—i.e., trust that the other side is genuinely committed, largely out of its own interest, to finding an accommodation" (Kelman 1997; Kriesberg 2001). The fact that they were able to develop this degree of trust enabled the parties to cooperate, at least for a while.

In many cases, the parties may do most of the trust building themselves by sending signals designed to reassure the other side. However, at times mediators take a more active role and attempt to directly influence the parties' beliefs. An example of a mediator attempting to build this degree of working trust comes from Secretary of State Henry Kissinger's negotiations with Egypt

Andrew H. Kydd is Associate Professor, Political Science Department at the University of Pennsylvania. Philadelphia, PA 19104–6215 (akydd@sas.upenn.edu).

An earlier version was presented at Yale in February 2005, and at the 2004 Annual Meeting of the American Political Science Association, Chicago, IL. A yet earlier draft was presented at the Mershon Center, Ohio State University, March 5, 2004, and at the Harvard Political Economy lunch, and I thank the participants at all events for helpful comments. I also thank Yoshiko Herrera, Yevgeniy Kirpichevskly, and Amit Modi for comments.

[1] For a recent literature review, see Wall, Stark, and Standifer (2001).

[2] For a review see Powell (2002).

[3] Foundational articles on mediated communication in economics include Myerson (1986) and Forges (1986). See also Myerson (1991).

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

and Israel over the ceasefire after the 1973 war. Both Egypt and Israel understood the need to build trust to overcome the legacy of warfare, including surprise attacks in 1967 and 1973. Egyptian President Anwar Sadat in his first message to Israeli Prime Minister Golda Meir (conveyed by Kissinger) said, "You must take my word seriously. When I made my initiative in 1971 I meant it. When I threatened war, I meant it. When I talk of peace now, I mean it." The Prime Minster responded that she would do her best "to establish trust and understanding between us." Later in the negotiations, Kissinger directly vouched for Sadat with the Israeli negotiators, saying, "I have to tell you honestly: my judgment is that Egypt is genuinely willing to make peace with Israel" (Kissinger 1982, 836, 844, 964). Although some of the Israelis "suspected a trick" from Sadat, Meir was willing to pursue the negotiations, which ultimately were concluded successfully, implemented, and led on to the historic peace agreement between Israel and Egypt in 1979.

Another example comes from U.S. mediation between India and Pakistan. In the course of several crises, U.S. mediators have attempted to prevent war in South Asia. During the 1990 Kashmir crisis the United States dispatched Robert Gates to dampen each side's enthusiasm for war (Hagerty 1998, 150–52). After the December 2001 bombing of the Indian Parliament by Pakistani jihadists, India and Pakistan headed toward war again and the United States sent Richard Armitage to mediate. He obtained a promise from Pakistan's leader General Musarraf to crack down on the jihadi groups and relayed this to the Indian Prime Minister Vajpayee, adding, "I believe him." The Indian national security adviser interpreted these remarks as a serious committment. The Indians responded immediately by committing to move some naval vessels away from Pakistan, which U.S. Secretary of State Powell conveyed to the Pakistani side. The crisis then subsided (Coll 2006, 134–135).

Mistrust, then, is a form of uncertainty that can cause conflict but is susceptible to mediation. However, unlike in the case of uncertainty about resolve or relative power, the theory of how mediation can overcome mistrust problems is not well developed. Basic questions remain to be addressed. What characteristics should a mediator have to build trust? In particular, what is the impact of the mediator's preferences over the issue in dispute? Should a mediator simply want to promote cooperation, without caring what form that cooperation takes? Must mediators be unbiased, and what is the role of a concern for future payoffs and reputation? Without answers to questions such as these, theories of mediation cannot provide well-grounded hypotheses for empirical evaluation nor effective prescriptive advice for practitioners of conflict resolution.

To begin to answer these questions, I develop a model of mediation in situations of mistrust. Two sides must decide whether to cooperate with each other, each fearing that the other side may defect. The mediator has formed a judgment about the trustworthiness of the two sides, and can share this judgment with the parties in an effort to reassure them. The key question is, when will the mediator communicate honestly, particularly if the mediator has come to think that the parties are untrustworthy and need to be warned against each other. The central result of the model is that to be credible in one-round situations (assuming no intrinsic commitment to honesty) a mediator must have some preferences over the issue in dispute, have a moderate ideal point on the issue, and not find war too costly. If the mediator is too focused on preventing conflict or does not care about the issue in dispute, it will be tempted to say whatever maximizes the likelihood of agreement, namely, that the two sides are trustworthy, even if this is not what the mediator actually believes. Similarly, if the mediator shares one side's preferences over the issue completely, it will be tempted to vouch for it with the other side, even if the side it is biased toward is likely to be untrustworthy. However, if the game is repeated, the mediator can acquire a reputational incentive to be honest which is sufficient to overcome a limited amount of bias.

Although the model was developed to apply specifically to mediation, it can also be interpreted more broadly to shed light on third-party monitoring of peace agreements as well. Some peace agreements are implemented in phases, where implementation of the early phases is monitored by third parties, and the two sides reserve the right to halt implementation if a violation is detected. The phased structure is specifically designed to overcome mistrust problems which would prevent radical steps toward peace that leave the parties vulnerable. A third-party monitor can play a trust-building role by providing information in this process. The third party in this context faces similar conflicting incentives as a mediator: reporting a violation will reduce the likelihood of cooperation, which the mediator seeks to promote, yet willingness to report violations is the only thing that gives a report of no violations any meaning.

MISTRUST AND MEDIATION

Among all the contexts in which mediation is employed and mistrust can cause conflict, international and civil war are especially important because of the tremendous human costs inflicted by these struggles.[4] There is a longstanding tradition of explaining international and civil conflict as a function of vulnerability combined with distrust. Hobbes explained civil strife by arguing that all men are vulnerable and that given the inherent uncertainty about the intentions and ambitions of

[4] Another social context in which trust is important is economic exchange. If one side must perform its end of the bargain first and then wait for the other side to fulfill its promise, the first mover must trust that the second one will actually do its part (Coleman 1990, 91). Similarly, if the quality of the products is not immediately apparent, then each side must trust that the other side has not unloaded shoddy goods on them if they are to be willing to pay full price (Akerlof 1970). Economic intermediaries such as retailers and rating agencies can be seen as a way of overcoming mistrust problems that could prevent exchange under uncertainty (Biglaiser 1993; Lizzeri 1999; Spulber 1996). Milgrom, North, and Weingast (1990) analyze the role of intermediaries in the revival of trade in the middle ages.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

others it makes sense even for defensively motivated individuals to act preemptively to destroy the power of others who might threaten them (Hobbes [1651] 1968, 184). Herz developed this argument in international relations and argued that states face a security dilemma, in which anarchy plus uncertainty about the intentions of others leads states to arm themselves, which harms the security of others leading them to respond in kind in a vicious circle (Herz 1950). This argument has become a cornerstone of realist thought in international relations (Glaser 1997; Jervis 1978) and has also been used to explain ethnic conflict (Posen 1993).[5] All of these approaches imply that conflict could be prevented or resolved through strategies of reassurance or trust building, and that this could be a role for mediators in the conflict resolution or prevention process.

The literature on mediation has not ignored the issues of trust and reassurance. A research tradition begun by Burton (1969) and Walton (1969) argues that conflict is driven or exacerbated by stereotypes and mistrust and that a form of unofficial mediation involving scholars of conflict resolution can help to overcome these problems in special workshop style sessions. Fisher (1972, 1983; Fisher and Keashly 1991), develops the idea and summarizes empirical applications, whereas Kelman (2000) has led one of the most sustained efforts along these lines focused on the Israeli–Palestinian dispute. Others have borrowed the concept of confidence-building measures from international security and applied it in the mediation of family and divorce proceedings (Landau and Landau 1997). Ross and Weiland (1996) found that mediators in an experimental setting facing situations of low trust resorted to trustbuilding strategies including the use of humor. Wehr and Lederach (1991) argue that in Central America cultural factors favor mediators who are trusted members of the community, even if they are partial to one side.

However, the mediation literature has not yet come to consensus on what makes for successful mediation, either in general or in the specific case of trust building. One prominent debate has concerned the role of mediator bias. Some scholars, such as Young (1967), have included impartiality in the very definition of mediation. Others, such as Saadia Touval, have argued that mediators are often biased and can perform their tasks as well if not better for it (Touval 1975; Touval and Zartman 1989). Thomas Princen (1991, 1992) has argued that weak mediators should be neutral but strong mediators from great powers are biased but effective.

This debate has arisen anew in the formal literature on mediation and asymmetric information about resolve and power. Drawing on the cheap-talk literature in economics (Farrell and Rabin 1996), in an earlier article I argued that mediators must be biased to be effective (Kydd 2003). In that model, a mediator is trying to convince one side to make a concession by providing information about the resolve of the other side. Only a mediator who is biased toward one party can credibly tell them that the adversary will not make peace without the concession, because such a mediator would not urge a concession on the party unless the mediator thought it was truly necessary. However, Rauchhaus (2005), based on a similar model, finds that "neutral" mediators can be credible about resolve. Meawhile, Smith and Stam (2003), adapting their random walk model of war to the mediation question, find that biased mediators are not credible on the subject of the disputants' power.

The discrepency is a result of differing definitions of bias. In my previous model, a mediator is biased if it shares one side's preference ordering over the issue space and unbiased if it is indifferent over the various issue resolutions. All mediators are assumed to find war costly. In Rauchhaus's case, the mediator is neutral if its ideal point on the issue space is located between the two possible negotiated solutions, and biased if it is outside that range. Rauchhaus's neutral mediator is therefore not the same as my unbiased mediator, so the results do not necessarily conflict. For Smith and Stam (2003), a mediator is biased toward peace if it prefers peace to war but has no preferences over the issue in dispute, and biased toward a player if it prefers one side's favorite issue resolution but does not care about the cost of war. A mediator that is biased toward peace according to Smith and Stam is unbiased according to my previous model, and both find such a mediator ineffective (as does Rauchhaus). However, Smith and Stam do not consider a mediator who is both biased toward one side and finds war to be costly, and so do not evaluate if a mediator who is biased according to my previous model could be credible in their model.[6]

The dispute about mediator bias highlights the need for careful and transparent definitions of terms and rigorous analysis. Indeed the impact of bias on mediation may well depend on the context and what the mediator is trying to accomplish.[7] In the model that follows, I show that having a moderate ideal point (similar to Rauchhaus's neutrality) is good for mediator credibility, while being indifferent to the issue in dispute or having a very high cost of war (my unbiasedness and Smith and Stam's bias toward peace) is fatal to it.

THE MODEL

There are three players: player 1, player 2, and the mediator. Players 1 and 2 represent two states, or a government and a rebel group, which have fought a conflict and are contemplating a peace accord. The issue space over which they have been fighting is denoted x conceived of as the unit interval. Players 1 and 2 have utility

[5] Jervis presents the Stag Hunt, in which trust is important, and the Prisoner's Dilemma, in which it is not, as alternative models of the security dilemma. I argue that the best model of the security dilemma combines these two games in an incomplete information setting, as in the trust game below, see also Kydd 2005.

[6] Other recent game theoretic approaches to mediation include Jarque, Ponsati, and Sakovics 2003 and O'Neill 2003.

[7] For instance, Holger Schmidt (2004) analyzes when third parties should be biased or impartial in a model of third-party monitoring and enforcement of peace agreements and an analysis of postwar cases of peace implementation.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

FIGURE 1. The Bargaining Space

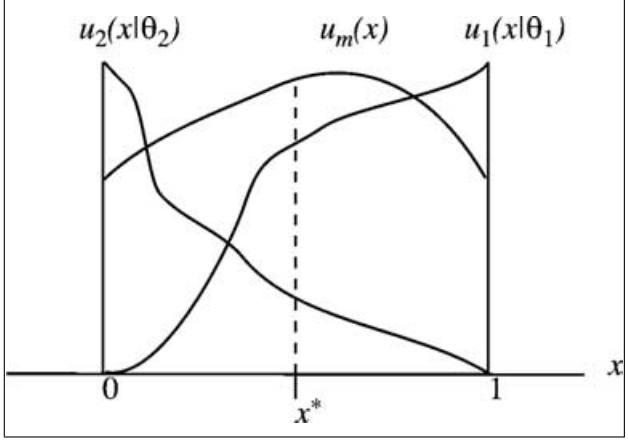


TABLE 1. Notation in the Game

$u_i(x\|\theta_i)$	Player i's utility function over the potential outcomes
θ_i	Player i's type
I_m	The mediator's ideal point
x^*	The deal on the table
m_i	Player i's military power
$p(x\|m_1, m_2)$	The likelihood of outcome x if there is conflict
c_i	Player i's costs of conflict
$w_i(m_1, m_2, \theta_i)$	Player i's war payoff
$g_i(\theta_i\|\mu_i)$	Density over player i's type
$f_i(\mu_i)$	Prior beliefs about μ_i
$h_i(\theta_i)$	Overall beliefs about player i's type
$S_i^N = \{U_i^N, T_i^N\}$,	Signals from nature about player i's type
$S_i^M = \{U_i^M, T_i^M\}$	Signals from the mediator about player i's type
η	Mediator's per round payoff
δ	Mediator's discount factor
γ	Mediator's likelihood of passing to the next round

functions over the interval, $u_1(x|\theta_1)$ and $u_2(x|\theta_2)$, which are conditional on their type, θ_i.[8] Player 1's ideal point is 1, whereas player 2's is 0, so player 1 likes higher issue resolutions and player 2 likes lower ones. The mediator also has a utility function, $u_m(x)$, with an ideal point I_m which can be anywhere in the interval. Player 1's (2's) utility is monotonically increasing (decreasing), whereas the mediator's is single peaked. I assume the mediator has no intrinsic preference for honesty in order to examine the hard case of when a mediator can be credible even without such a preference. I normalize the utilities for the ideal points to one and the utilities of the bargainers for the other side's ideal point to zero. Let the negotiated deal on the table be $x^* \in (0, 1)$. The bargaining space is illustrated in Figure 1. Notation in the game is summarized in Table 1 and proofs are relegated to the Appendix unless otherwise noted.

[8] The type refers to the shape of their utility function, as discussed next.

Each side has a stock of militarily relevant resources, m_i. While they are fighting, they maintain a level of military resources geared for war, m_i^w. If they sign and implement a peace deal, they will have to lower their guard to some extent, to a level $m_i^p < m_i^w$. This decline in military resources could be a function of the demobilization of rebel or state armies, which often takes place at the end of civil and international wars (Walter 2002). Lowering one's guard, however, induces a certain level of vulnerability, because if the other side secretly fails to disarm and attacks while one is in a weakened state, one's prospects in the war will be worse than if one had remained fully prepared.

I model conflict as a lottery that leads to some outcome in the interval.[9] In the usual winner take all model, player 1 wins with probability $\frac{m_1}{m_1+m_2}$ and player 2 wins with probability $\frac{m_2}{m_1+m_2}$ and the winner imposes its ideal point. However, given that most conflicts do not result in absolute victory, a more general formulation in which interior issue resolutions occur with positive probability seems more realistic. This is especially important in models of mediation because mediators may actively prefer interior issue resolutions and wish to make them more likely. Therefore, let the probability that any issue resolution x results from a conflict be given by a density conditional on the military resources of each side, $p(x|m_1, m_2)$. I assume that the cumulative distribution function (CDF) $P(x|m_1, m_2)$ is decreasing in m_1 and increasing in m_2.[10] This means that the likelihood that the result of a war falls below any given threshold increases as player 2 gets more powerful, and decreases as player 1 gets more powerful. This implies that the greater a side's military power, the more favorable outcomes it can expect. For instance, increasing m_1 reduces the likelihood of low outcomes, making higher outcomes more likely which is good for player 1. Each side also suffers a cost of war, c_i. Player i's payoff for war is therefore

$$w_i(m_1, m_2, \theta_i) = \int_0^1 u_i(x|\theta_i)p(x|m_1, m_2)\,dx - c_i.$$

In choosing whether to implement the negotiated deal, the two opponents face a simple two by two game. They may cooperate or defect as illustrated in Table 2. Player 1's payoffs are listed first; player 2's, second; and the mediator's, third.

With respect to the mediator's payoffs, I define two terms which will prove important in later results. First, if the mediator does not care about the issue resolution, such that $u_m(x) = 1, \forall x$, the mediator is described as *issue indifferent*. This is equivalent to Smith and Stam's (2003) "biased towards peace," or my earlier "unbiased," but is more specific and grounded in the fundamentals of the game. Second, if the mediator prefers

[9] For the debate over whether one-time lotteries are acceptable models for war initiation, see Wagner 2000 and Powell 2004.

[10] One common distribution that satisfies this criteria is the β density, $\beta(x|m_1, m_2) = \frac{x^{m_1-1}(1-x)^{m_2-1}}{\int_0^1 x^{m_1-1}(1-x)^{m_2-1}dx}$. The mean is player 1's relative power $\frac{m_1}{m_1+m_2}$.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

TABLE 2. The Mediation Game

		Player 2: Cooperate	Player 2: Defect
Player 1	Cooperate	$u_1(x^*\|\theta_1)$, $u_2(x^*\|\theta_2)$, $u_m(x^*)$	$w_1(m_1^p, m_2^w, \theta_1)$, $w_2(m_1^p, m_2^w, \theta_2)$, $w_m(m_1^p, m_2^w)$
	Defect	$w_1(m_1^w, m_2^p, \theta_1)$, $w_2(m_1^w, m_2^p, \theta_2)$, $w_m(m_1^w, m_2^p)$	$w_1(m_1^w, m_2^w, \theta_1)$, $w_2(m_1^w, m_2^w, \theta_2)$, $w_m(m_1^w, m_2^w)$

mutual cooperation and mutual defection to either side exploiting the other, such that

$$\min\left\{u_m(x^*),\, w_m\left(m_1^w, m_2^w\right)\right\} > \max\left\{w_m\left(m_1^w, m_2^p\right),\, w_m\left(m_1^p, m_2^w\right)\right\},$$

the mediator is described as *exploitation averse*. The exploitation averse mediator, in addition to wanting to promote an agreement, prefers that neither side be exploited. Exploitation aversion is a property of the payoffs in the normal form game, the main-diagonal payoffs are higher than the off-diagonal payoffs. However, it is ultimately related to the mediator's ideal point, I_m, and the location of the deal x^*. Having a moderate ideal point in the issue space proves to be a necessary condition for exploitation aversion, so exploitation aversion can be thought of as a form of unbiasedness, as I will discuss next. The mediator's preferences are known to the parties.[11]

Turning to the antagonists, we can easily rank their war outcomes. From the assumption that $P(x|m_1, m_2)$ is decreasing in m_1 and increasing in m_2 we know that $P(x|m_1^w, m_2^p)$ stochastically dominates $P(x|m_1^w, m_2^w)$ which in turn dominates $P(x|m_1^p, m_2^w)$ for any increasing utility function, such as player 1's, whereas the reverse order holds for decreasing utility functions such as player 2's (Levy 1992). Therefore, defecting while the other side cooperates beats mutual defection which in turn beats cooperating while the other side defects, the worst payoff in the game, as stated in Lemma 1.

Lemma 1. *Unilateral defection beats mutual defection which beats unilateral cooperation for both bargainers,* $w_1(m_1^w, m_2^p, \theta_1) > w_1(m_1^w, m_2^w, \theta_1) > w_1(m_1^p, m_2^w, \theta_1)$, *and* $w_2(m_1^w, m_2^p, \theta_2) < w_2(m_1^w, m_2^w, \theta_2) < w_2(m_1^p, m_2^w, \theta_2)$.

The relationship between the payoff for mutual cooperation, $u_i(x^*|\theta_i)$, and the other payoffs may vary. If either side prefers unilateral defection to mutual cooperation, it has a dominant strategy to defect, and the only equilibrium in the game is mutual defection. If both sides prefer mutual cooperation to unilateral defection, an equilibrium exists (in the complete information game) in which they both cooperate and implement the deal. I call players that prefer mutual cooperation to unilateral defection *trustworthy*, and those with the opposite ordering, *untrustworthy*.[12] Two trustworthy players can cooperate, but the presence of an untrustworthy player prevents cooperation.

Why would some types be trustworthy and others not? The answer relates to the shape of their preferences over the issue space, and whether they have extreme or moderate ambitions. For players with extreme goals, the value for the compromise would be much lower than the value of total victory. Returning to war with an advantage therefore promises great gains. More moderate types, by contrast, have abandoned extremist ambitions and place a relatively high value on the compromise in comparison to total victory. In international relations, states with such preferences are often called status quo powers, or security seekers, whereas states with more extreme goals are often termed expansionist or greedy. For instance, in Figure 1, player 1 has a utility function that assigns a relatively high valuation to the negotiated deal and the increase in utility from higher outcomes is relatively slight. For player 2, however, the deal is valued relatively low in comparison to total victory, and player 2 might well prefer to exploit player 1's cooperation and continue the war. Player 1 is therefore more likely to be trustworthy than player 2 (as illustrated in the figure).[13]

Note, the distinction between trustworthy and untrustworthy actors is different from, but related to the familiar distinction between types with high or low resolve in models of bargaining. In bargaining models, the uncertainty is over how much the other side can be pushed to make a concession. Types with a high payoff for war, in comparison to the concession one would like to extract from them, have high resolve. The payoff for war is in turn a function of relative power and the costs of war; increasing relative power and decreasing the costs of war increases resolve. These two underlying

[11] A possible avenue for advance is to incorporate uncertainty about the mediator's preferences. This would presumably lead to signaling behavior by mediators to establish their "bona-fides," which does get discussed in the participant literature on mediation.

[12] Trustworthy actors have the preference ordering associated with the Assurance Game or Stag Hunt. Untrustworthy actors may have Prisoner's Dilemma payoffs, or other orderings depending on how bad the negotiated deal is.

[13] It can easily be shown that in the case where p is the β density, if the negotiated deal reflects the balance of power, $x^* = \frac{m_1}{m_1+m_2}$, then any risk neutral or risk averse actor is trustworthy. Risk acceptant actors could be trustworthy or untrustworthy, and actors that are risk acceptant over some ranges and risk averse over others could also be either trustworthy or untrustworthy.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

variables also affect trustworthiness; raising an actor's payoff for war by increasing its relative power and decreasing its costs of war tends to make it untrustworthy (Kydd 2005). However, other factors are also important in determining trustworthiness, most importantly, how much the parties could gain if they overturn the status quo or renege on the deal and are successful. In Figure 1, player 1 stands to gain little, whereas player 2 could gain much more. This kind of difference in utility functions is orthogonal to relative power and the costs of fighting. A player could have high resolve, because of high relative power, but be trustworthy, because it stands to gain little from a successful war. Such a player could not be bullied for a concession but would not renege on the deal either. Conversely, a player could have relatively little relative power but much to gain from a successful war, so that if the other side made itself vulnerable in implementing a peace deal, the player would take advantage of this to attack. Thus resolve and trustworthiness are correlated through their dependence on relative power and the costs of war, but not perfectly so because of the independent effect of the shape of the utility function.

If the game featured complete information, so the player's types, θ_i, were known, the role of the mediator would be limited. As noted earlier, if either side were untrustworthy, the only equilibrium would be mutual defection. If both sides were trustworthy, a pareto optimal equilibrium would exist in which both sides cooperate. The mediator could help focus expectations on this equilibrium, but the players would already know of its existence. To generate a role for a trust-building mediator, we need to add mistrust to the model, in the form of uncertainty about whether the other side is disposed to honor the negotiated settlement or not.

Along these lines, we can introduce uncertainty about trustworthiness into the game by positing a space of utility functions for player 1 indexed by $\theta_1 \in [0, 1]$, such that if $\theta_1 < \theta_1'$ then $u_1(x^*|\theta_1) - w_1(m_1^w, m_2^p, \theta_1) < u_1(x^*|\theta_1') - w_1(m_1^w, m_2^p, \theta_1')$ and similarly for player 2. This means that for higher values of θ_i, the player has a higher valuation for the negotiated settlement as opposed to reneging on the deal. The important property of such a space of utility functions is that there will be a threshold for θ_i below which player i is untrustworthy, and above which it is trustworthy. By making the players uncertain about the θ_i value for the opponent's utility function, they become uncertain if they can trust the other side.[14]

Specifically, assume that there is a distribution over the player's types, $g_i(\theta_i|\mu_i)$. This could represent the fact that the actors have somewhat different utility functions for different issues, or that even if one were to find out all there is to know about a player, their utility function would still be somewhat uncertain because of errors and randomness. The mean of each player's distribution, however, tells us something about how likely they are to be trustworthy. If a player has a high mean value for θ_i, then it is likely to be trustworthy, if the mean is lower, the player is less likely to be trustworthy. There are prior distributions over the means, $f_i(\mu_i)$, which summarize the player's initial beliefs about how likely the other players are to be trustworthy.[15] This uncertainty structure implies that θ_i will be distributed according to

$$h_i(\theta_i) = \int_0^1 g_i(\theta_i|\mu_i) f_i(\mu_i)\, d\mu_i.$$

If the mediator is to build trust between the parties, it must have some information to share with them about the other side's type. One way to conceive of this is to posit that the mediator observes the behavior of the parties and forms its own judgment about their trustworthiness. Kissinger, for instance, in his 1973 shuttle diplomacy had many opportunities to observe the behavior of the leaders of Egypt and Israel, as well as their key subordinates. These leaders had no opportunities to observe each other face to face, and even lower level direct meetings had only recently begun in the Sinai. These opportunities enabled Kissinger to make an independent judgment about how likely Egypt was to be willing to make peace, and convey this to Israel. Even when leaders do have contacts with each other, such as at Camp David in 2000, mediators typically have more freedom to move among each side's camp and observe how they react to the ebb and flow of negotiations, and can therefore form their own judgments about each side's commitment to peace. These judgments, if credibly conveyed, could in turn affect the parties' beliefs. Note, the mediator is not simply conveying promises of good behavior, because such promises would be made by untrustworthy types as well, whether to the mediator or directly to the other side. Instead, the mediator is conveying its opinion on the credibility of such promises, formed from its interaction with the other side.

Formally, assume that the mediator receives a signal from nature about the parties in which if player i seems to have a utility function such that θ_i is greater than some threshold, θ_i^0, the mediator receives a signal T_i^N indicating player i is trustworthy, whereas if player i seems to have $\theta_i < \theta_i^0$, then the mediator gets the U_i^N signal indicating player i is untrustworthy. Let the signal $S_i^N \in \{U_i^N, T_i^N\}$ be an ordered set where the order corresponds to the likelihood of being trustworthy, so $U_i^N \prec T_i^N$. I assume that the players know the mediator's information about themselves, but do not know the mediator's information about the other side. The mediator's posterior beliefs are derived from Bayes' Rule and denoted $f_i(\mu_i|U_i^N)$ and $f_i(\mu_i|T_i^N)$ after the untrustworthy and trustworthy signals, respectively, and are shown in the Appendix. These can be substituted into the expression for h_i to derive posterior

[14] A numerical example of such a family of functions is $u_1(x|\theta_1) = x + (1-2\theta_1)x(x-1)$, and $u_2(x|\theta_2) = 1 - x + (1-2\theta_2)x(x-1)$. These functions move smoothly upwards as θ_i increases from zero to 1, raising the value of the negotiated settlement as opposed to the ideal points. These functions are either risk acceptant or risk averse over their entire range, but this need not be the case in general.

[15] This is the standard random utility model (McFadden 1976; Signorino 1999).

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

beliefs about θ_i. Effectively, receiving the U_i^N signal shifts the mediator's beliefs downwards, so that the party concerned is believed less likely to be trustworthy, whereas receiving the T_i^N signal shifts the beliefs upwards, toward higher values of θ_i, so the party is thought more likely to be trustworthy. The posterior beliefs are also more concentrated about their means, reflecting greater certainty on the mediator's part after getting new information.

An important fact about the posterior beliefs is that if the mediator receives a U_i^N signal, it's posterior estimate that player i's utility function index θ_i falls below any given threshold is higher than it would have been if it had received the opposite T_i^N signal. This is a simple way of summing up the fact that observing behavior that indicates untrustworthiness increases one's belief that the player will be untrustworthy in the future. This result is stated in the following lemma.

Lemma 2. *The posterior belief that player i's type falls below a certain threshold, $H_i(\theta_i|S_i^N)$, is decreasing in S_i^N, that is, $H_i(\theta_i|U_i^N) > H_i(\theta_i|T_i^N)$.*

After the mediator updates its beliefs, it then communicates with each player privately about the trustworthiness of the other player. The mediator can say that the other side is likely to be trustworthy, T_i^M, corresponding to the T_i^N signal, or that the party is likely to be untrustworthy, U_i^M, reporting the U_i^N signal. Let the mediator's communication be denoted $S_i^M \in \{T_i^M, U_i^M\}$. After the mediator communicates with each player, if the mediator is believed to be telling the truth, the parties' beliefs will shift to mirror the mediator's. The mediator will of course take this into account when deciding whether or not to be honest.

After the mediator's communications, the two players simultaneously choose to cooperate or defect, and the game ends.

EQUILIBRIA IN THE GAME

I solve for perfect Bayesian equilibria. In cheap-talk games of this kind there are two broad categories of equilibria, "babbling" and "truth telling." In babbling equilibria, the mediator's signals are uncorrelated with its information, and the players retain and act upon their prior beliefs. Babbling equilibria are possible regardless of the initial conditions.[16] In truth-telling equilibria, the mediator faithfully passes on its information, the players update their beliefs and act accordingly. For such an equilibrium to hold, the mediator's payoff for sending a signal must be conditional on the mediator's beliefs, such that if the mediator received the T_i^N signal, it would prefer to send the T_i^M signal rather than the U_i^M signal, but if it received the U_i^N signal, it would prefer to send the U_i^M signal rather than the T_i^M signal. Interestingly, successful lying is never in equilibrium in this kind of game, given that the mediator's preferences are known. If the mediator said the reverse of the truth in equilibrium, the parties would simply adjust their interpretation of the mediator's statements to reverse their meaning. If it said one thing all the time, the parties would ignore the mediator (as in a babbling equilibrium).[17] The main focus of analysis is therefore to determine the conditions under which truth telling is sustainable in equilibrium. I first discuss the player's behavior and then the mediator's.

The Players' Behavior

All types of player 1 with $w_1(m_1^w, m_2^p, \theta_1) > u_1(x^*|\theta_1)$ have a dominant strategy to defect and similarly for player 2. Types with $w_1(m_1^w, m_2^p, \theta_1) < u_1(x^*|\theta_1)$ could cooperate given certain beliefs and expectations about behavior from the other side. Equilibrium behavior of the players is described in the following theorem.

Theorem 1. *In any equilibrium there will be cutoff points, θ_i^* below which types will defect and above which types will cooperate. One possible equilibrium in which no type cooperates is $\theta_1^* = \theta_2^* = 1$. If an equilibrium exists in which some types cooperate, the cutoff points can be derived from the following relation for player 1,*

$$u_1(x^*|\theta_1^*) \geq w_1(m_1^w, m_2^p, \theta_1^*) + \left(\frac{1}{1 - H_2(\theta_2^*|S_2^M)} - 1\right) \times \left[w_1(m_1^w, m_2^w, \theta_1^*) - w_1(m_1^p, m_2^w, \theta_1^*)\right], \quad (1)$$

and a similar one for player 2.[18]

A "noncooperative" equilibrium with $\theta_i^* = 1$ is possible regardless of beliefs. If the other side is expected to defect regardless of its payoffs, each side has an incentive to defect as well, to avoid the payoff from being exploited. If the players start out very suspicious of each other, this may be the only equilibrium. If the players are not that suspicious, other equilibria may be possible in which there is some chance that each player cooperates. Since each equilibrium cutoff point is an increasing function of the other one, if there are multiple equilibria they can be ordered by the likelihood of mutual cooperation. I will focus on the "most cooperative" one in which the likelihood of mutual cooperation is highest (θ_i^* is lowest), given that its Pareto superiority has obvious focal appeal, and it has convenient and quite general comparative static properties (Milgrom and Roberts 1994; Milgrom and Shannon

[16] The game can only go off the equilibrium path in a babbling equilibrium in which the mediator is expected to send one signal regardless of the signal received, but actually sends the other one. Therefore assumptions about off equilibrium path beliefs do not affect the conditions under which truth-telling is possible. For completeness, I assume that signals that have probability zero are perceived to be uncorrelated with the player's types. This keeps beliefs unchanged, supporting the babbling equilibrium.

[17] In cheap-talk games in which the sender knows the value of an uncertain parameter exactly, it can often only truthfully reveal a range in which it falls. Even in these games, however, the mediator never lies about the correct range in equilibrium.

[18] Note, the cutoff points depend on the signals received. However, given that the players know what signals are sent about themselves (and obviously which ones they receive), in a truth-telling equilibrium, there is complete information about what signals are sent.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

1994). I restrict attention to the case where the posterior distributions, after the mediator's information is revealed, could support an equilibrium with $\theta_i^* < 1$ and hence some chance of cooperation. If this were not the case, mediation would be pointless.

The central comparative static of interest with respect to the player's behavior is how the thresholds for cooperation respond to the mediator's information, assuming it to be honestly relayed. This is given in the following lemma.

Lemma 3. *In a truth-telling equilibrium, each player is more likely to cooperate if it is reassured about the other side, that is, $\theta_i^*(S_j^M)$, is declining in S_j^M.*

Proof. From Lemma 2, we know that $H_j(\theta_j^*|S_j^N)$ is decreasing in S_j^N and in a truth-telling equilibrium the player's beliefs will be the same as the mediator's, so S_j^M can be substituted for S_j^N. Since θ_i^* is increasing in $H_j(\theta_j^*|S_j^M)$ from Equation 1, the result is obtained. ■

What this means is that if the mediator tells a player that the other side is trustworthy, that player will be more likely to cooperate than if the mediator told it that the other side was untrustworthy. This is a natural result of the player becoming more trusting of the other side, and hence more willing to take a risk and cooperate.

The Mediator's Behavior

The mediator's payoff, denoted η, as a function of the signals it receives, S_i^N, and those it sends, S_i^M, is

$$\begin{aligned}\eta&\left(S_1^M, S_2^M|S_1^N, S_2^N\right)\\ &= u_m(x^*)\left[1 - H_1\left(\theta_1^*\left(S_2^M\right)|S_1^N\right)\right]\left[1 - H_2\left(\theta_2^*\left(S_1^M\right)|S_2^N\right)\right]\\ &\quad + w_m\left(m_1^w, m_2^p\right)H_1\left(\theta_1^*\left(S_2^M\right)|S_1^N\right)\left[1 - H_2\left(\theta_2^*\left(S_1^M\right)|S_2^N\right)\right]\\ &\quad + w_m\left(m_1^p, m_2^w\right)\left[1 - H_1\left(\theta_1^*\left(S_2^M\right)|S_1^N\right)\right]H_2\left(\theta_2^*\left(S_1^M\right)|S_2^N\right)\\ &\quad + w_m\left(m_1^w, m_2^w\right)H_1\left(\theta_1^*\left(S_2^M\right)|S_1^N\right)H_2\left(\theta_2^*\left(S_1^M\right)|S_2^N\right). \quad (2)\end{aligned}$$

In a babbling equilibrium, the θ_i^* will not be affected by the S_j^M because the players will disregard the mediator and act on their prior beliefs. The mediator's payoff is therefore unaffected by its communication and it will be willing to play its role in the equilibrium. In a truth-telling equilibrium, the mediator's payoff will depend on its communication, and we can derive the conditions under which the mediator will tell the truth, as stated in the following theorem.

Theorem 2. *In the single-round mediation game, for a truth-telling equilibrium to be possible, the mediator must be exploitation averse, and its best (worst) payoff cannot be too much higher (lower) than the other three.*

For the mediator to have an incentive to be honest, it must prefer that the two sides either both cooperate or both defect rather than that either side exploit the other. If this were not the case, the mediator would benefit by encouraging one side to cooperate even if it thought the other side was untrustworthy and would defect. In addition, the mediator's best payoff cannot be too much greater than the other three, or it will have an incentive to say whatever maximizes the likelihood of achieving that best payoff. Similarly, it's worst payoff cannot be so bad that it will say anything that reduces the likelihood of receiving it.[19]

The conditions in Theorem 2 are constraints on the payoffs in the normal form mediation game, but these payoffs are in turn based on more fundamental preferences over the outcomes, $u_m(x)$, the mediator's ideal point I_m, and its costs of war c_m. By examining how these factors affect the payoffs, we can determine the impact of these more basic variables on the ability of mediators to be credible trust builders.

First, consider the shape of the mediator's utility function over the outcomes, $u_m(x)$. The function cannot be too flat, or the second condition in Theorem 2 will be violated.

Corollary 1. *In the single-round mediation game, if the mediator's utility function is too flat, a truth-telling equilibrium will not be possible.*

In the extreme case, if the mediator is issue indifferent, so that $u_m(x) = 1, \forall x$, its payoff from Equation 2 will reduce to a simple function of the likelihood of mutual cooperation. It will therefore have an incentive to say whatever will maximize the chance that both sides cooperate. If the players will believe the mediator, it will have an incentive to vouch for both of them, because this makes them more likely to cooperate. This incentive will hold regardless of the mediator's information. Therefore the mediator who just wants to promote cooperation will vouch for the players regardless of its beliefs about their trustworthiness, and so it will not be capable of sustaining a truth-telling equilibrium. Hence it will have no ability to reassure the players and no impact on the likelihood of cooperation. Note, this result in the trust-building context is the same as that found in case of communication about power and resolve. In my earlier model, and in Smith and Stam's (2003), a mediator who just wants to avoid war (biased in favor of peace according to Smith and Stam, unbiased according to me) will not be effective. Purely issue-indifferent mediators may seem rare, in that they have absolutely no preferences whatsoever over the issue in dispute. However, the mediator need not have perfectly flat preferences to be afflicted by this problem, the preferences simply need to be flat enough so that the war payoffs are all similarly valued at about $1 - c_m$ and the value for the deal is about 1.

This highlights the relation between the flatness of the utility function and the mediator's cost of war. As the following corollary indicates, if the cost of war is too great, the effect is the same as if the mediator has flat preferences.

[19] It bears repeating that the model assumes that the mediator has no innate attachment to honesty per se. Such an attachment could be easily modeled by adding a parameter to the mediator's payoff representing a cost for making a statement known to be false. This would expand the range of truth-telling equilibria. The model, in effect, identifies the incentives for honesty or dishonesty present in the situation itself, combined with the mediator's policy preferences.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

Corollary 2. *In the single-round mediation game, if the mediator's cost of war c_m is too great, a truth-telling equilibrium will not be possible.*

If the cost of war, c_m, is too great, the mediator's preferences will approximate the case in which the mediator is indifferent among the issue resolutions but just wants to avoid war. In this case the temptation to say whatever maximizes the chance of peace, namely, that the two parties are trustworthy, will be irresistible. It might seem that most mediators would have costs of war below this threshold, because the mediator's cost of conflict is so much less than that of the players. It is true that the mediator does not directly suffer the death and destruction, not to mention the financial expense involved in war. However, the mediator also typically cares much less about the issues being fought over than the parties do. The flatter the mediator's utility function, the lower the upper bound on the costs of war. The less the mediator cares about what the parties are fighting over, the less it must care about the cost of war if it is to be able to tell the truth. In the limit, if the mediator is issue indifferent, any positive costs of war at all will make it unable to be honest. In practice, therefore, this constraint is not trivial. Many well-intentioned mediators who do not really care about where a border gets drawn in an empty mountainous region but who abhor the waste and destruction of war could be tempted to lie in order to encourage cooperation. In this case, the parties would be well advised to get a mediator who cares more about the outcome and less about the suffering of war.

The mediator must care about the outcome, but not be too biased. Consider the mediator's ideal point, I_m.

Corollary 3. *In the single-round mediation game, for a truth-telling equilibrium to be possible, the mediator's ideal point I_m must be between a lower bound and upper bound, $I_m \in [\underline{x}, \overline{x}]$ where $\underline{x} > 0$ and $\overline{x} < 1$. That is, the mediator cannot be too biased toward either side.*

If the mediator's ideal point is too far to the right, then its preferences over the war outcomes in the game will be $w_m(m_1^w, m_2^p) > w_m(m_1^w, m_2^w) > w_m(m_1^p, m_2^w)$, which is the same as player 1's ordering. We can think of this as an effective indicator that the mediator is biased in favor of player 1, because it shares player 1's preference ordering. This obviously violates exploitation aversion, so the mediator will not be able to be honest to player 2 about player 1's trustworthiness. The same logic applies if the mediator's ideal point is too far to the left, the mediator will share player 2's preference ordering and prefer to have player 2 exploit player 1 rather than have both defect. Therefore, if the mediator's ideal point is too extreme, if the mediator is too biased, it will be incapable of supporting a truth-telling equilibrium. If the ideal point is within the appropriate range, I define it to be *moderate*. The concept of a moderate ideal point is similar to Rauchhaus's definition of neutrality, having an ideal point between the ideal points of the players, or between two relevant offers on the table (Rauchhaus 2005).[20]

A moderate ideal point may seem like a common enough attribute for a mediator, especially well-intentioned mediators such as United Nations officials or Scandinavian diplomats. However, not all participants in war are normatively equal, and sometimes grave human rights abuses are perpetrated by one side and not by the other. It is conceivable that when mediating between, for example, the Khmer Rouge and Cambodian Government, or between Uganda and the Lord's Resistance Army, the mediator could share the government's preference to execute a surprise attack and disarm the guerilla forces so their leaders could be brought to justice.[21] Such a mediator would not be credible to the rebel forces, and so could not build their trust in the government. Similarly, mediators in hostage situations may not actually prefer that a moderate compromise be worked out between the hostage takers and the authorities, but could easily prefer that the hostage takers be lured into the open where they can be targeted by snipers or overwhelmed by assault, so long as the hostages could be saved.

In the one-round game, therefore, if a mediator is to be credible in a trust-building role, too much bias is fatal. This result contrasts with my earlier finding that bias is necessary for honesty in conveying information about resolve (Kydd 2003). This discrepancy is a result of certain differences in the models. In my earlier model, the mediator was assumed to share one side's preference ranking over the possible issue resolutions, or be indifferent to the issue. The possibility of an interior ideal point was not considered, and this turns out to be essential to truth-telling in the trust-building context.[22] With an interior ideal point, the mediator can be in a sense biased toward both sides, in that the mediator shares with each a preference that they not be exploited.

Real preferences over the issue space, a low enough cost of conflict, and a moderate ideal point are necessary to produce the appropriate incentives in a mediator to support honesty in a truth-telling equilibrium. But what if the mediator does not have these characteristics, and so would not be honest in a one time interaction? Could the desire to preserve a reputation so as to ensure future employment prove adequate to ensure honesty?

[20] The costs of war do not affect these constraints. This is because in the definition of the war payoff, the cost of war for the mediator is the same in each of the three war outcomes, and so drops out of the comparison that determines the rank ordering of the war payoffs.

[21] Especially when they are under indictment at the International Criminal Court, as is LRA leader Joseph Kony.

[22] In my earlier model, an interior ideal point would likely lessen the mediator's credibility. However, if the mediator had a middling ideal point and one side was stronger than the other, and so was trying to extract a concession that would be viewed unfavorably by the mediator, the mediator would likely be credible to the weaker side in advising whether to grant the concession, because the mediator would share the weaker side's preference not to grant the concession unless it were necessary.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

REPETITION AND REPUTATION

Consider a repeated version of the game. The players are chosen anew each round, so the information structure is the same. The mediator is the same each round, with a discount factor δ.[23] The mediator is "hired" for the round, provided it has never said that a player is trustworthy who subsequently defects. If the mediator does make this mistake, it is never hired again. Note, such a mistake is easily visible because each side knows what the mediator tells it and can see whether the other side cooperates. If the mediator vouches for the other side and it then defects, a player is free to advertise this to the world, harming the mediator's reputation.[24] For all rounds in which the mediator is not hired, it receives a payoff of zero. I consider perfect Bayesian equilibria, so that the mediator's choice at each round must be optimal given its information. As in the one-round game, untrustworthy types will defect in any equilibrium, whereas trustworthy types can cooperate if they are trusting enough.

The payoff to the mediator for the equilibrium strategy is a function of the per-round payoff and the likelihood of passing to the next round. Equation 2 gives the mediator's payoff conditional on the signals received and sent. To derive the ex-ante per-period payoff, this must be expanded assuming that signals sent match signals received. The expansion is shown in the Appendix. Denote the ex-ante per-round payoff for telling the truth η_t. The likelihood that the mediator gets to the next round is the likelihood that the player(s) the mediator vouches for cooperates. The ex-ante likelihood of getting to the next round by telling the truth is denoted γ_t and the formula is also in the Appendix. The equilibrium payoff in the game can then be written as

$$\eta_t + \delta\gamma_t\eta_t + (\delta\gamma_t)^2\eta_t + \cdots = \frac{\eta_t}{1-\delta\gamma_t}.$$

This is also the continuation payoff from any point forward, assuming honest behavior from then on.

To determine when the equilibrium is sustainable, we must consider the payoffs for deviations ex-post, after the mediator receives its information. Denote the payoff from telling the truth in any round given the signals received as $\eta_t(S_1^N, S_2^N)$ and the corresponding payoff from a contemplated deviation as $\eta_d(S_1^N, S_2^N)$. The likelihood of getting to the next round will also be a function of the signals and the contemplated deviation, denote these as $\gamma_t(S_1^N, S_2^N)$ and $\gamma_d(S_1^N, S_2^N)$. The payoff for sticking to the truth can then be written as

$$\eta_t(S_1^N, S_2^N) + \delta\gamma_t(S_1^N, S_2^N)\frac{\eta_t}{1-\delta\gamma_t},$$

whereas the payoff for a one-round deviation[25] is

$$\eta_d(S_1^N, S_2^N) + \delta\gamma_d(S_1^N, S_2^N)\frac{\eta_t}{1-\delta\gamma_t}.$$

Following the equilibrium payoff will beat the deviation if

$$\eta_d(S_1^N, S_2^N) - \eta_t(S_1^N, S_2^N) \leq \left(\gamma_t(S_1^N, S_2^N) - \gamma_d(S_1^N, S_2^N)\right)\delta\frac{\eta_t}{1-\delta\gamma_t}. \quad (3)$$

Note that for the equilibrium to hold, if a deviation improves the payoff in the current round it must reduce the likelihood of getting to the next round, or if it improves the likelihood of getting to the next round, it must reduce the payoff in the current round. If a deviation improved both the current round payoff and the likelihood of getting to the next round, the equilibrium would be unsustainable. Also, as usual in such games, the discount factor, δ, must not be too small or the mediator will have an incentive to do whatever maximizes the current round payoff.

The effect of repeating the game is to make truth-telling equilibria possible when they were not before, as reflected in the following theorem.

Theorem 3. *In the repeated mediation game, a truth-telling equilibrium can be possible even with a biased or an issue indifferent mediator, that is if* $I_m \notin [\underline{x}, \bar{x}]$ *or* $u_m(x) = 1, \forall x$.

Intuitively, the fact that the game is repeated can give the mediator an incentive to be honest. The mediator knows that if it is caught in a lie, saying a player is trustworthy who then defects when a trustworthy player would have cooperated, it will not be hired again. Being hired again is valuable because of the opportunity to get more benefits for fostering successful cooperation. So the mediator may not wish to jeopardize its chances of getting to the next round by vouching for the players when it has received information that they are not that trustworthy. Note, however, that the effect of bias and issue indifference is still negative; if either becomes too pronounced then credibility will be undermined. But reputation can make up for limited amounts of bias or issue indifference.

[23] An alternative structure would be to have the same mediator with the same parties over time, as in a prolonged negotiation or peace implementation. This would complicate things by allowing the parties to update from each other's behavior as well as the mediator's information, and introduce the potential for complicated player strategies involving reciprocity. My conjecture is that a similar dynamic would operate to give the mediator further incentives for honesty, however.

[24] The mediator could lie and claim that it warned the player about the other side, but such claims would be obviously self-interested and hence not likely to be credible.

[25] If no one round deviation is profitable, no longer one will be. If a one round deviation is not profitable, an n round deviation will be beat by an $n-1$ round deviation where the mediator returns to honesty one round earlier. This will unravel to the first round, so any finite deviation will be beat by the equilibrium. The payoff for an infinite deviation will be the limit of the payoffs of finite deviations of increasing lengths, which are decreasing, all of which are beat by the equilibrium, so the infinite deviation will be beat by the equilibrium as well.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

DISCUSSION AND CONCLUSION

Systematic empirical testing of the arguments presented here will be difficult. First of all, the unit of analysis in typical datasets on bargaining and mediation is an entire negotiation from beginning to end, which can take months if not years (e.g., Brecher and Wilkenfeld 2000). These are coded according to certain characteristics of the actors, the identity of the mediator, previous attempts at mediation, previous agreements, and whether they succeeded or failed according to various metrics. In the course of a negotiation, the mediator involved may do many things at different times, and more than one mediator may be brought in. In the model presented here, however, only one task of the mediator is examined, building trust through speech acts vouching for the players. The unit of analysis should therefore be speech acts by the mediator, and the relevant effects are the beliefs of the parties, and their subsequent actions. In a typical negotiation that gets recorded in a dataset as a single observation, there will implicitly be thousands of speech acts by the mediator and the parties to the negotiations, and the individual effects of any small subset of these will be impossible to distinguish.

Even if a dataset with the appropriate unit of analysis were available, selection effects would be formidable. Some mediators impose themselves on the parties, but others need acceptance by both sides to function. Those that need acceptance will be selected by the parties based on anticipated effectiveness, among other criteria. This would imply that even if my model were correct, many mediators in the sample would have the appropriate characteristics identified by the model, but the ones that did not, having been selected because of some feature known to the parties that makes them effective, would also be effective; hence, the overall result would suggest that the criteria the model suggests are not essential. This selection effect could be overcome by modeling the selection stage. For instance, Holger Schmidt's (2004) study of third-party intervention in implementation examines first whether the intervenors selected have the characteristics predicted by the theory, and then separately whether they are in fact effective.

An example of the quantitative literature that begins to address issues raised by the model is Bercovitch and Houston (1993, 317) who find, in a quantitative study of 364 post-1945 mediation efforts, that "another strong effect in our data concerns the importance of a continuing relationship, especially one that may extend into the future. Mediators from the same bloc tend to be more successful and to use a different pattern of strategies than other mediators. Mediation works best when the parties and the mediator share some bonds and are part of a recognizable network of interdependence". The different strategies mentioned are "communication facilitation" strategies, which correspond more closely to trust building than the other two categories, "procedural," and "directive." Unfortunately, the problems just identified apply here, in that the communication facilitation strategy includes a vast array of things besides trust building, and there is no selection model.

While awaiting better systematic testing, the model can shed light on debates in the literature on mediation and find application to qualitative information derived from more detailed narratives of mediation efforts. The model supports the intuition that mediators should be unbiased to build trust; however, it clarifies the issue by identifying *which kind of unbiasedness is helpful* (moderate ideal points) and *which harmful* (indifference to the issue). Countless scholars and practitioners have noted the advantage of neutrality in mediation without being very rigorous about what they mean by it and why it works. Young (1967, 81) argues that "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)". In fact, however, the mediator must have "something to gain" because it must have clear preferences for moderate outcomes and so wish to achieve some favored issue resolution. This is what translates into a desire to see that neither side gets exploited, which is essential for credibility. Indeed, of all the findings of the model, the fact that mediators who simply want to avoid war will have a serious credibility problem and may therefore be incapable of lessening its likelihood seems the most counterintuitive and absent from the writings of practitioners.

Repetition can provide an additional incentive for honesty. Mediators that are biased, but not too biased, can still be honest brokers if they could lose their position by being caught vouching for players who subsequently betray the other side. This provides a theoretical rationale for Wehr and Lederach's observation that the parties to conflict in Central America prefer mediators who have ties to the community even if they are somewhat biased, over outsiders who are neutral. Wehr and Lederach (1991, 87) write,

> The insider-partial is the "mediator from within the conflict" whose acceptability to the conflictants is rooted not in distance from the conflict or objectivity regarding the issues, but rather in connectedness and trusted relationships with the conflict parties. The trust comes partly from the fact that the mediators do not leave the postnegotiation situation . . . They must continue to relate to conflictants who have trusted their commitment to a just and durable settlement.

The model clarifies the strategic logic of this case. A somewhat biased mediator in a repeated setting may be honest, whereas in a one round game, a mediator who is indifferent over the issue outcomes and just wants to resolve the conflict will not be credible. Therefore the parties may be right to prefer a moderately biased intermediary who has lasting ties to the community.

In international conflict resolution, different mediators will have different concerns for reputation versus immediate success. When President Clinton mediated between Arafat and Barak at Camp David in 2000, the reward for a successful settlement would have been tremendous, whereas Clinton's concern about future mediation opportunities would have been minimized

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

by uncertainty that there would ever be a next time, at least of such magnitude, at least for him. However, long-term mediators associated with international organizations such as the United Nations could perhaps have a stronger incentive to preserve a reputation for honesty. States, such as the Scandinavian countries, that specialize in international conflict resolution could also establish incentives for their diplomats to preserve the national reputation for credibility.

Finally, it is important to emphasize that there are many other tasks that mediators perform besides trying to build trust between the parties. Mediators need credibility in all the tasks they undertake that involve attempting to persuade the parties of something. Some of the results obtained here about the mediator's ideal point, preferences, and cost of conflict may extend to other contexts, whereas others will not. Each context needs to be examined carefully before a complete picture of the requisites of mediator credibility can be drawn.

APPENDIX

Proof of Lemma 2

For convenience I will drop the subscript denoting the player, since it is the same throughout. Recall that θ^0 is the cutoff point between the U^N and T^N signals, and let θ^* be the the new threshold under consideration. The posterior beliefs over μ are $f_i(\mu_i|U_i^N) = \frac{G(\theta^0|\mu)f(\mu)}{H(\theta^0)}$ and $f_i(\mu_i|T_i^N) = \frac{(1-G(\theta^0)|\mu)f(\mu)}{1-H(\theta^0)}$. We wish to show that the posterior beliefs over t have the following relation,

$$\int_0^{\theta^*}\int_0^1 g(\theta|\mu)\frac{G(\theta^0|\mu)}{H(\theta^0)}f(\mu)\,d\mu\,d\theta > \int_0^{\theta^*}\int_0^1 g(\theta|\mu)\frac{1-G(\theta^0|\mu)}{1-H(\theta^0)}f(\mu)\,d\mu\,d\theta.$$

Simplifying, we get

$$\int_0^{\theta^*}\int_0^1 g(\theta|\mu)\left(\frac{G(\theta^0|\mu)}{H(\theta^0)} - \frac{1-G(\theta^0|\mu)}{1-H(\theta^0)}\right)f(\mu)\,d\mu\,d\theta > 0$$

$$\int_0^{\theta^*}\int_0^1 g(\theta|\mu)\frac{G(\theta^0|\mu)-H(\theta^0)}{H(\theta^0)(1-H(\theta^0))}f(\mu)\,d\mu\,d\theta > 0$$

$$\int_0^{\theta^*}\int_0^1 g(\theta|\mu)(G(\theta^0|\mu)-H(\theta^0))f(\mu)\,d\mu\,d\theta > 0$$

$$\int_0^1 G(\theta^*|\mu)(G(\theta^0|\mu)-H(\theta^0))f(\mu)\,d\mu > 0$$

$$\int_0^1 G(\theta^*|\mu)G(\theta^0|\mu)f(\mu)\,d\mu - H(\theta^0)H(\theta^*) > 0$$

$$\int_0^1 G(\theta^*|\mu)G(\theta^0|\mu)f(\mu)\,d\mu - \int_0^1 G(\theta^*|\mu)f(\mu)\,d\mu\int_0^1 G(\theta^0|\mu)f(\mu)\,d\mu > 0.$$

This is the covariance of $G(\theta^*|\mu)$ and $G(\theta^0|\mu)$. Given that these are both decreasing functions of a common underlying variable, μ, the covariance will be positive, as proven in Schmidt 2003.

Proof of Theorem 1

Let the likelihood that player j cooperates in equilibrium be α_j. Player 1's payoff for cooperation is $\alpha_2 u_1(x^*|\theta_1) + (1-\alpha_2)w_1(m_1^p, m_2^w, \theta_1)$, whereas the payoff for defection is $\alpha_2 w_1(m_1^w, m_2^p, \theta_1) + (1-\alpha_2)w_1(m_1^w, m_2^w, \theta_1)$. Cooperation beats defection if player 1's value for the deal is above a threshold:

$$u_1(x^*|\theta_1^*) \geq w_1(m_1^w, m_2^p, \theta_1^*) + \left(\frac{1}{\alpha_2} - 1\right) \times \left[w_1(m_1^w, m_2^w, \theta_1^*) - w_1(m_1^p, m_2^w, \theta_1^*)\right].$$

In a truth-telling equilibrium the other side's likelihood of cooperation is $1 - H_j(\theta_j^*|S_j^M) = \alpha_j$. In a babbling equilibrium the prior beliefs $1 - H_j(\theta_j^*)$ can be substituted.

Proof of Theorem 2

If the mediator is not exploitation averse, then $\min\{u_m(x^*), w_m(m_1^w, m_2^w)\} < \max\{w_m(m_1^p, m_2^w), w_m(m_1^w, m_2^p)\}$. Pick the case where $u_m(x^*) < w_m(m_1^w, m_2^p)$. If $w_m(m_1^w, m_2^w) > w_m(m_1^p, m_2^w)$, then the mediator's payoff increases with $H_1(\theta_1^*(S_2^M)|S_1^N)$, so the mediator has a dominant strategy to send the U_2^M signal and discourage player 1 from cooperating. If $w_m(m_1^w, m_2^w) < w_m(m_1^p, m_2^w)$, then receiving the T_2^N signal indicating player 2 is trustworthy can only increase the incentive to send the U_2^M signal to discourage player 1 from cooperating. The mediator would want player 1 to cooperate when player 2 is untrustworthy, and not to cooperate when player 2 is trustworthy. In neither case is a truth-telling equilibrium possible. The symmetry of the situation means the same analysis applies to the other cases.

Assume that the mediator's best payoff is mutual cooperation. As the mediator's payoff for mutual cooperation increases, from Equation 2, eventually the mediator will have a dominant strategy to encourage cooperation by sending T_i^M signals regardless of the signals received. If the best payoff were mutual defection, the same incentives would eventually mandate U_i^M signals regardless of information. A similar analysis implies that if, say, $w_m(m_1^w, m_2^p)$ is the worst payoff, if it grows too bad the mediator will have a dominant strategy to minimize its likelihood by sending the U_1^M and T_2^M signals, and similarly if $w_m(m_1^p, m_2^w)$ is the worst.

The Values of η_t and γ_t

The mediator's ex-ante per round payoff assuming truthtelling is

$$\eta_t = u_m(x^*)A + w_m(m_1^w, m_2^p)B + w_m(m_1^p, m_2^w)C + w_m(m_1^w, m_2^w)D,$$

where

$$\begin{aligned} A = {} & (1 - H_1(\theta_1^*(U_2^M)|U_1^N))(1 - H_2(\theta_2^*(U_1^M)|U_2^N))H_1(\theta^0)H_2(\theta^0) \\ & + (1 - H_1(\theta_1^*(U_2^M)|T_1^N))(1 - H_2(\theta_2^*(T_1^M)|U_2^N))(1 - H_1(\theta^0))H_2(\theta^0) \\ & + (1 - H_1(\theta_1^*(T_2^M)|U_1^N))(1 - H_2(\theta_2^*(U_1^M)|T_2^N))H_1(\theta^0)(1 - H_2(\theta^0)) \\ & + (1 - H_1(\theta_1^*(T_2^M)|T_1^N))(1 - H_2(\theta_2^*(T_1^M)|T_2^N))(1 - H_1(\theta^0))(1 - H_2(\theta^0)). \end{aligned}$$

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

$$B = H_1\left(\theta_1^*\left(U_2^M\right)|U_1^N\right)\left(1-H_2\left(\theta_2^*\left(U_1^M\right)|U_2^N\right)\right)H_1(\theta^0)H_2(\theta^0)$$
$$+H_1\left(\theta_1^*\left(U_2^M\right)|T_1^N\right)\left(1-H_2\left(\theta_2^*\left(T_1^M\right)|U_2^N\right)\right)\left(1-H_1(\theta^0)\right)H_2(\theta^0)$$
$$+H_1\left(\theta_1^*\left(T_2^M\right)|U_1^N\right)\left(1-H_2\left(\theta_2^*\left(U_1^M\right)|T_2^N\right)\right)H_1(\theta^0)\left(1-H_2(\theta^0)\right)$$
$$+H_1\left(\theta_1^*\left(T_2^M\right)|T_1^N\right)\left(1-H_2\left(\theta_2^*\left(T_1^M\right)|T_2^N\right)\right)\left(1-H_1(\theta^0)\right)\left(1-H_2(\theta^0)\right).$$

$$C = \left(1-H_1\left(\theta_1^*\left(U_2^M\right)|U_1^N\right)\right)H_2\left(\theta_2^*\left(U_1^M\right)|U_2^N\right)H_1(\theta^0)H_2(\theta^0)$$
$$+\left(1-H_1\left(\theta_1^*\left(U_2^M\right)|T_1^N\right)\right)H_2\left(\theta_2^*\left(T_1^M\right)|U_2^N\right)\left(1-H_1(\theta^0)\right)H_2(\theta^0)$$
$$+\left(1-H_1\left(\theta_1^*\left(T_2^M\right)|U_1^N\right)\right)H_2\left(\theta_2^*\left(U_1^M\right)|T_2^N\right)H_1(\theta^0)\left(1-H_2(\theta^0)\right)$$
$$+\left(1-H_1\left(\theta_1^*\left(T_2^M\right)|T_1^N\right)\right)H_2\left(\theta_2^*\left(T_1^M\right)|T_2^N\right)\left(1-H_1(\theta^0)\right)\left(1-H_2(\theta^0)\right).$$

and

$$D = H_1\left(\theta_1^*\left(U_2^M\right)|U_1^N\right)H_2\left(\theta_2^*\left(U_1^M\right)|U_2^N\right)H_1(\theta^0)H_2(\theta^0)$$
$$+H_1\left(\theta_1^*\left(U_2^M\right)|T_1^N\right)H_2\left(\theta_2^*\left(T_1^M\right)|U_2^N\right)\left(1-H_1(\theta^0)\right)H_2(\theta^0)$$
$$+H_1\left(\theta_1^*\left(T_2^M\right)|U_1^N\right)H_2\left(\theta_2^*\left(U_1^M\right)|T_2^N\right)H_1(\theta^0)\left(1-H_2(\theta^0)\right)$$
$$+H_1\left(\theta_1^*\left(T_2^M\right)|T_1^N\right)H_2\left(\theta_2^*\left(T_1^M\right)|T_2^N\right)\left(1-H_1(\theta^0)\right)\left(1-H_2(\theta^0)\right).$$

The ex-ante likelihood of getting into the next round is

$$\gamma_t = 1-\{(1-H_1(\theta^0))[H_1\left(\theta_1^*\left(U_2^M\right)|T_1^N\right)H_2(\theta^0)$$
$$+H_1\left(\theta_1^*\left(T_2^M\right)|T_1^N\right)(1-H_2(\theta^0))]$$
$$+(1-H_2(\theta^0))[H_2\left(\theta_2^*\left(U_1^M\right)|T_2^N\right)H_1(\theta^0)$$
$$+H_2\left(\theta_2^*\left(T_1^M\right)|T_2^N\right)(1-H_1(\theta^0))]\}.$$

Proof of Theorem 3

An equilibrium is possible when the improvement in the per round payoff from vouching for a player are stronger when in receipt of a T_i^N signal, and the decrease in the likelihood of getting to the next round is lower, than after receipt of a U_i^N signal. First, consider the likelihood of getting to the next round.

We can write out the likelihood of getting to the next round more fully as a function of the signals received and sent as follows.

$$\gamma\left(T_1^M, T_2^M|S_1^N, S_2^N\right) = \left(1-H_1\left(\theta_1^*\left(T_2^M\right)|S_1^N\right)\right)\left(1-H_2\left(\theta_2^*\left(T_1^M\right)|S_2^N\right)\right)$$
$$\gamma\left(U_1^M, T_2^M|S_1^N, S_2^N\right) = 1-H_2\left(\theta_2^*\left(U_1^M\right)|S_2^N\right)$$
$$\gamma\left(T_1^M, U_2^M|S_1^N, S_2^N\right) = 1-H_1\left(\theta_1^*\left(U_2^M\right)|S_1^N\right)$$
$$\gamma\left(U_1^M, U_2^M|S_1^N, S_2^N\right) = 1.$$

The likelihood of getting to the next round is decreasing in the signal sent, S_i^M, regardless of the signal received; that is, vouching for a player introduces some chance that that player's defection will end one's mediation career. The signal received determines how much it decreases, however. The key result is that the likelihood of getting to the next round exhibits decreasing differences in the signals sent and received; that is, $\gamma(U_i^M|U_i^M) - \gamma(T_i^M|U_i^M) > \gamma(U_i^M|T_i^N) - \gamma(T_i^M|T_i^N)$. This means that the decrease in the likelihood of getting to the next round from vouching for a player is smaller if the mediator has received the signal that the other side actually is trustworthy.

Now return to Equation 2, and consider the case where the mediator is biased toward player 1, so $w_m(m_1^w, m_2^p) > w_m(m_1^w, m_2^w)$ and $u_m(x^*) > w_m(m_1^p, m_2^w)$. The mediator's payoff is increasing in S_1^M, because sending the T_1^M signal causes player 2 to be more likely to cooperate, increasing the likelihood of mutual cooperation and player 1 exploiting player 2. However, if $u_m(x^*) > w_m(m_1^w, m_2^p)$, the payoff exhibits increasing differences; that is, $\eta(T_1^M, S_2^M|U_1^N, S_2^N) - \eta(U_1^M, S_2^M|U_1^N, S_2^N) < \eta(T_1^M, S_2^M|T_1^N, S_2^N) - \eta(U_1^M, S_2^M|T_1^N, S_2^N)$. In other words, the increase in the payoff of vouching for player 1 is stronger when the signal received indicates that player 1 is trustworthy, than when the signal indicates otherwise. The same condition holds if the mediator is issue indifferent.

Putting these two facts together means that after receiving the U_1^N signal, the increase in the payoff for deviating from the U_1^M signal to the T_1^M signal will be small, and the decrease in the likelihood of getting to the next round will be large. This provides a disincentive to lie and means that the condition in Equation 3 may hold. Conversely, if the T_1^N signal is received, the payoff loss from deviating from T_1^M to U_1^M will be large and the increase in likelihood of getting to the next round will be small, once again punishing dishonesty and making Equation 3 more likely to be satisfied. These considerations have the potential to make the overall payoff conditional on the signals received, making truth-telling equilibria possible.

REFERENCES

Akerlof, George A. 1970. "The Market for "Lemons": Quality, Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (August): 488–500.

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 (October): 297–321.

Biglaiser, Gary. 1993. "Middlemen as Experts." *RAND Journal of Economics* 24 (Summer): 212–23.

Blainey, Geoffrey. 1988. *The Causes of War*. Third edition. New York: The Free Press.

Brecher, Michael, and Jonathan Wilkenfeld. 2000. *A Study of Crisis*. Ann Arbor: University of Michigan Press.

Burton, John W. 1969. *Conflict and Communication: The Use of Controlled Communication in International Relations*. New York: The Free Press.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press.

Coll, Steve. 2006 "The Stand-Off." *The New Yorker*: 13 (February): 126–39.

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

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

Fisher, Ronald J. 1972. "Third Party Consultation: A Method for the Study and Resolution of Conflict." *Journal of Conflict Resolution* 16 (March): 67–94.

Fisher, Ronald J. 1983. "Third Party Consultation as a Method of Intergroup Conflict Resolution: A Review of Studies." *Journal of Conflict Resolution* 27 (June): 301–34.

Fisher, Ronald J., and Loraleigh Keashly. 1991. "The Potential Complementarity of Mediation and Consultation within a Contingency Model of Third Party Intervention." *Journal of Peace Research* 28 (February): 29–42.

Forges, Francoise. 1986. "An Approach to Communication Equilibria." *Econometrica* 54 (November): 1375–85.

Fukuyama, Francis. 1995. *Trust: The Social Virtues and the Creation of Prosperity*. New York: The Free Press.

Gambetta, Diego, ed. 1988. *Trust: Making and Breaking Cooperative Relations*. New York: Basil Blackwell.

Glaser, Charles. 1997. "The Security Dilemma Revisited." *World Politics* 50 (October): 171–202.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press

Hagerty, Devin T. 1998. *The Consequences of Nuclear Proliferation: Lessons from South Asia*. Cambridge, MA: MIT Press.
Hardin, Russell. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.
Herz, John H. 1950. "Idealist Internationalism and the Security Dilemma." *World Politics* 2 (January): 157–80.
Hobbes, Thomas. 1968 (1651). *Leviathan*. New York: Penguin.
Jarque, Xavier, Clara Ponsati, and Jozsef Sakovics. 2003. "Mediation: Incomplete Information Bargaining with Filtered Communication." *Journal of Mathematical Economics* 39 (September): 803–30.
Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30 (January): 167–214.
Kelman, Herbert C. 1997. "Some Determinants of the Oslo Breakthrough." *International Negotiation* 2 (September): 183–94.
Kelman, Herbert C. 2000. "The Role of the Scholar-Practitioner in International Conflict Resolution." *International Studies Perspectives* 1 (December): 273–88.
Kissinger, Henry A. 1982. *Years of Upheaval*. Boston: Little, Brown and Company.
Kriesberg, Louis. 2001. "Mediation and the Transformation of the Israeli-Palestinian Conflict." *Journal of Peace Research* 38 (May): 373–92.
Kydd, Andrew. 2003. "Which Side Are You On? Bias, Credibility and Mediation," *American Journal of Political Science* 47 (October): 597–611.
Kydd, Andrew H. 2005. *Trust and Mistrust in International Relations*. Princeton: Princeton University Press.
Landau, Daryl, and Sy Landau. 1997. "Confidence Building Measures in Mediation." *Mediation Quarterly* 15 (Winter): 97–103.
Larson, Deborah Welch. 1997. *Anatomy of Mistrust: U.S.-Soviet Relations During the Cold War*. Cornell Studies in Security Affairs Ithaca, NY: Cornell University Press.
Levy, Haim. 1992. "Stochastic Dominance and Expected Utility: Survey and Analysis," *Management Science* 38 (April): 555–93.
Lizzeri, Alessandro. 1999. "Information revelation and certification intermediaries." *RAND Journal of Economics* 30 (Summer): 214–31.
McFadden, Daniel. 1976. "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement* 5 (Summer): 363–90.
Milgrom, Paul, and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62 (January): 157–80.
Milgrom, Paul, and John Roberts. 1994. "Comparing Equilibria." *American Economic Review* 84 (June): 441–59.
Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Medieval Law Merchant, Private Judges and Champaign Fairs." *Economics and Politics* 2 (March): 1–23.
Myerson, Roger. 1986. "Multistage Games with Communication." *Econometrica* 54 (March): 323–58.
Myerson, Roger, 1991. "Analysis of Incentives in Bargaining and Mediation." In *Nesotiation Analysis*, ed. H. Peyton Young. Ann Arbor. University of Michigan Press, 67–85.
O'Neill, Barry. 2003. "Mediating National Honor: Lessons from the Era of Duelling." *Journal of Institutional And Theoretical Economics* 159 (March): 229–47.
Posen, Barry R. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35 (Spring): 27–47.
Powell, Robert. 2002. "Bargaining Theory and International Conflict." *Annual Review of Political Science* 5: 1–30.
Powell, Robert. 2004. "Bargaining and Learning while Fighting." *American Journal of Political Science* 48 (April): 344–61.
Princen, Thomas. 1991. "Camp David: Problem Solving or Power Politics As Usual." *Journal of Peace Research* 28 (February): 57–69.
Princen, Thomas. 1992. *Intermediaries in International Conflict*. Princeton: Princeton University Press.
Rauchhaus, Robert W. 2005. "Asymmetric Information, Mediation and Conflict Management." Manuscript, University of California Santa Barbara.
Ross, William H., and Carole Wieland. 1996. "Effects of Interpersonal Trust and Time Pressure on Managerial Mediation Strategy in a Simulated Organizational Dispute." *Journal of Applied Psychology* 81 (May): 228–48.
Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica* 50 (January): 97–109.
Schmidt, Holger. 2004. "When (and Why) Do Brokers Have to be Honest? Impartiality and Third Party Support for Peace Implementaion after Civil Wars 1945–1999." Manuscript. Georgetown University.
Schmidt, Klaus D. 2003. "On the Covariance of Monotone Functions of a Random Variable." Manuscript. Technische Universität Dresden.
Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93 (June): 279–97.
Smith, Alastair, and Alan Stam. 2003. "Mediation and Peacekeeping in a Random Walk Model of Civil and Interstate War." *International Studies Review* 5 (December): 115–35.
Spulber, Daniel F. 1996. "Market Microstructure and Intermediation." *Journal of Economic Perspectives* 10 (Summer): 135–52.
Touval, S. 1975. "Biased Intermediaries: Theoretical and Historical Considerations." *Jerusalem Journal of International Relations* 1 (Fall): 51–69.
Touval, Saadia, and I. William Zartman. 1989. "Mediation in International Conflicts." In *Mediation Research: The Process and Effectivencess of Third Party Intervention*, ed. Kenneth Kressel and Dean G. Pruitt. Hoboken: Jossey-Bass, 155–37.
Wagner, R. Harrison. 2000. "Bargaining and War." *American Journal of Political Science* 44 (July): 469–84.
Wall, James A., John B. Stark, and Rhetta L. Standifer. 2001. "Mediation: A Current Review and Theory Development." *Journal of Conflict Resolution* 45: 370–91.
Walter, Barbara F. 2002. *Committing to Peace: The Successful Settlement of Civil Wars*. Princeton: Princeton University Press.
Walton, Richard E. 1969. *Interpersonal Peacemaking: Confrontations and Third Party Consulation*. Reading, MA: Addison-Wesley.
Wehr, Paul, and John Paul Lederach. 1991. "Mediation Conflict in Central America." *Journal of Peace Research* 28 (February): 85–98.
Young, Oran R. 1967. *The Intermediaries: Third Parties in International Crises*. Princeton: Princeton University Press.

https://doi.org/10.1017/S0003055406062290 Published online by Cambridge University Press