Mediation, Arbitration and Negotiation

Maria Goltsman

Department of Economics, University of Western Ontario, Social Science Centre,
London, Ontario N6A 5C2, Canada, mgoltsma@uwo.ca

Johannes Hörner

Kellogg School of Management, Managerial Economics and Decision Sciences, 2001
Sheridan Road, Evanston, IL 60208, USA, j-horner@kellogg.nwu.edu

Gregory Pavlov

Department of Economics, Boston University, 270 Bay State Road, Boston, MA
02215, USA, gpavlov@bu.edu

Francesco Squintani

Department of Economics, Universita' degli Studi di Brescia, Via San Faustino 74B,
25122 Brescia, Italy, squintan@eco.unibs.it

MEDIATION, ARBITRATION AND NEGOTIATION

ABSTRACT. We compare three common dispute resolution processes – negotiation, mediation, and arbitration – in the framework of Crawford and Sobel (1982). Under negotiation, the two parties engage in (possibly arbitrarily long) face-to-face cheap talk. Under mediation, the parties communicate with a neutral third party who makes a non-binding recommendation. Under arbitration, the two parties commit to conform to the third party recommendation. We characterize and compare the optimal mediation and arbitration procedures. Both mediators and arbitrators should optimally filter information, but mediators should also add noise to it. We find that unmediated negotiation performs as well as mediation if and only if the degree of conflict between the parties is low.

1. INTRODUCTION

Mediation is a procedure of dispute resolution broadly adopted in a number of contexts, from international crises, to legal confrontations, and business transactions. Despite the relevance and pervasiveness of mediation, applied experts report that the theoretical characterization of optimal mediation techniques is largely under-developed.[1] This paper characterizes the properties of optimal mediation and arbitration, for the cases in which the source of conflict is informational, and transfers cannot be established to the disputants. Furthermore, we compare the performance of mediation and arbitration with unmediated negotiation.

Despite the possibly complex information transmission strategies that a mediator can adopt, we find that in our context a mediator acts optimally by gathering information from the parties, filtering it, and strategically adding noise to it.[2] Although it is widely

---

[1] The latest survey by Wall, Stark and Standifer (2001) concludes: "What techniques should mediators use? And how can mediation be improved [...]? It is hoped that some of these questions will be studied in the next decade."

[2] The role of mediators in providing disputants with information has been highlighted by Touval (1996). Among other techniques adopted, the mediator can propose threats or punishments, determine which points are negotiable, encourage concessions (Wall, Stark and Standifer, 2001), or attempt to reframe the opponent in a more positive light (Umbreit, 1993).

believed that a successful mediator should establish credible reports to the conflicting parties, we find that a mediator that reports precisely all the information transmitted would not act optimally. At the same, we show that there is benefit in mediation over unmediated negotiation if and only if the degree of conflict between the parties is high. When the degree of conflict is low, the conflicting parties can achieve the optimal outcome of mediation with just two rounds of unmediated negotiation. These findings resonate with well-documented stylized facts in the empirical literature on negotiation (Bercovich and Jackson 2001, Carnevale and Pruitt 1992, Wall and Lynn, 1993), that show that parties are less likely to reach an agreement without a mediator when the intensity of conflict is high than when it is low.

We study conflicts that arise because of informational advantage by one party, and where the opposite party retains the ultimate right of making the final decision. For example, in international conflicts, disproportionate military strength may endow one party with the ability of imposing a final decision. The opposite party may nevertheless have some private information on the best course of action to resolve the crisis, for example on its willingness to accept concessions to avoid military confrontation. The mediator's task is to try and elicit private information and avoid the military option. We model these situations by adopting the standard classic framework by Crawford and Sobel (1982).[3] Among the two parties, we distinguish an *informed party* and an uninformed *decision-maker*. Each party's utility decreases with the distance to his most preferred outcome. There is a conflict of interest: in any state of the world, the informed party's most preferred decision is higher than the decision-maker's.

Within this set-up, we consider the three most common means by which the parties resolve disputes outside of court. In unfacilitated *negotiation*, or cheap talk, the two parties directly and voluntarily exchange information back and forth, and attempt to reach an acceptable agreement, ratified by the decision maker.[4] Under *mediation*, a

---

[3]The model by Crawford and Sobel (1982) has been a foundation for theoretical and applied work on communication in a variety of fields, including political economy (Grossman and Helpman, 2001) and finance (Morgan and Stocken, 2003).

[4]Although Crawford and Sobel considered the case in which the informed party has only one opportunity to send a message to the decision-maker before the latter party makes his choice, it is known that this assumption may be restrictive (see Krishna and Morgan, 2004). Accordingly, when

neutral third party is called to propose an agreement after hearing the arguments of each party. The neutral third party has no authority to impose a settlement, and merely suggests an agreement that must be mutually acceptable. In contrast, under *arbitration*, the neutral third party renders a final decision. While the agreement to arbitrate is voluntary and the arguments presented are unverifiable, the arbitrator's decision is binding, in the sense that courts or military action will enforce it against a possibly reluctant decision maker.

When studying mediation, the revelation principle (Myerson, 1982) implies that it is sufficient to consider a particular kind of mechanism with one round of mediated communication. The informed party makes a report to the mediator, who then makes a recommendation to the informed party. Further, the mediator's strategy is such that the informed party truthfully reveals his private information, and the decision-maker follow the recommendation. We prove that the optimality of an incentive compatible mechanism is equivalent to the optimality (from the informed party's point of view) of the action assigned to the lowest possible state. This result, akin to the revenue equivalence theorem in auction theory (see Myerson, 1981), allows to describe optimal mediated communication protocols, and to compare the optimality properties of mediation with negotiation and arbitration.

We prove that in any optimal mediated communication protocol, the mediator must choose his recommendation randomly for some reports. Further, we demonstrate that the communication protocol developed by Blume and Board (2006) actually achieves the optimum of mediated communication.[5] In this mediation protocol, the informed party reveals the state of the world precisely to the mediator. Subsequently, the mediator filters this information into a coarser description of the state of the world, and optimally adds noise. With some probability, he reports the filtered information, otherwise he sends a completely uninformative message. The decision maker then acts

---

considering negotiation, we allow both parties to engage in an arbitrarily large number of rounds of communication.

[5]Ganguly and Ray (2005) also provide an example of mediation rule that improves upon Crawford and Sobel's equilibrium, and Myerson (2007) even provides a numerical tool to compute such rules. See also Kawamura (2006). None of these papers establish optimality of their examples.

on the basis of reports. The two actions that follow the informative or uninformative message are chosen so as to be identical for the lowest range of report, i.e., the mediator actually sends a deterministic recommendation for sufficiently low reports. This recommendation maximizes the informed party's payoff for the lowest possible state of the world, and thus makes the mechanism optimal.

The intuition for why a successful mediator must randomize over recommendations is simple. A mediator that would be a mere relay or censor of information would be of no value here, as such transmission or censoring of information could be directly performed by the informed party.[6] In particular, a perfectly credible mediator could not improve upon the (most informative) equilibrium outcome by Crawford and Sobel. In our environment, the mediator can only create value by controlling the flow of information between the parties. This role of mediation has already been pointed out in other contexts by Brown and Ayres (1994), Ayres and Nalebuff (1997) and Mitusch and Strausz (2005). In practice, this corresponds to the mediator's role in "collecting and judiciously communicating select confidential material" (Raiffa, 1982, 108-09), and "communicating selected information back and forth" (Singer, 1990). Obviously, the role for mediation that we identify cannot be performed by holding joint, face-to-face sessions with both parties, but requires private and separate caucuses, a practice that is often, but not always, followed by mediators. Indeed, such caucuses are thought to encourage parties to share confidential information with the mediator. Although mediators frequently commit not to share what is learned with the other party, they commonly use these caucuses to explore settlements possibilities. Interestingly, the practical literature on mediation emphasizes the particular usefulness of this kind of mediation in situations in which the intensity of conflict is high (see for example, the survey by Wall and Lynn, 1993). In international affairs, shuttle diplomacy is a related practice that has become popular since Henry Kissinger's efforts in the Middle East

---

[6]Censoring by the mediator may be valuable in situations in which both parties have private information, as the censoring may require knowledge of both reports, and could not be performed by either party on its own. Indeed, in the practice of "confidential listening", each side reveals his settlement position to the mediator in a private caucus, and the mediator then reveals whether there is any overlap or common ground upon which to reach a settlement, without any further detail.

in the early 1970s and the Camp David negotiations mediated by Jimmy Carter, in which a third party conveys information back and forth between parties, providing suggestions for moving the conflict toward resolution (see, for example, Kydd, 2006).

For sake of comparison, we assume that these procedures are designed to maximize the ex ante welfare of the uninformed party, the decision-maker. (In some cases, this turns out to also maximize the ex ante welfare of the informed party.) In particular, we exactly characterize the performance of both optimal arbitration and mediation. Clearly, mediation dominates negotiation: any outcome that can be achieved without a mediator can also be achieved with a mediator that plainly reports the informed party's messages to the decision-maker. In turn, arbitration dominates mediation, as any outcome that can be achieved by a mediator can also be achieved by an arbitrator that always makes a recommendations in agreement with the final decision of the decision maker.

Most importantly, we show that, relative to negotiation, mediation performs better when the intensity of conflict is high, but yields the same ex-ante welfare when the intensity of conflict is low.[7] Furthermore, we find that for small enough intensities, there is no need for more than two rounds of communication to achieve the mediated communication optimum. In this case the optimum is precisely achieved by the optimal 'monotonic' equilibrium proposed by Krishna and Morgan (2004). When the intensity of conflict is intermediate, instead, the beneficial role of the mediator cannot be replicated by however many rounds on unmediated communication.

Arbitration has been already studied in the literature on delegation (Holmström, 1977; Melumad and Shibano, 1991; Alonso and Matouschek, 2004 and 2006). However, all these earlier contributions have restricted attention to deterministic mechanisms. Since we show that the optimal mediation scheme is stochastic, a comparison of the optimal properties of both procedures calls for allowing stochastic arbitration

---

[7]When the intensity of conflict is too high, all the three mechanisms fail to induce any information transmission and induce the same outcome.

mechanisms as well. We prove here that the optimal protocol is indeed deterministic, and is therefore the one identified in the earlier papers.[8] We conclude by noting that arbitration always performs better than negotiation and mediation. This result supports widespread views that a third party intermediary able to force an agreement on the parties may be more effective than one acting as a mere consultant (see, for example, the surveys by Wall and Lynn, 1993 and by Wall, Stark and Standifer, 2001).

The paper is organized as follows. Section 2 introduces the set-up and formally defines optimal arbitration, mediation and negotiation. Section 3 studies arbitration. Section 4 analyzes mediation, and Section 5 examines negotiation. Concluding comments are in Section 6. All formal proofs are in the appendix.

2. MODEL

There are two players, the informed party and the decision-maker. The payoffs of both players depend on the state of nature $\theta \in \Theta = [0, 1]$ and the action $y \in Y = \Re$. The informed party knows θ; the decision-maker does not know θ, and his prior is uniform on Θ. While the informed party is better informed about the state of nature, he lacks the decision-making ability: that is, y is chosen by the decision-maker.

We assume that the utility function of the decision-maker equals $v(y, \theta) = -(y - \theta)^2$, and that of the informed party equals $u(y, \theta) = -(y - (\theta + b))^2$ where $b > 0$. For any given θ, the informed party's preferred action is $y = \theta + b$, while the decision-maker's preferred action is $y = \theta$. The utility of each party in state θ decreases in the distance from the preferred action given θ to the action that is actually taken.

In this setting, we will study three different classes of communication procedures: arbitration, mediation and negotiation. Under arbitration the parties communicate privately or publicly with a neutral trustworthy arbitrator, who then enforces a binding decision. Mediation is similar to arbitration, but the recommendation of the mediator is not binding: the decision-maker is free to disobey. Finally, under unmediated

---

[8]In coincident work, and following a different approach, Kováč and Mylovanov (2007) generalize this last result to more general environments.

negotiation, the parties cannot employ a mediator or arbitrator, and communicate only by sending direct messages to each other.

Let us formally introduce the optimization problems that are solved in each of the three cases, looking at arbitration first. The revelation principle (Myerson, 1982) allows us to restrict attention to arbitration protocols whereby the informed party reports of the state of the world to the arbitrator, who makes the final decision. Further, in equilibrium the informed party finds it optimal to announce the true state. Let an arbitration rule be a probability distribution p on $Y \times \Theta$, with the interpretation that $p(y, \theta)$ is the joint probability that the informed party reports state θ to the arbitrator and the arbitrator recommends action y to the decision-maker.

Definition 1. *An **optimal arbitration rule** p is a probability measure on $Y \times \Theta$ that solves the following problem:*

$$\max_{p(\cdot)} V = -\int_{Y \times \Theta} (y - \theta)^2 \, p(dy, d\theta)$$

subject to

$$(IC - IP) \qquad \theta = \arg\max_{\hat{\theta} \in \Theta} \left[-\int_Y (y - (\theta + b))^2 \, p\left(dy, \hat{\theta}\right) \right], \ \forall \theta \in \Theta;$$

$$(PROB) \qquad 1 = \int_Y p(dy, \theta), \ \forall \theta \in \Theta.$$

The constraint $(PROB)$ reflects the fact that θ is uniformly distributed on Θ. The constraint $(IC - IP)$ reflects the fact that the informed party should find it optimal to tell the truth. An arbitration rule that satisfies $(IC - IP)$ is called incentive compatible.

By the revelation principle, one can restrict attention to mediation protocols whereby the informed party reports the state of the world to the mediator, and the mediator makes a recommendation to the decision-maker. Further, we can assume without loss

of generality that the report is truthful, and the recommended action is incentive compatible (see Myerson, 1982). Formally, a mediation rule is a probability distribution p on $Y \times \Theta$, with the interpretation that $p(y, \theta)$ is the joint probability that the informed party reports state θ to the mediator and the mediator recommends action y to the decision-maker.

Definition 2. *An **optimal mediation rule** p is a probability measure on $Y \times \Theta$ that solves the following problem:*

$$\max_{p(\cdot)} V = - \int_{Y \times \Theta} (y - \theta)^2 \, p(dy, d\theta)$$

subject to

$$(IC - IP) \qquad \theta = \arg\max_{\hat{\theta} \in \Theta} \left[-\int_Y (y - (\theta + b))^2 \, p\left(dy, \hat{\theta}\right) \right], \ \forall \theta \in \Theta;$$

$$(PROB) \qquad 1 = \int_Y p(dy, \theta), \ \forall \theta \in \Theta;$$

$$(IC - DM) \qquad y = \frac{\int_\Theta \theta p(y, d\theta)}{\int_\Theta p(y, d\theta)}, \ \forall y \in Y \text{ s.t. } \int_\Theta p(y, d\theta) > 0.$$

The constraint $(IC - DM)$ states that the decision-maker never has an incentive to deviate from an action that is prescribed to him by the mediator (the right-hand side of the equality is the expectation of θ given the recommendation y, which is the action that maximizes the decision-maker's payoff when the mediator recommends y). A mediation rule that satisfies $(IC - IP)$ and $(IC - DM)$ is called incentive compatible.

Finally, negotiation means that the informed party and the decision-maker engage in several rounds of unmediated communication, sending a message to the other party at each round. Similarly to Aumann and Hart (2003), a negotiation protocol will include two sets, I and D, and $T \in N \cup \{\infty\}$, where I and D are the sets of admissible messages of the informed party and the decision-maker, respectively, and T is the length of the protocol. The protocol will define a game with incomplete information

with $T + 2$ stages that proceeds as follows. At stage 0, Nature selects the state θ and informs the informed party. At each of the stages $1, ..., T$, the informed party and the decision-maker simultaneously choose a message, and their choices become commonly known at the end of the stage. At stage $T + 1$, the decision-maker selects an action. The payoffs for the decision-maker and the informed party are $v(y, \theta)$ and $u(y, \theta)$ respectively, where y is the action, and θ is the true state of nature. A negotiation protocol will be called finite if $T < \infty$.[9]

Definition 3. *An **optimal negotiation protocol** (I, D, T) solves the following problem:*

$$\max_{p(\cdot), I, D, T} V = - \int_{Y \times \Theta} (y - \theta)^2 \, p(dy, d\theta)$$

subject to

$$p \text{ is the outcome distribution of a Bayesian-Nash equilibrium}$$

$$of \text{ the game induced by the protocol } (I, D, T)$$

It is immediate from the statement of these optimization problems that any outcome that can be achieved with mediation can be replicated with arbitration. It follows that the decision-maker always at least weakly gains from arbitration relative to mediation. Also, the revelation principle implies that any equilibrium of any negotiation protocol is outcome equivalent to a truthful equilibrium of some incentive compatible mediation rule. It follows that the decision-maker always at least weakly gains from mediation relative to negotiation.

The following fact is also worth noting. Consider any feasible mediation rule, and let $V = - \int_{Y \times \Theta} (y - \theta)^2 p(dy, d\theta)$ be the ex-ante expected utility of the decision-maker, and $U(\theta) = - \int_Y (y - (\theta + b))^2 p(dy, \theta)$ be the expected utility of the informed party given θ when the mediation rule p is in place. Crawford and Sobel (1982) prove that

$$V = E_\theta U(\theta) + b^2$$

---

[9]See Forges (1986) for an early application to economics of long negotiation, and Gerardi (2004) for recent work on unmediated communication.

As a consequence, an incentive compatible mediation rule p ex ante Pareto dominates an incentive compatible rule q if and only if the decision-maker's ex-ante expected utility under p is higher than under q. This fact will allow us to maximize the expected utility of the decision-maker, with the understanding that the resulting mediation rule will be Pareto optimal. Since any equilibrium of any negotiation protocol is outcome equivalent to a truthful equilibrium of some incentive-compatible mediation rule, the same statement holds also for the optimal negotiation protocol.

3. ARBITRATION

To solve for the optimal arbitration rule, we first develop a tractable way to deal with the incentive compatibility constraint for the informed party. Let $y\left(\widehat{\theta}\right) = \int_Y y p\left(dy, \widehat{\theta}\right)$ and $\sigma^2\left(\widehat{\theta}\right) = \int_Y \left(y - y\left(\widehat{\theta}\right)\right)^2 p\left(dy, \widehat{\theta}\right)$ be the conditional expectation and the variance of y given a message $\widehat{\theta}$. Then an expected payoff of the informed party of type θ who reported a message $\widehat{\theta}$ in the mechanism p is

$$\int_Y -(y-(\theta+b))^2 p\left(dy, \widehat{\theta}\right) = -\sigma^2\left(\widehat{\theta}\right) - \left(y\left(\widehat{\theta}\right) - (\theta+b)\right)^2$$

Namely, the fact that the informed party has a quadratic loss function implies that the informed party cares only about the expectation and the variance of the action.

One apparent benefit of this representation is that the constraint $(IC - IP)$ can be stated in terms of $(y(\theta), \sigma^2(\theta))$ only. In addition, notice that the variance of y enters the utility function of the informed party in a quasi-linear way, and thus it does not interact with the type θ. Taking advantage of this fact, we can show (Lemma 2 in Appendix A1) that the incentive compatibility for the informed party is equivalent to two conditions: the expected action is non-decreasing in the state, and the informed party's equilibrium payoff in any state θ can be expressed as a function of his payoff in state 0 and of the expected action in the states below θ. This result is analogous to a well-known result in mechanism design for environments where the preferences are quasi-linear in money.

This representation allows us to prove the following theorem.

Theorem 1. *The optimal arbitration rule selects the preferred action of the informed party in the set* $\left[0, \max\left\{1-b, \frac{1}{2}\right\}\right]$. *Formally, it satisfies:*

$$
y\left(\theta\right) \;=\; \begin{cases} \theta + b, & if \quad \theta \in [0, \max\{1 - 2b, 0\}) \\ \max\left\{1 - b, \frac{1}{2}\right\}, & if \quad \theta \in [\max\{1 - 2b, 0\}, 1] \end{cases} \;;
$$

$$
\sigma^2\left(\theta\right) \;=\; 0, \;\forall \theta \in [0, 1]\,;
$$

$$
U\left(0\right) \;=\; \begin{cases} 0, & if \quad b \leq \frac{1}{2}; \\ -\left(\frac{1}{2} - b\right)^2, & if \quad b > \frac{1}{2}. \end{cases}
$$

Observe that when the preference divergence parameter b is above $\frac{1}{2}$, the optimal arbitration rule is a flat one (the same decision is enforced no matter what the informed party reports). For future reference notice that for these values of the parameter communication is useless in the arbitration model, and, consequently, it is useless in the mediation and negotiation models.

When $b \leq \frac{1}{2}$, the optimal arbitration rule is deterministic. It implements the most preferred action of the informed party for low states of the world, and is constant at $1 - b$ for high states of the world. See Figure 1 for an illustration.
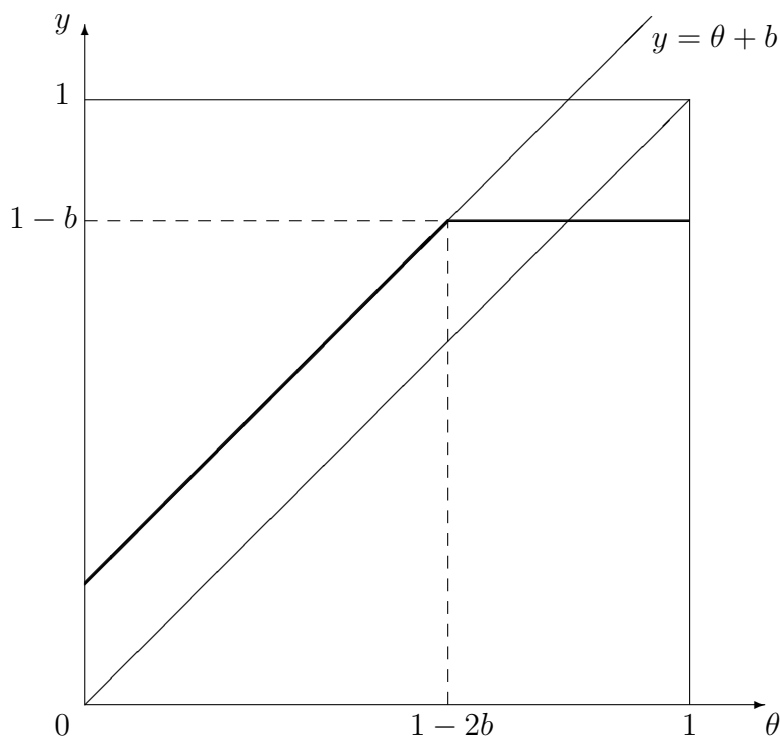
Figure 1 : Optimal arbitration

Melumad and Shibano (1991) already established the optimality of such a rule among deterministic mechanisms. Optimization over deterministic mechanisms amounts to a choice of a set of actions for an informed party to choose from. The optimal mechanism can be viewed as a delegation of the decision to the informed party with a limited form of discretion: the informed party can enforce any decision he likes, as long as it does not exceed $1 - b$. Since the informed party's most preferred action in any state of the world is higher than that of the decision maker, it pays to impose an upper bound on the allowable actions. On the other hand, it turns out that the best way to make use of the informed party's information in case of the low states is to grant a complete freedom of choice of the action to the informed party.[10]

---

[10]For additional intuition and results on optimal delegation see also Holmström (1977), Alonso and Matouschek (2004, 2006).

Our result demonstrates that this delegation rule remains optimal even if we allow for stochastic mechanisms. The tradeoff here is between an implementation of expected action functions which are more desirable for the decision maker and incentive costs due to an increased variance of the mechanism. It turns out that this tradeoff is always resolved in favor of using mechanisms with the smallest possible variance, i.e. deterministic mechanisms. The intuition for this is in Appendix A3.

4. MEDIATION

In this section, we look for the optimal mediation rule. We first note that the optimal arbitration rule (always recommending the action $y = \frac{1}{2}$) is feasible when $b > \frac{1}{2}$. Since the mediation problem is more constrained than the arbitration problem, this rule also has to be the optimal mediation rule. So we focus on finding a solution for $b \in \left(0, \frac{1}{2}\right]$.

We will proceed as follows. First, we will derive an upper bound on the objective function. Next, we show that some of the mechanisms already proposed in the literature achieve this upper bound for certain values of b.

Lemma 1.

(a) If a mediation rule p is incentive compatible, then $V \leq -\frac{1}{3}b(1 - b)$;

(b) An incentive compatible mediation rule is optimal if and only if $U(0) = 0$.

Lemma 1 can be compared to the revenue equivalence theorem in standard mechanism design. However, while in the revenue equivalence theorem, the revenue is pinned down by the utility of the lowest type and the allocation, here the welfare from an incentive compatible mechanism is determined only by the utility of the lowest type. The reason for this difference is that in our problem, the mechanism designer is facing two sets of incentive compatibility constraints, one for the informed party and one for the decision-maker. The incentive compatibility constraints for the decision-maker allow us to express $y(\theta)$, which plays the role of an "allocation", as a function of $U(0)$, the lowest type's utility.

Lemma 1 immediately implies that some of the procedures that have been proposed in the literature as improvements upon one-shot negotiation are, in fact, optimal. One of them is described below.

Theorem 2. *For every $b < \frac{1}{2}$, an optimal mediation rule is such that the mediator randomizes between two actions in each state. With some probability μ, he recommends action b, and with probability $1 - \mu$ he recommends action a_i when $\theta \in [\theta_i, \theta_{i+1})$, $i = 0, ..., N$, where*

$$
\begin{aligned}
\theta_0 &= 0; \\
\theta_i &= 2bi^2 - \left(2bN^2 - 1\right) \frac{2i - 1}{2N - 1}, \ i = 1, ..., N; \\
a_i &= b(i+1) - 2bi(N - i) + \frac{(2 - b)i}{2N - 1}, \ i = 0, ..., N - 1, \\
\mu &= 1 - \frac{1 - 2b}{4(1 - b)} \left(\frac{1}{N - 1} - \frac{1}{N} - \frac{2 - b}{bN - 1} + \frac{2 - b}{bN - b + 1} \right)
\end{aligned}
$$

and N is such that

$$
\frac{1}{2N^2} \leq b < \frac{1}{2(N - 1)^2}
$$

It is straightforward to verify that this mediation rule is feasible and achieves $V = -\frac{1}{3}b(1 - b)$. See Figure 2 for an illustration.
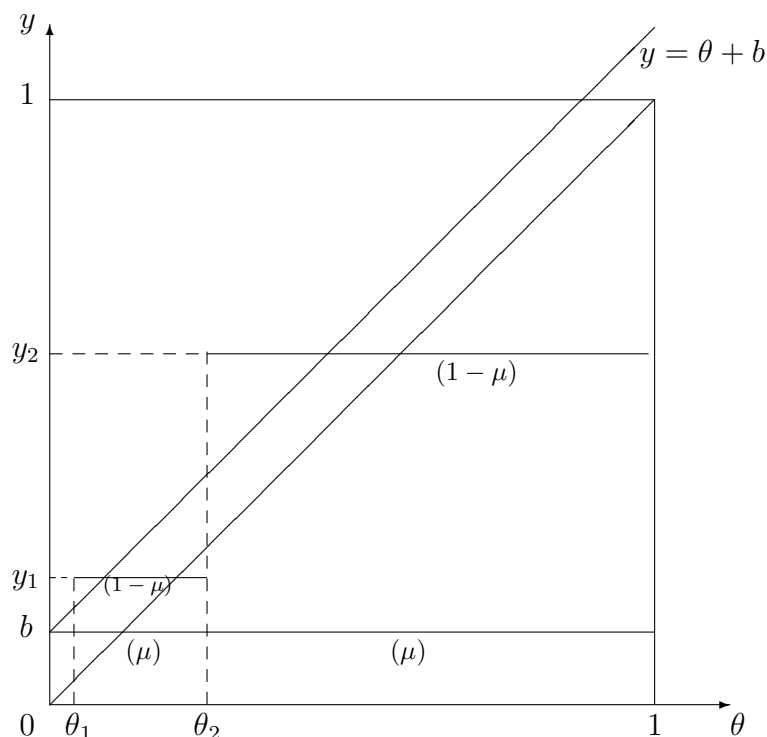
Figure 2 : Optimal mediation

It is immediate to verify that, as the bias tends to zero, so does the probability μ. However, μ is not monotonic in the bias. Rather, for each value of N, it is concave and equal to zero for the two extreme values of bias that are consistent with N. For these extreme (and nongeneric) values, the mediation rule replicates the most informative equilibrium of Crawford and Sobel.

The above rule appears in Blume and Board (2006), who propose it as an improvement upon the most informative Crawford and Sobel equilibrium, but do not prove that it is optimal, or interpret it as a mediation procedure. They propose the following simple interpretation. Imagine that the informed party sends one message from the interval $[0, 1]$ to the decision-maker, but the decision-maker gets his message only with probability $1 - \mu$: with probability μ, the message that the decision-maker gets is a

random draw from the uniform distribution on $[0, 1]$. It is straightforward to show that this procedure is equivalent to the mediation rule formally described above.

Theorem 2 highlights the fact that the primary role of the mediator is to filter the information provided by the informed party. In fact, the optimal mediator's only function is to introduce noise into communication. As Blume and Board (2006) note, introducing noise can have opposing effects on the amount of information transmitted. On the one hand, the direct effect of noise is to make the message received by the decision-maker less informative. On the other hand, the presence of noise relaxes the incentive compatibility for the informed party by weakening the link between his message and the decision-maker's reaction, which makes it easier to motivate the informed party to transmit more information. Theorem 2 shows that the second effect dominates the first one; moreover, simply introducing an optimal amount of noise into communication is optimal in the class of all mediation rules.

The optimal mediation rule described above is not unique. In particular, Lemma 1 implies that another mediation rule that has been proposed in the literature is also optimal when $b \leq \frac{1}{8}$. This is the mediation rule of Krishna and Morgan (2004), which can be implemented with two rounds of cheap talk and is discussed in more detail in the next section.

Observe also that in the optimal mediation problem the constraints are convex in p, and the objective function is linear. Therefore, the set of optimal mediation rules is convex, so that there is in general a continuum of optimal mediation rules.

5. Negotiation

The central result of this section is the following one.

Theorem 3. *Finite negotiation achieves the optimal mediated outcome if and only if* $b \leq 1/8$.

The "if" part of the theorem is easy to show on the basis of our Lemma 1. Krishna and Morgan (2004)'s optimal 'monotonic' equilibrium in a two-period negotiation protocol exists if and only if $b \leq 1/8$, and achieves value $U(0) = 0$: the type-0 informed

party achieves the optimal utility. In light of Lemma 1, for a mediation scheme to be optimal, it is necessary and sufficient that $U(0) = 0$. Hence the optimal monotonic two-period negotiation equilibrium by Krishna and Morgan performs as well as the optimal mediation scheme.

For purposes of illustration, we sketch the construction of the optimal 'monotonic' equilibrium developed by Krishna and Morgan (2004). The reader is referred to that paper for the details in the construction. In the first period of the negotiation protocol, the informed party signals whether the state is above or below some threshold θ^*. Simultaneously, the informed party and the decision-maker exchange messages in a meeting, so as to emulate a public randomization device with probabilities p and $1-p$.[11] In the second round of communication, if the informed party's message indicates that the state is below θ^*, a partitional equilibrium is played, as in Crawford and Sobel (1982). Given the number of elements in the partition N such that $1/\left[2\left(N+1\right)^2\right] \leq b < 1/[2\left(N\right)^2]$, and the set of thresholds $\{\theta_i : i = 0, \ldots, N-1\}$, with $\theta_i = 2bi^2$, $\theta_{N-1} = \theta^*$, the informed party reports in which interval $[\theta_i, \theta_{i+1}]$ the state lies, and the decision-maker takes the corresponding action $a_i = [\theta_i + \theta_{i+1}]/2$. These thresholds assure that $\theta_0 = 0$, $\theta_1 = 2b$ and hence $a_0 = b$ so that, optimally, $U(0) = 0$.

If the informed party's message reported that the state is above $\theta^* = \theta_{N-1}$ in the first round of communication, the continuation play depends on the outcome of the simultaneous exchange of messages. With some probability p, no further communication occurs and the decision-maker takes his action accordingly: $a_{N-1}^* = [1+\theta_{N-1}]/2$. With probability $1-p$, the informed party further signals whether the state is in some lower interval $[\theta_{N-1}, \theta_N]$, or upper interval $[\theta_N, 1]$, upon which the decision-maker takes his action: $a_{N-1} = [\theta_{N-1} + \theta_N]/2$ and $a_N = [1+\theta_N]/2$ respectively.

Krishna and Morgan prove that such equilibria exist for $b \leq 1/8$ with the values θ_N and p that satisfy the following two indifference conditions. The type-θ_N sender

---

[11]Such meetings in which parties simultaneously exchange messages is called a jointly controlled lottery. The reader is referred to Aumann and Hart (2003) and Krishna and Morgan (2004) for a formal definition.

is indifferent between the outcome a_{N-1} and a_N, and the type-θ_{N-1} sender is indifferent between the outcome a_{N-2} and the lottery determining the outcome a_{N-1} with probability $(1-p)$ and the outcome a_{N-1}^* with probability p.

The "only if" part is considerably more involved, and its proof is relegated to the appendix. To gain some intuition for this part, recall from Lemma 1 that for a mediation scheme to be optimal, it is necessary and sufficient that the lowest informed party's type be mapped into the action b (with probability 1). Since preferences are quadratic, any lottery y over actions can be summarized by its first two moments. We may thus represent the preferences of the different informed party's types by their indifference curves in the plane defined by the lotteries' expectation and variance. Figure 3 shows the indifference curves going through the lottery that is degenerate on the action b for two different types $\theta' > \theta$. Observe that if the informed party's type θ is indifferent between the (degenerate) lottery b and some nondegenerate lottery, then type θ' strictly prefers this nondegenerate lottery to b. To put it differently, if in some equilibrium the informed party's type θ is mapped into the action b, then so must be all lower types. Furthermore, in equilibrium, there can be at most one type indifferent between the action b and some nondegenerate lottery.
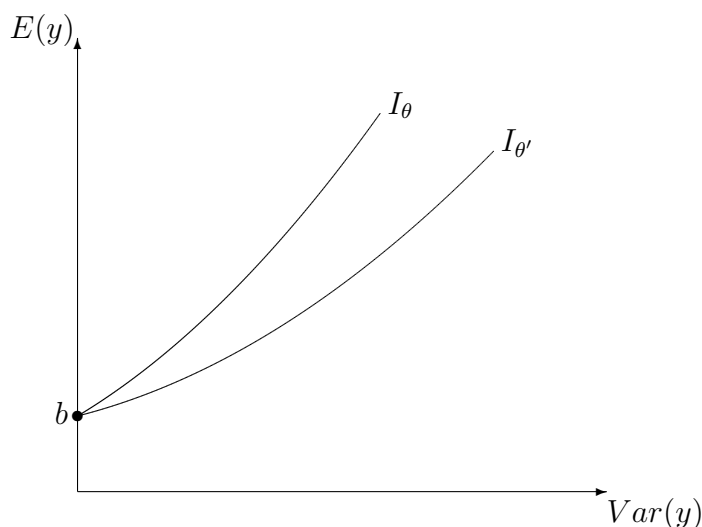
Figure 3: Indifference curves for θ, θ' $(\theta' > \theta)$

Suppose that, in some equilibrium, the action b is chosen when the state is 0, so that the optimal mediated outcome is achieved. Then it must be that, for some θ^*, the action b is finally chosen for almost all states in $[0, \theta^*]$. Furthermore, the action b cannot be finally adopted with positive probability when the state is larger than θ^*. For this to conform with the decision-maker's equilibrium beliefs, it must be that $\theta^* = 2b$.

Consider now the choice of the type-θ^* informed party. By following any of the strategies that achieves action b, the informed party bears a cost $(b - \theta^* - b)^2 = (\theta^*)^2$. But this type of informed party may be better off by upsetting the equilibrium, deviating and mimicking the higher types. At every information set, the informed party can always choose the message that leads to the lottery over actions with the lowest expected value, among all messages that do not necessarily lead to the action b. All the actions that may be realized by this lottery are above b, which is the action supposed to be associated in equilibrium with informed party types smaller than θ^*. Further, such a strategy must eventually lead to actions no larger than $(\theta^* + 1)/2$ – the expected

value over types in $(\theta^*, 1)$. Therefore, we must have

$$(\theta^*)^2 \leq \left(\frac{\theta^* + 1}{2} - \theta^* - b \right)^2,$$

$$\text{or } b \leq 1/8,$$

because $\theta^* = 2b$.

More succinctly, the 'only if' part of the above theorem follows because quadratic preferences imply that such an equilibrium be monotonic, in the sense that the set of states for which the action b should finally be chosen constitute an initial interval (i.e., an interval containing 0). If the bias is large, this imposes a significant cost on the informed party when the state is close enough to the upper end of this interval, the informed party may be better off pretending that the state of the world is larger. This intuition suggests that the result should extend to the case of communication of unbounded length, although we do not prove this here.

6. Discussion and Conclusion

We have compared the performance of three common dispute resolution processes – arbitration, mediation and negotiation – in the framework of Crawford and Sobel (1982). Under arbitration, the two parties commit to conform to the decision of a neutral third party. Under mediation instead, compliance with the third party's suggested settlement is voluntary. Finally, under unfacilitated negotiation, the two parties engage in (possibly arbitrarily long) face-to-face cheap talk. We have characterized and compared the optimal arbitration and mediation schemes, and identified necessary and sufficient conditions for negotiation to perform as well as mediation. The optimal mediation scheme corresponds to the communication protocol developed by Blume and Board (2006). Thus, we find that mediators may act optimally in dispute resolution by filtering the unmediated communication and introducing noise to it. We have found that mediation performs better than negotiation when the conflict of interest is intermediate, whereas a mediator is unnecessary and two rounds of communication suffice when the conflict of interest is low.

While our most direct applications are in the contexts of international crises, legal confrontations, and business transactions, as mentioned earlier, the mediation rules we characterize do not only refer to the practice of professional mediators. Other economic institutions can play the role of mediators. For instance, it has been argued that the use of consultants during firm restructuring (Mitusch and Strausz, 1999), or of legislative committees (Li, 2007) fulfill this function. In other settings, however, it is hard to see what institutions would accomplish this purpose. For instance, the communication between investors and financial analysts appears to be unmediated. This raises the question of why this is the case. Our paper offers one possible explanation. Namely, it may be the case that the "Chinese walls" that separate by law the brokerage and investment divisions of banks are sufficiently effective in mitigating the analysts' biases to render mediation useless (as is the case, formally, when $b < 1/8$). Another important difference is that monetary transfers (i.e. payments) are standard practice in this setting, which we do not allow here.

Alternatively, this may suggest that other models of cheap talk, such, as for instance, Dewatripont and Tirole's (2005), might be better suited than Crawford and Sobel's, depending on the particular application. As a first step towards characterizing the optimal properties of mediation, this paper illustrates the potential and the limitations of mediation and negotiation in the set-up by Crawford and Sobel. Understanding when and how mediation or negotiation are likely to be effective beyond the classical framework by Crawford and Sobel is a further, formidable and fundamental challenge.

References

[1] Alonso, R., and N. Matouschek (2004): "Relational Delegation," manuscript, Northwestern University.

[2] Alonso, R., and N. Matouschek (2006): "Optimal Delegation," manuscript, Northwestern University.

[3] Aumann, R., and S. Hart (2003): "Long Cheap Talk," *Econometrica* 71(6), 1619-1660.

[4] Ayres, I. and B. J. Nalebuff (1997): "Common Knowledge As A Barrier To Negotiation", Yale ICF Working Paper No. 97-01.

[5] Bercovich, J. and R. Jackson (2001): "Negotiation or Mediation? An Exploration of Factors Affecting the Choice of Conflict Management in International Conflict", *Negotiation Journal*, 17(1), 59-77.

[6] Blume, A., and O. Board (2006): "Noisy Talk," manuscript, University of Pittsburgh.

[7] Brown, J. G. and I. Ayres (1994): "Economic Rationales for Mediation", *Virginia Law Review*, 80, 323-402.

[8] Carnevale, P.J. and D.G. Pruitt (1992): "Negotiation and Mediation," *Annual Review of Psychology* 43, 531-82.

[9] Crawford, V., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica* 50(6), 1431-1451.

[10] Dewatripont, M. and J. Tirole (2005): "Modes of Communication," *Journal of Political Economy*, 113(6), 1217-1238.

[11] Forges, F. (1990): "Equilibria with communication in a job market example," *Quarterly Journal of Economics,* 105, 375-398.

[12] Ganguly, C. and I. Ray (2005): "Can Mediation Improve Upon Cheap Talk? A Note," manuscript, University of Birmingham.

[13] Gerardi, D. (2004): "Unmediated Communication in Games with Complete and Incomplete Information," *Journal of Economic Theory*, 114, 104-131.

[14] Grossman, G. and E. Helpman (2001): *Special Interest Politics*, Cambridge: MIT Press.

[15] Holmström, B. (1977): "On Incentives and Control in Organizations," PhD Dissertation, Stanford University.

[16] Kawamura, K. (2006): "Anonymity, Equal Treatment, and Overconfidence: Constraints on Communication May Enhance Information Transmission," manuscript, Nuffield College, Oxford University.

[17] Kováč, E. and T. Mylovanov (2007): "Stochastic Mechanisms in Settings without Monetary Transfers: Regular Case," working paper, University of Bonn.

[18] Krishna, V. and J. Morgan (2004): "The Art of Conversation: Eliciting Information from informed partys through Multi-stage Communication," *Journal of Economic Theory* 117, 147-179.

[19] Kydd, A. (2006): "When Can Mediators Build Trust", *American Political Science Review*, 100, 449-462.

[20] Li, T. (2007): "The Messenger Game: Strategic Information Transmission through Legislative Committees", working paper, Harvard University.

[21] Melumad, N. and T. Shibano (1991): "Communication in Settings with No Transfers," *The RAND Journal of Economics* 22(2), 173-198.

[22] Mitusch, K. and R. Strausz (1999): "Mediators and Mechanism Design: Why Firms Hire Consultants," discussion paper 1999/15, Free University of Berlin.

[23] Mitusch, K. and R. Strausz (2005): "Mediation in Situations of Conflict and Limited Commitment," *Journal of Law, Economics, and Organization*, 21(2), 467-500.

[24] Morgan, J. and P. Stocken (2003): "An Analysis of Stock Recommendations," *The RAND Journal of Economics* 34(1), 183-203.

[25] Myerson, R. (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics* 10(1), 67-81.

[26] Myerson, R. (1981): "Optimal Auction Design, " *Mathematics of Operations Research* 6, 58-73.

[27] Myerson, R. (2007): "crawfsob.xls", http://home.uchicago.edu/~rmyerson/research/index.html.

[28] Raiffa, H. (1982): *The Art and Science of Negotiation*, Harvard University Press.

[29] Singer, L. R. (1990): *Settling Disputes: Conflict Resolution in Business, Families, and the Legal System*, Westview Press.

[30] Touval, S. (1996): "Coercive Mediation on the road to Dayton," *International Mediation,* 1, 547-70.

[31] Umbreit M.S. (1993): "Crime Victims and Offenders in Mediation: An Emerging Area of Social Work Practice," *Social Work* 38, 69-73.

[32] Wall, J.A. and A. Lynn (1993): "Mediation: A Current Review," *Journal of Conflict Resolution* 36, 160-94.

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

7. Appendix A: Arbitration

7.1. **A1. Lemma 2.**

Lemma 2. $\{y(\theta), \sigma^2(\theta)\}_{\theta \in \Theta}$ *satisfy* $(IC - IP)$ *if and only if*

(i) $y(\theta)$ *is non-decreasing;*

(ii) $-\sigma^2(\theta) = U(\theta) + (y(\theta) - (\theta + b))^2$, *and* $U(\theta) = U(0) + \int_0^\theta 2\left(y\left(\tilde{\theta}\right) - \left(\tilde{\theta} + b\right)\right) d\tilde{\theta}.$

Proof. <**Only If**>

(i) From incentive compatibility for every $\theta, \theta' \in \Theta$ we have

$$-\sigma^2(\theta) - (y(\theta) - (\theta + b))^2 \geq -\sigma^2(\theta') - (y(\theta') - (\theta + b))^2;$$

$$-\sigma^2(\theta') - (y(\theta') - (\theta' + b))^2 \geq -\sigma^2(\theta) - (y(\theta) - (\theta' + b))^2.$$

Adding up and rearranging we get

$$(\theta - \theta')(y(\theta) - y(\theta')) \geq 0.$$

(ii) By the Envelope Theorem we have

$$U(\theta) = U(0) + \int_0^\theta 2\left(y\left(\tilde{\theta}\right) - \left(\tilde{\theta} + b\right)\right) d\tilde{\theta}.$$

Hence

$$-\sigma^2(\theta) = U(\theta) + (y(\theta) - (\theta + b))^2.$$

<**If**>

We need to show that for every $\theta, \theta' \in \Theta$,

$$\left(-\sigma^2(\theta) - (y(\theta) - (\theta + b))^2\right) - \left(-\sigma^2(\theta') - (y(\theta') - (\theta + b))^2\right) \geq 0.$$

Notice that

$$-\sigma^2\left(\theta'\right) - \left(y\left(\theta'\right) - \left(\theta + b\right)\right)^2 = -\sigma^2\left(\theta'\right) - \left(y\left(\theta'\right) - \left(\theta' + b\right)\right)^2 - 2y\left(\theta'\right)\left(\theta' + b\right)$$

$$+\left(\theta' + b\right)^2 + 2y\left(\theta'\right)\left(\theta + b\right) - \left(\theta + b\right)^2 = U\left(\theta'\right) - \int_\theta^{\theta'} 2\left(y\left(\theta'\right) - \left(\widetilde{\theta} + b\right)\right)d\widetilde{\theta}.$$

So

$$U\left(\theta\right) - U\left(\theta'\right) + \int_\theta^{\theta'} 2\left(y\left(\theta'\right) - \left(\widetilde{\theta} + b\right)\right)d\widetilde{\theta} = \int_\theta^{\theta'} 2\left(y\left(\theta'\right) - y\left(\widetilde{\theta}\right)\right)d\widetilde{\theta} \geq 0.$$

∎

7.2. **A2.   Proof of Theorem 1.** By Lemma 2 the optimal arbitration rule has to solve the following simplified problem:

$$\max_{y(\cdot),\sigma^2(\cdot),U(0)} V = \int_0^1 \left(-\sigma^2\left(\theta\right) - \left(y\left(\theta\right) - \theta\right)^2\right)d\theta.$$

subject to

$$y\left(\theta\right) \text{ is non-decreasing;} \qquad\qquad\qquad (MON)$$

$$\sigma^2\left(\theta\right) = -U\left(0\right) - \int_0^\theta 2\left(y\left(\widetilde{\theta}\right) - \left(\widetilde{\theta} + b\right)\right)d\widetilde{\theta} - \left(y\left(\theta\right) - \left(\theta + b\right)\right)^2; \quad (ENV)$$

$$\sigma^2\left(\theta\right) \geq 0, \ U\left(0\right) \leq 0 \qquad\qquad\qquad\qquad (NONNEG)$$

The proof of Theorem 1 proceeds through a series of lemmas.

Lemma 3. *If $\left(y\left(\theta\right), \sigma^2\left(\theta\right), U\left(0\right)\right)$ are feasible, then*

$$V\left(y\left(\theta\right), \sigma^2\left(\theta\right), U\left(0\right)\right) = U\left(0\right) + 2\int_0^1 y\left(\theta\right)\left(1 - \theta - b\right)d\theta + b^2 - \frac{1}{3}.$$

Proof. Substitute constraint (ENV) into the objective function and change the order of integration in the double integral. ∎

Lemma 4.   *Let $b \in \left[0, \frac{1}{2}\right]$. Mechanism $\left(y\left(\theta\right), \sigma^2\left(\theta\right), U\left(0\right)\right)$ is optimal.*

Proof. Assume there exists a mechanism $\left(\widehat{y}(\theta), \widehat{\sigma}^2(\theta), \widehat{U}(0)\right)$ which achieves a strictly higher welfare than the mechanism $(y(\theta), \sigma^2(\theta), U(0))$.

By Lemma 3 we have

$$0 < V\left(\widehat{y}(\theta), \widehat{\sigma}^2(\theta), \widehat{U}(0)\right) - V(y(\theta), \sigma^2(\theta), U(0)) =$$

$$= \widehat{U}(0) - U(0) + 2 \int_0^1 (\widehat{y}(\theta) - y(\theta))(1 - \theta - b)\, d\theta.$$

Also

$$0 < V\left(\widehat{y}(\theta), \widehat{\sigma}^2(\theta), \widehat{U}(0)\right) - V(y(\theta), \sigma^2(\theta), U(0)) =$$

$$= -\int_0^1 \left((\widehat{y}(\theta) - \theta)^2 + \widehat{\sigma}^2(\theta)\right) d\theta + \int_0^1 \left((y(\theta) - \theta)^2 + \sigma^2(\theta)\right) d\theta \leq$$

$$\leq -\int_0^1 (\widehat{y}(\theta) - \theta)^2\, d\theta + \int_0^1 (y(\theta) - \theta)^2\, d\theta =$$

$$= 2 \int_0^1 (\widehat{y}(\theta) - y(\theta))(\theta - y(\theta))\, d\theta - \int_0^1 (\widehat{y}(\theta) - y(\theta))^2\, d\theta <$$

$$< 2 \int_0^1 (\widehat{y}(\theta) - y(\theta))(\theta - y(\theta))\, d\theta.$$

Adding up two inequalities,

$$0 < \widehat{U}(0) - U(0) + 2 \int_0^1 (\widehat{y}(\theta) - y(\theta))(1 - b - y(\theta))\, d\theta.$$

Substituting $y(\theta)$ and using (ENV) we get

$$0 < \widehat{U}(0) - U(0) + 2 \int_0^{1-2b} (\widehat{y}(\theta) - y(\theta))(1 - 2b - \theta)\, d\theta =$$

$$= \widehat{U}(0) + \int_0^{1-2b} \frac{d\widehat{U}(\theta)}{d\theta}(1 - 2b - \theta)\, d\theta - U(0) - \int_0^{1-2b} \frac{dU(\theta)}{d\theta}(1 - 2b - \theta)\, d\theta =$$

$$= \widehat{U}(0) - \widehat{U}(0)(1 - 2b) + \int_0^{1-2b} \widehat{U}(\theta)\, d\theta - U(0) + U(0)(1 - 2b) - \int_0^{1-2b} U(\theta)\, d\theta =$$

$$= 2b\left(\widehat{U}(0) - U(0)\right) + 2 \left(\int_0^{1-2b} \left(\widehat{U}(\theta) - U(\theta)\right) d\theta\right).$$

However, this is not possible since $\widehat{U}(\theta) \leq U(\theta) = 0$ for every $\theta \in [0, 1 - 2b]$. ∎

Lemma 5. *Let $b > \frac{1}{2}$. If $\left(\widehat{y}\left(\theta\right), \widehat{\sigma}^2\left(\theta\right), \widehat{U}\left(0\right)\right)$ are optimal, then $\widehat{y}\left(\theta\right)$ is constant on $(0,1)$.*

Proof. Suppose that $\widehat{y}\left(\theta\right)$ is not constant on $\left(0,1\right)$, i.e. $\exists \theta, \theta' \in (0,1)$ such that $\theta' > \theta$, $\widehat{y}\left(\theta'\right) > \widehat{y}\left(\theta\right)$.

Consider the following policy:

$$
\begin{aligned}
y_1\left(\theta\right) &= \widehat{y}\left(0\right) \text{ for every } \theta \\
\sigma_1^2\left(\theta\right) &= \widehat{\sigma}^2\left(0\right) \\
U_1\left(0\right) &= \widehat{U}\left(0\right).
\end{aligned}
$$

Obviously $(y_1\left(\theta\right), \sigma_1^2\left(\theta\right), U_1\left(0\right))$ satisfy constraint (MON) and $(NONNEG)$. Moreover, this policy achieves a strictly higher value of the objective function than the original policy, since, by Lemma 3,

$$
\begin{aligned}
V\left(\widehat{y}, \widehat{\sigma}^2\right) &= \widehat{U}\left(0\right) + 2\int_0^1 \widehat{y}\left(\theta\right)\left(1 - \theta - b\right) d\theta + b^2 - \frac{1}{3} \\
&< \widehat{U}\left(0\right) + 2\int_0^1 \widehat{y}\left(\theta\right) d\theta \int_0^1 \left(1 - \theta - b\right) d\theta + b^2 - \frac{1}{3} \\
&< \widehat{U}\left(0\right) + 2\widehat{y}\left(0\right) \int_0^1 \left(1 - \theta - b\right) d\theta + b^2 - \frac{1}{3} \\
&= V\left(y_1, \sigma_1^2\right).
\end{aligned}
$$

The first inequality is due to (MON) and the fact that $\widehat{y}\left(\theta\right)$ is not constant on $(0,1)$; the last inequality is due to (MON) and the fact that $\int_0^1 \left(1 - \theta - b\right) d\theta = \frac{1}{2} - b < 0$. So the original policy is suboptimal. ∎

Lemma 6. *Let $b > \frac{1}{2}$. Mechanism $(y\left(\theta\right), \sigma^2\left(\theta\right), U\left(0\right))$ is optimal.*

Proof. By Lemma 5, if $\widehat{y}\left(\theta\right)$ is a part of an optimal policy, then it is constant on $(0,1)$. Without loss of generality, we can restrict attention to policies such that $\widehat{y}\left(\theta\right)$

is constant on $[0, 1]$. Take any such policy $\left(\widehat{y} \left(\theta \right), \widehat{\sigma}^2 \left(\theta \right), \widehat{U} \left(0 \right) \right)$. Then

$$V \left(y \left(\theta \right), \sigma^2 \left(\theta \right), U \left(0 \right) \right) - V \left(\widehat{y} \left(\theta \right), \widehat{\sigma}^2 \left(\theta \right), \widehat{U} \left(0 \right) \right)$$

$$= \int_0^1 \left(-\sigma^2 \left(\theta \right) - \left(y \left(\theta \right) - \theta \right)^2 \right) d\theta - \int_0^1 \left(-\widehat{\sigma}^2 \left(\theta \right) - \left(\widehat{y} \left(\theta \right) - \theta \right)^2 \right) d\theta$$

$$\geq - \int_0^1 \left(y \left(\theta \right) - \theta \right)^2 d\theta + \int_0^1 \left(\widehat{y} \left(\theta \right) - \theta \right)^2 d\theta = \widehat{y}^2 \left(0 \right) - \widehat{y} \left(0 \right) + \frac{1}{4} \geq 0,$$

where the first inequality follows from the fact that $\widehat{\sigma}^2 \left(\theta \right) \geq \sigma^2 \left(\theta \right) = 0$. ∎

The proof of Theorem 1 follows from Lemmas 4 and 6. ∎

7.3. **A3. Intuition for Theorem 1.** To gain the intuition for why the optimal arbitration rule is deterministic, consider, for example, reducing $y \left(\theta \right)$ by a small $\delta > 0$ on an interval $\left[0, \overline{\theta} \right] \subseteq \left[0, 1 - 2b \right]$. The distance between the expected action and the decision-maker's most preferred action is thus reduced from b to $b - \delta$ on this interval. The expected gain for the decision-maker is

$$\int_0^{\overline{\theta}} \left(- \left(b - \delta \right)^2 \right) d\theta - \int_0^{\overline{\theta}} \left(-b^2 \right) d\theta = \left(2b - \delta \right) \delta \overline{\theta}.$$

In order to implement this new expected action function in an incentive compatible way we need to adjust a conditional variance function. First, we need to adjust $\sigma^2 \left(\cdot \right)$ on the interval $\left[0, \overline{\theta} \right]$. Differentiation of (ENV) yields

$$\tfrac{d}{d\theta} \sigma^2 \left(\theta \right) = -2 \left(y \left(\theta \right) - \left(\theta + b \right) \right) \tfrac{d}{d\theta} y \left(\theta \right) \big|_{y \left(\theta \right) = \theta + b - \delta} = 2\delta.$$

Intuitively, since we are implementing expected actions which are below the most preferred actions of the informed party, we need to discourage him from choosing higher actions by increasing their variance. The variance for the types on the interval $\left[0, \overline{\theta} \right]$ is thus $\sigma^2 \left(\theta \right) = 2\delta\theta$.

Second, the variance for the types immediately above $\overline{\theta}$ cannot be lower than the variance $\sigma^2 \left(\overline{\theta} \right) = 2\delta\overline{\theta}$, since otherwise the type $\overline{\theta}$ could slightly overstate his type and receive his most preferred expected action, $\overline{\theta} + b$, at a smaller variance. The incentive

compatibility of the original mechanism implies that the variance for all the types above $\bar{\theta}$ is at least as large as the variance of the type $\bar{\theta}$.

The expected loss for the decision-maker from the variance is thus at least

$$\int_0^{\bar{\theta}} 2\delta\theta d\theta + \int_{\bar{\theta}}^1 2\delta\bar{\theta}d\theta = \left(2 - \bar{\theta}\right)\delta\bar{\theta}.$$

Hence the net benefit for the decision-maker is at most

$$\left((2b - \delta) - \left(2 - \bar{\theta}\right)\right)\delta\bar{\theta} < 0$$

where the inequality follows from $\delta > 0$, $\bar{\theta} < 1$ and $b \leq \frac{1}{2}$.

8. Appendix B: Mediation

8.1. **B1. Proof of Lemma 1.**

Proof. By $(IC - DM)$ and $(PROB)$,

$$(1) \qquad \int_\Theta y\left(\theta\right)d\theta = \int_{Y\times\Theta} yp\left(dy, d\theta\right) = \int_{Y\times\Theta} \theta p\left(dy, d\theta\right) = \frac{1}{2}.$$

By $(IC - DM)$,

$$(2) \qquad cov\left(\theta, y\left(\theta\right)\right) = cov\left(\theta, y\right) = cov\left(E_\theta\left[\theta \mid y\right], y\right) = cov\left(y, y\right) = var\left(y\right),$$

By Lemma 3 (see Appendix A.2) and equations (1) and (2),

$$\begin{aligned} V &= U\left(0\right) + 2\int_0^1 y\left(\theta\right)\left(1 - \theta - b\right)d\theta + b^2 - \frac{1}{3} \\ &= U\left(0\right) - 2\int_0^1 y\left(\theta\right)\theta d\theta + 1 - b + b^2 - \frac{1}{3} \\ (3) \qquad &= U\left(0\right) - 2var\left(y\right) + \frac{1}{6} - b + b^2. \end{aligned}$$

On the other hand,

$$
\begin{aligned}
V &= -E\left(y - \theta\right)^2 \\
&= -E\left[\left(\left(\theta - \frac{1}{2}\right) - \left(y - \frac{1}{2}\right)\right)^2\right] \\
&= -var\left(\theta\right) + 2cov\left(y, \theta\right) - var\left(y\right) \\
&= var\left(y\right) - var\left(\theta\right) = var\left(y\right) - \frac{1}{12},
\end{aligned}
$$

(4)

where the second equality follows from (1), the third equality follows from (2), and the last equality holds because θ is uniformly distributed.

Combining (3) and (4), we get

$$
U\left(0\right) = 3var\left(y\right) - \frac{1}{4} + b - b^2.
$$

Since $U\left(0\right) \leq 0$, we have

(5)
$$
var\left(y\right) \leq \frac{1}{12} - \frac{1}{3}b + \frac{1}{3}b^2.
$$

Substituting (5) into (4), we get

$$
V \leq -\frac{1}{3}\left(b - b^2\right).
$$

This holds with equality if and only if $U(0) = 0$. ∎

9. Appendix C: Negotiation

9.1. **C1. Proof of Theorem 3.** First, let us prove an auxiliary result about a helpful property of quadratic preferences.

Lemma 7. *Let $\theta_1, \theta_2 \in [0, 1]$, $\theta_1 < \theta_2$. Let l be a lottery on Y such that l does not put probability one on action b, and θ_1 weakly prefers l to action b. Then θ_2 strictly prefers l to action b.*

Proof. Recall that the utility of a lottery l with mean y and variance σ^2 for the informed party of type θ equals $U\left(\theta\right) = -\sigma^2 - (y - (\theta + b))^2$. Consequently, type θ weakly prefers

l to action b if and only if

$$\sigma^2 + (y - b)^2 \leq 2\theta(y - b),$$

which implies that $y \geq b$, no matter what θ is. So if $\theta_2 > \theta_1$ and the inequality above holds weakly for θ_1, then it has to hold strictly for θ_2. ∎

We restrict attention to canonical equilibria in the sense of Aumann and Hart (2003): that is equilibria in which revelations by the informed party alternate with jointly controlled lotteries. For expositional simplicity, let us suppose that the players, instead of conducting jointly controlled lotteries, have access to a randomization device that sends messages at the jointly controlled lottery stages, so that at each stage, either the informed party or the device sends one public message.

First, let us introduce some notation. Let p be an optimal mediation rule, and suppose that p is implementable with finite cheap talk. Let $\Theta_1 := \{\theta \in \Theta : p(b|\theta) = 1\}$. We know that $\Theta_1 \neq \emptyset$, because $0 \in \Theta_1$. Let N be the set of all possible sequences of messages that can be observed in the equilibrium that implements p, and let $\mu(\cdot|\theta)$ be the probability distribution over N conditional on the state being θ. Let $P(\cdot|n)$ be the decision-maker's posterior upon observing $n \in N$, that is, for $\Theta \subseteq [0,1]$, $n \in N$,

$$P(\Theta|n) = \frac{\int_{\Theta} \mu(n|\theta) \, dF(\theta)}{\int_{[0,1]} \mu(n|\theta) \, dF(\theta)}, \text{ if } \int_{[0,1]} \mu(n|\theta) \, dF(\theta) > 0$$

Let us also assume that

$$P(\Theta|n) = 1 \text{ if } \int_{[0,1]} \mu(n|\theta) \, dF(\theta) = 0 \text{ and } [\mu(n|\theta) > 0 \Rightarrow \theta \in \Theta]$$

The last assumption implies that if a particular path of play can only appear in one state θ^*, then upon observing this path of play, the DM concludes that the state is θ^* with probability one (this restriction on conditional probabilities seems somewhat arbitrary, but it is commonly made in signaling models with a continuum of types when talking about separating equilibria).

Finally, let $n(t)$ be the restriction of sequence $n \in N$ to the first t stages (including stage t), and let n_t be the message sent at stage t according to sequence n. We can

also define $\mu\left(n\left(t\right)|\theta\right) := \int_{n' \in N : n'(t) = n(t)} d\mu\left(n|\theta\right)$, the probability that $n\left(t\right)$ realizes in equilibrium given θ.

Lemma 8. $\Theta_1 = [0, 2b]$.

Proof. For any period $t = 0, ..., T$ and partial history $n(t)$, let

$$A(n(t)) = \{\theta \in [0, 1] : \exists n' \in \text{ support } \mu(\cdot|\theta), n'(t) = n(t)\}$$

be the set of types whose equilibrium behavior is consistent with partial history $n(t)$. Let $\Theta_1(n(t)) = \{\theta \in A(n(t)) :$ for a.e. (with respect to $\mu(\cdot|\theta)$) $n' \in N$ s.t. $n'(t) = n(t), E(\theta|n') = b\}$ be the set of types that, following the history $n(t)$, get action b with probability one. Let us prove that for every $t = 0, ..., T$ and $n(t)$ such that $\Theta_1(n(t)) \neq \emptyset$,

    (a) $\Theta_1(n(t)) = [0, \theta(n(t))] \cap A(n(t))$, for some $\theta(n(t)) \geq b$;

    (b) $E(\theta|n(t), \Theta_1(n(t))) = b$.

The proof will be by induction, starting from $t = T$. Take any partial history $n(T - 1)$ such that $\Theta_1(n(T - 1)) \neq \emptyset$. Suppose, without loss of generality, that T is a revelation stage. By Lemma 1 of Crawford and Sobel (1982), the equilibrium of the subgame following the history $n(T - 1)$ is partitional. In particular, since $\Theta_1(n(T - 1)) \neq \emptyset$, there exists an interval $[a(n), \theta(n)\rangle$ (closed or open on the right) such that, after the history $n(T - 1)$, all types in this interval, and only them, choose messages that lead to action b; that is, $\Theta_1(n) = [a(n), \theta(n)\rangle \cap A(n(t))$. Moreover, $a(n)$ can be taken to be 0. Suppose not, that is, $a(n) > \theta$, for some $\theta \in A(n(t))$. Then in the partitional equilibrium of the subgame that we consider, type θ achieves an action lower than b, which is strictly worse for it than action b. But it could have achieved action b if it played like type $a(n)$ – a contradiction. It also has to be the case that $E(\theta|n, \Theta_1(n)) = E(\theta|n) = b$, and, consequently, that $\theta(n) \geq b$.

Now suppose that the statement is true for all partial histories of length $t + 1, ..., T$, and let us prove it for partial histories of length t. Consider any $n(t)$ such that $\Theta_1(n(t)) \neq \emptyset$. By definition, $\Theta_1(n(t)) \subseteq A(n(t))$. There are two cases to consider:

(a) t is a revelation stage. We have to prove that

$$\theta', \theta'' \in \Theta_1(n(t)), \ \theta \in (\theta', \theta'') \cap A(n(t)) \Rightarrow \theta \in \Theta_1(n(t))$$

and that

$$\theta' \in \Theta_1(n(t)), \ \theta \in \cap A(n(t)), \ \theta < \theta' \Rightarrow \theta \in \Theta_1(n(t))$$

Suppose $\theta', \theta'' \in \Theta_1(n(t))$ and $\theta \in A(n(t))$. Then both θ' and θ'' choose contin-uation strategies at stage t that guarantee action b with probability one. Incentive compatibility implies that θ also has to choose a continuation strategy that guarantees b with probability one – otherwise either θ' or θ'' has an incentive to imitate θ. This means that $\theta \in \Theta_1(n(t))$.

Now, suppose that $\theta' \in \Theta_1(n(t))$, $\theta \in \cap A(n(t))$ and $\theta < \theta'$. This means that θ' chooses a continuation strategy at stage t that guarantees action b with probability one. If θ chooses a strategy that results in a different lottery over actions, then, by Lemma 7, θ' should strictly prefer to imitate θ – a contradiction. This means that θ also chooses a continuation strategy at stage t that guarantees action b with probability one, so $\theta \in \Theta_1(n(t))$.

This proves that $\Theta_1(n(t)) = [0, \theta(n(t))] \cap A(n(t))$.

Since t is a revelation stage,

$$\Theta_1(n(t)) = \left(\bigcup_{n' \in N: n'(t) = n(t)} \Theta_1(n'(t+1)) \right) \setminus B,$$

where $B \subseteq [0,1]$ includes at most one type. To see this, note that it follows from the definition that $\Theta_1(n(t)) \subseteq \bigcup_{n' \in N: n'(t)=n(t)} \Theta_1(n'(t+1))$. Now suppose that $\Theta_1(n(t)) \subset \bigcup_{n' \in N: n'(t)=n(t)} \Theta_1(n'(t+1))$, and take any $\theta \in \bigcup_{n' \in N: n'(t)=n(t)} \Theta_1(n'(t+1)) \setminus \Theta_1(n(t))$. By the definition of $\Theta(n(t))$, it must be the case that type θ is randomizing at stage t between messages that will result in action b with probability one, and messages that results in some other lottery. But with quadratic preferences, there can be at most one such type. To see this, suppose, by way of contradiction, that there are two types, θ_1 and θ_2, both in $\bigcup_{n' \in N: n'(t)=n(t)} \Theta_1(n'(t+1)) \setminus \Theta_1(n(t))$, such that after history $n(t)$, type

θ_i is randomizing between messages that will result in action b with probability one, and messages that results in some other lottery (call it l_i). Without loss of generality, suppose that $\theta_1 < \theta_2$. Then it must be the case that θ_1 is indifferent between action b and lottery l_1, so, by Lemma 7, θ_2 strictly prefers l_1 to b and, consequently, to l_2. This means that imitating θ_1 is a profitable deviation for θ_2 – a contradiction. This proves that B contains at most one type.

So

$E(\theta|n(t), \Theta_1(n(t))) = \int_{n' \in N: n'(t) = n(t)} E[\theta|n'(t+1), \Theta_1(n(t))] d\mu(n'(t+1)|n(t), \Theta_1(n(t)))$

$= \int_{n' \in N: n'(t) = n(t)} \{ E[\theta|n'(t+1), \Theta_1(n'(t+1))] P[\Theta_1(n'(t+1))|n'(t+1), \Theta_1(n(t))] + E[\theta|n'(t+1),$

$\Theta_1(n(t)) \setminus \Theta_1(n'(t+1))] P[\Theta_1(n(t)) \setminus \Theta_1(n'(t+1))|n'(t+1), \Theta_1(n(t))] d\mu(n'(t+1)|n(t), \Theta_1(n(t)))$

$= b,$

where the last equality follows from the fact that

$$P(\Theta_1(n'(t+1))|n'(t+1), \Theta_1(n(t))) = \begin{cases} 1, \; if \; n'(t+1) \in support \; \mu(\cdot|\Theta_1(n(t))); \\ 0, \; otherwise \end{cases}$$

and from the induction hypothesis. It follows immediately that $\theta(n(t)) \geq b$.

(b) t is a jointly controlled lottery stage. Then it follows from the definition of $\Theta_1(n(t))$ that

$$\Theta_1(n(t)) = \bigcap_{n' \in N: n'(t) = n(t)} \Theta_1(n'(t+1)) = \bigcap_{n' \in N: n'(t) = n(t)} [0, \theta(n'(t+1))] \cap A(n'(t+1))$$

If t is a jointly controlled lottery stage, then for any $n', n'' \in N$ such that $n'(t) = n''(t) = n(t)$, $A(n'(t+1)) = A(n''(t+1)) = A(n(t))$. So

$$\Theta_1(n(t)) = A(n(t)) \cap \bigcap_{n' \in N: n'(t) = n(t)} [0, \theta(n'(t+1))] = A(n(t)) \cap [0, \theta(n(t))]$$

where $\theta(n(t)) = inf_{n' \in N: n'(t) = n(t)} \theta(n'(t+1))$. Furthermore, the first equality above, together with the fact that $\forall n' \in N : n'(t) = n(t), E(\theta|A(n(t)) \cap [0, \theta(n'(t+1))]) = b$ implies that $E(\theta|A(n(t)) \cap [0, \theta(n(t))]) = b$.

So we have proved that for every $t = 0, ..., T$ and $n(t)$ such that $\Theta_1(n(t)) \neq \emptyset$,

(a) $\Theta_1(n(t)) = [0, \theta(n(t))] \cap A(n(t))$, for some $\theta(n(t)) \geq b$;

(b) $E(\theta|n(t), \Theta_1(n(t))) = b$.

In particular, if $t = 0$, then $n(t)$ is an empty history, $\Theta_1(t) = \Theta_1$ by definition, and $A(n(t)) = [0,1]$. Consequently, $\Theta_1 = [0, \theta_0]$ for some $\theta_0 \geq b$, and $E(\theta|\Theta_1) = b$. It follows immediately that $\theta_0 = 2b$. ∎

Proof of Theorem 3 (Only If). Because any type θ smaller than and sufficiently close to $2b$ strictly prefers $(2b + 1)/2$ over b, it follows that any such type θ also strictly prefers to the outcome b any non-degenerate distribution q over actions with support contained in $[b, (2b + 1)/2]$.

Consider any strategy μ such that, after any history $n(t - 1)$ such that t is a revelation stage and $supportP(\theta|n(t-1)) \cap [2b, 1] \neq \emptyset$, the sender chooses a message m that minimizes $E[\theta|n(t-1), m']$ among the messages m' such that $supportP(\theta|n(t-1), m') \cap [2b, 1] \neq \emptyset$. Then at the first stage (without loss of generality, suppose that this is a revelation stage), this strategy calls for sending a message m such that $E[\theta|m] \leq [2b + 1]/2 = E(\theta|[2b, 1])$; and by the law of iterated expectations, for every stage t, $E[\theta|n(t-1), m] \leq E[\theta|n(t-1)]$. It is clear that since $\Theta_1 = [0, 2b]$, this strategy cannot lead to any action that is lower than $2b$. On the other hand, at the terminal stage T, for any history n that can realize if strategy μ is followed, $E(\theta|n) \leq E(\theta|n(t-1)) \leq ... \leq E(\theta|n(1)) \leq (2b + 1)/2$, so the action that will be executed cannot exceed $(2b + 1)/2$. It follows that the strategy μ induces a lottery over actions whose support is contained in $[2b, (2b + 1)/2]$, and a type $\theta = 2b - \varepsilon$ for $\varepsilon > 0$ small enough will prefer following this strategy to the strategy that induces action b with certainty.

It follows that p is not incentive compatible. ∎