Discussion Paper No. 24

Mediation in Situations of Conflict and Limited Commitment

Kay Mitusch*
Roland Strausz**

June 2004

*Kay Mitusch, Free University Berlin, Department of Economics, Boltzmannstr. 20, D-14195 Berlin, Germany, mitusch@wz-berlin.de
**Roland Strausz, Free University Berlin, Department of Economics, Boltzmannstr. 20, D-14195 Berlin, strausz@fu-berlin.de

Sonderforschungsbereich/Transregio 15 · www.gesy.uni-mannheim.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Konrad Stahl, Ph.D. · Department of Economics · University of Mannheim · D-68131 Mannheim,
Phone: +49(0621)1812786 · Fax: +49(0621)1812785

Mediation in Situations of Conflict and Limited Commitment

Kay Mitusch*and Roland Strausz*

June 10, 2004

Abstract

We study the reasons and conditions under which mediation is beneficial when
a principal needs information from an agent to implement an action. Assuming
a strong form of limited commitment, the principal may employ a mediator who
gathers information and makes non–binding proposals. We show that a partial rev-
elation of information is more effective through a mediator than through the agent
himself. This implies that mediation is strictly helpful if and only if the likelihood
of a conflict of interest is positive but not too high. The value of mediation depends
non–monotonically on the degree of conflict. Our insights extend to general models
of contracting with imperfect commitment.

Keywords: Contracting, Non-Commitment, Revelation Principle
JEL Classification No.: D82

1   Introduction

On January the 17th, 1998 Saddam Hussein announces a halt to all UN weapons inspections in Iraq. During the following month all direct negotiations between the US and Iraq end in vain and military action seems unavoidable. Yet, on February 20th the US authorizes the UN secretary general Kofi Annan to fly to Iraq on a last–ditch mediation effort, where he announces a deal with Bagdad. After a two–day review of the deal the US accepts and strikes are averted. How come that the secretary general succeeded in securing a deal that was not attainable in direct negotiations? How come that about 8 months later air strikes were launched after a new stand-off and mediation did not take place? These are the questions that this paper addresses.

More generally, we are interested in the role of mediation in situations of conflict. Casual observation suggests that mediators are often indispensable for settling conflicts between sovereign nations and resolving labor disputes between employers and employees. Also in every day life mediators play an important, albeit less formal role in defusing many quarrels between family members, friends, and colleagues. Overall the popularity of mediation is increasing. Smith (1995), for example, reports that in 1990 American businesses started using mediation to resolve legal disputes. The increased interest in mediation is also reflected in modern education. Nowadays mediation skills are taught to sixth graders in elementary school and students at Harvard Law School.

We want to study the rationale behind mediation and are especially interested in the circumstances under which institutions may arise that perform the role of mediators. This paper focuses on a specific explanation for mediation: limited commitment in contracting. Indeed, although the secretary general was sent with a highly restricted offer to Bagdad, he still had some discretion over the final deal and the US needed two additional days to review the final offer. In short, the US was not committed to the restricted offer of Kofi Annan. This paper argues that it is this limited commitment power that renders mediation profitable. Indeed, if full commitment had been available, the US could have committed itself to behave just as Kofi Annan and therefore arrive at an identical deal. The theoretical counterpart of this simple idea is the revelation principle in mechanism design: With full commitment the party that offers the contract can imitate the behavior of any other player and, hence, such players are not useful. Without such commitment the revelation principle fails and additional players may alleviate contracting.

Since we focus on limited commitment as a potential beneficial source for mediation, we perform our analysis in a model in which contractual commitment is simply impossible.

More specifically, we contrast mediation to non–mediation in a model of cheap talk based on Crawford and Sobel (1982) in which the uninformed party, the principal, has all bargaining power. We derive necessary and sufficient conditions under which mediation is strictly helpful to the principal. More importantly, we provide a straightforward and general intuition for this result that applies to other settings in which the contractual ability of the uninformed party is limited (e.g. Dewatripont 1986; Hart and Tirole 1988 and Laffont and Tirole 1988;1990; Bester and Strausz 2001):

The non–standard feature of contracting settings with limited commitment is that from an ex ante point of view an uninformed principal may not want to obtain all information from her agent. Rather, the principal is better off if she obtains only a *partial* revelation of information. We show that a principal is more effective in obtaining a partial revelation of information if she uses a mediator than if she communicates directly with the agent herself. Partial revelation with direct communication requires that the principal cannot uniquely identify the agent's message with the agent's private information. Hence, the principal must receive messages in some stochastic way. Yet, if the principal wants to induce the agent to perform this randomization, she must ensure that the randomizing agent is kept indifferent between the allocations that his messages lead to. This is not the case if the principal employs a mediator to perform the randomization on part of the agent, because with a mediator a specific type of agent must only prefer the *mixture over allocations* that is designed for him, but need not be *indifferent between the allocations* over which the randomization occurs.[1] Hence, with a mediator the principal is less restricted in inducing a partial revelation of information and this may render mediation beneficial.

We show that mediation is only helpful if the incentives between the conflicting parties are partially aligned such that it is unsure whether a genuine conflict of interests exists.[2] We obtain three cases. First, if the ex ante probability of conflict is relatively small, mediators are helpful in increasing the amount of information that is revealed in equilibrium. In this case the mediator becomes more valuable as the ex ante probability of conflict rises. Second, when the probability of conflict lies in an intermediate range, the principal without a mediator would be unable to induce her agent to reveal any infor-

---

[1] Myerson (1985) and Forges (1986) already identified the function of the mediator as a garbling device, but do not explain why the agents are unable to perform the garbling themselves by sending messages randomly and under which circumstances this function may be helpful.

[2] Our results may therefore explain the observation of Matthews and Postlewaite (1989) that in a two–sided bargaining setting mediation is ineffective, since in a bargaining setting the incentives of the parties are diametrically opposed.

mation. Yet, with the help of a mediator information revelation is possible and desirable to the principal. Last, if the likelihood of a conflict is large, then even a mediator is unable to induce information revelation in equilibrium. Hence, the value of mediation in this range is zero. We show that the value of mediation changes continuously over the three different regions and is non–monotonic in the degree of conflict.

Given that mediators may alleviate contracting, this paper has an important implication for the general theory of contracting with limited commitment. Our result implies that in general contracting parties benefit from using a third party as a mediator. An important question is therefore whether mediators should be included in the analysis of optimal contracts. The existing literature (e.g. Laffont and Tirole 1988,1990, Dewatripont 1986, Hart and Tirole 1988) excludes mediators from its analysis. Yet, since contract theory intends to study how economic agents use contracts optimally, a consideration of mediators seems natural: If contracting parties gain by using mediators, there does not seem a reason why they will not do so.[3] This line of reasoning leads to a further observation. If mediation is generally helpful to contracting parties, then one may expect the existence of economic institutions that play this role. For instance, the use of mediation by American businesses to resolve legal disputes resulted in a completely new type of services by so–called centers for Alternative Dispute Resolution (ADR).[4] Similarly, in Mitusch and Strausz (1999) we explain consultants as playing the role of mediators in a situation of conflict within the firm. Apart from consultants, one may use a similar argument to motivate the existence of lawyers as mediating between a privately informed defendant and the court and a regulatory agency as mediating between a privately informed firm and a government susceptible to the ratchet effect.

2 Related Literature

Our result that mediation may be helpful is not new. Indeed, this paper combines and contrasts two strands of literature: the contracting literature with limited commitment (e.g. Laffont and Tirole 1988, 1991 and Hart and Tirole 1988) and the literature on communication (e.g. Myerson 1985, 1986 and Forges 1986). It is therefore worthwhile to review the similarities and differences between the two bodies of literature and to position

---

[3]The question is also relevant if one is only interested in the set of implementable allocations, as for instance in Crawford and Sobel (1982).

[4]Brown and Ayres (1994) also emphasize the mediator's role in controlling the flow of information. Based on this idea they provide additional rationales for ADR.

our own paper in relation to them. Both strands of literature study an implementation problem in which the ex ante contracting possibilities are limited. In this sense the literature on communication is more rigorous and excludes any form of commitment before communication takes place. In contrast, the literature on contracting with limited commitment allows still some limited form of contractual commitment. A further, more pragmatic difference is that the literature on communication takes a rather abstract and general approach, while the literature on contracting with limited commitment is much more application driven.

The reason for these differences may be found in the different objectives and origins of the theories. The literature on communication was developed to provide a general and uniform framework to analyze the power of communication in games with multiple players and multi–sided asymmetric information. Its generality was extended to repeated games (Forges 1988) and multi–stage games with repeated acquisition of private information (Myerson 1988). However, this literature does not address the exact nature of beneficial mediation and its intuition. Hence, the added value of this paper with respect to the communication literature is that we explicitly show how, why, and under what conditions mediation facilitates communication.

Yet, both in its focus — delivering a positive explanation of existing institutions as mediators — and its methodology — allowing the principal to optimize among contracts and equilibria — the current paper is more related to contract theory. Overall the literature on contracting with limited commitment is more pragmatic than the literature on communication, as it grew out of the concern that the standard theory of contracting made unrealistic assumptions concerning the contract designer's commitment. For instance, Dewatripont (1986) noted that many ex ante optimal contracts tend to exhibit ex post inefficiencies and argued that, in reality, contracting parties will renegotiate these inefficiencies away. In the same vein Baker et al. (1999, p. 56) assert that "decision rights in organizations are not contractible: the boss can always overturn a subordinate's decision, so that formal authority resides at the top." This effectively results in a limited commitment on part of the principal.

Directly related to the current paper is a recent paper by Krishna and Morgan (2004). The authors study different forms of communication in the cheap talk setting of Crawford and Sobel. In particular, they demonstrate that the use of multiple communication stages enable players to reach outcomes that are Pareto superior to any outcome with a single

stage of communication.[5] Moreover, they also illustrate in a numerical example that mediators may alleviate the cheap talk problem even further. With respect to Kishna and Morgan (2004) the added value of the current paper is therefore to analyze the mediator much more carefully and provide an intuition for its beneficial role. It is thereby instructive to concentrate first on a single stage of communication. This restriction leads to explicit conditions under which a mediator performs strictly better. Section 7 then demonstrates that our results are robust when we allow for more stages of communication. We will show this by providing an argument that is based on the recent work of Aumann and Hart (2003).

3   Model and Preliminaries

Consider two players, a principal and an agent. The principal must implement an option $y \in \mathbb{R}$, which affects both players. The effect of the implemented option y depends on the state of the world on which the agent is privately informed. For simplicity, we assume that the agent's information may only take on two values. With probability $1 - \pi$ he possesses the information 1 and with probability $\pi \in (0, 1)$ his private information is 2.

We suppose that both players have each some preferred option $y \in \mathbb{R}$ and the farther the implemented option is from this preferred option the more they dislike it and increasingly so. We capture this idea by assuming that the players have Von–Neumann Morgenstern utility functions which are strictly concave and attain a maximum on $\mathbb{R}$.[6] The utility functions depend on the private information of the agent. We write the principal's utility function as $V_i(y)$ when the agent has the private information $i = 1, 2$. Similarly, we denote agent i's utility function as $U_i(y)$. For technical reasons we assume that the utility functions are well defined over $\mathbb{R}$ and three times continuously differentiable.

Moreover, we adopt a *monotonicity condition* concerning the agent's utility functions:

$$U_1'(y) < U_2'(y) \quad \text{for all } y \in \mathbb{R}.$$

The condition is similar to a standard sorting condition in screening models and will fulfill a similar role. First of all, it implies that the preferred options of the agents
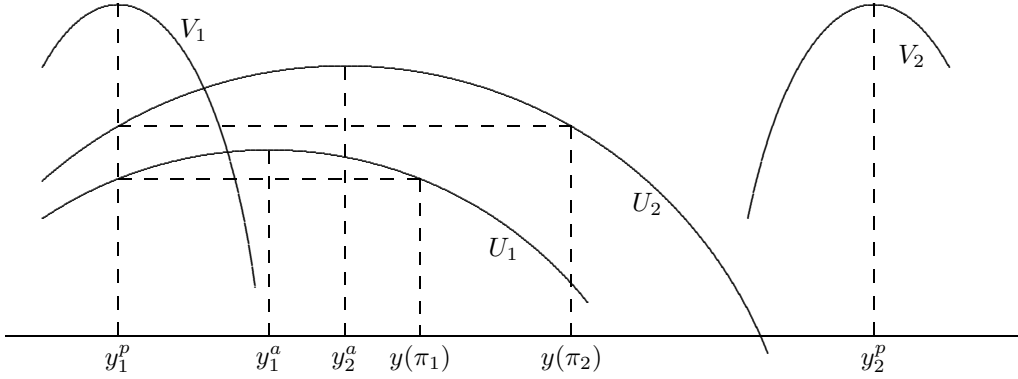
$$y_i^a \equiv \arg\max_y U_i(y)$$

Figure 1: An example

exhibit the ordering $y_1^a < y_2^a$. It implies further that if one agent is indifferent between two distinct allocations y_1 and y_2, the other agent must have a strict preference. More generally:

Lemma 1 *For any $y_1 < y_2$ the following holds.*

$$\text{If } U_1(y_1) \leq U_1(y_2) \quad then \quad U_2(y_1) < U_2(y_2).$$
$$\text{If } U_2(y_1) \geq U_2(y_2) \quad then \quad U_1(y_1) > U_1(y_2).$$

We denote similarly the principal's preferred options by

$$y_i^p \equiv \arg \max_y V_i(y).$$

We make no explicit assumptions about the relation between $V_1(y)$ and $V_2(y)$, and hence between y_1^p and y_2^p, except that $y_1^p \neq y_2^p$ so that information about i is of interest for the principal. Figure 1 illustrates the utility functions U_i and V_i for the ordering $y_1^p < y_1^a < y_2^a < y_2^p$.

This paper studies games of cheap talk. In such games the principal is unable to commit to some implementation function ex ante.[7] This distinguishes the current model from standard principal agent models with adverse selection.[8] As a consequence the implemented option will in the end only depend on the beliefs of the principal concerning

---

[7]I.e., the principal is also unable to commit to any form of conditional payments.

[8]We have nevertheless chosen the connotation principal and agent rather than receiver and sender, since we follow the standard approach of principal–agent theory and allow the principal to select among different equilibria. Moreover, we give the principal all bargaining power in connection with the mediator.

the agent's private information. Since there exist only two possible types of private information, these beliefs are fully described by some $\rho \in [0,1]$, representing the probability that the agent is of type 2. Given a belief ρ, the principal implements

$$y(\rho) \equiv \arg\max_y \ (1-\rho)V_1(y) + \rho V_2(y).$$

The following lemma gives already some indication about the possible outcome and will be helpful in the subsequent analysis.

Lemma 2 *For any belief $\rho \in [0,1]$ the option that the principal will choose, $y(\rho)$, lies in between y_1^p and y_2^p, i.e.*

$$\min\{y_1^p, y_2^p\} \leq y(\rho) \leq \max\{y_1^p, y_2^p\}.$$

Moreover, if $y_1^p < y_2^p$, then $y(\rho)$ is monotonic increasing. The function $y(\rho)$ is monotonically decreasing if $y_1^p > y_2^p$.

4 Contractual Commitment

Before analyzing the cheap talk version of the model, it is helpful to analyze the model as a standard mechanism design problem and assume that the principal can commit contractually to a mechanism before she asks her agent for information. In this version of our model the classical revelation principle applies and the optimal mechanism may be found in the set of direct mechanisms that induce the agent to reveal his information truthfully. Consequently, an optimal mechanism is the solution to the problem,

$$\max_{y_1, y_2} \ (1-\pi)V_1(y_1) + \pi V_2(y_2)$$
$$\text{s.t.} \qquad U_1(y_1) \geq U_1(y_2) \tag{1}$$
$$U_2(y_2) \geq U_2(y_1) \tag{2}$$

where inequalities (1) and (2) represent the two incentive compatibility constraints. There are no individual rationality constraints, as we assume that the principal must choose some option and no outside option exists. Alternatively, one may assume that the outside options are so low that they will not be binding, i.e. agent i's outside option is smaller than $\min\{U_i(y_1^p), U_i(y_2^p)\}$.

The solution depends on the severity of the conflict of interest between principal and agent. The following proposition identifies two extremes:

Proposition 1 *1. If $y_1^p > y_2^p$, the pooling mechanism $y_1 = y_2 = y(\pi)$ is optimal.*

2. If $y_1^p < y_2^p$, $U_1(y_1^p) \geq U_1(y_2^p)$, and $U_2(y_2^p) \geq U_2(y_1^p)$, the optimal contract is the principal's first best $(y_1, y_2) = (y_1^p, y_2^p)$.

The intuition behind Proposition 1 is straightforward. In the first case, the incentive problem between principal and agent is extremely severe. The two incentive compatibility conditions (1) and (2) imply $y_1 \leq y_2$. Yet, if $y_1^p > y_2^p$, the principal prefers to set y_1 greater than y_2. The interests of the principal and agent are diametrically opposed and the principal is unable to benefit from a separation of types. Taking the idea of a direct mechanism literally, this result implies that it is optimal for the principal to commit not to use the information which the agent's message represents. That is, if $y_1^p > y_2^p$, the principal is unable to induce information revelation in a beneficial way.

On the other hand, when preferences fulfill the conditions of Proposition 1.2 we obtain the other extreme. Here the principal can extract the agent's information costlessly and implement her first best, $y_1 = y_1^p$ and $y_2 = y_2^p$. In this setting the incentive problem is trivial and there does not exist a genuine conflict of interest between agent and principal.

Proposition 1 has two important consequences concerning our analysis of beneficial mediation. First, it shows that if $y_1^p > y_2^p$ then even with contractual commitment a principal cannot do better than offering a single, pooling contract $y = y(\pi)$. Obviously, this must also hold if the principal is unable to commit to her option y, as with full commitment she can imitate any behavior under non–commitment. Therefore, if $y_1^p > y_2^p$, the principal cannot benefit from employing a mediator. Second, if $y_1^p < y_2^p$, $U_1(y_1^p) \geq U_1(y_2^p)$, and $U_2(y_2^p) \geq U_2(y_1^p)$, the principal has no ex post incentive to deviate from the contractual implementation, y_1^p resp. y_2^p, since it is her first best. Consequently, also without commitment to y the principal will be able to implement it. Naturally, the principal cannot do better than achieving her first best, and a mediator will not be helpful. The two arguments lead to the conclusion that, under non–commitment, the only remaining constellation in which the mediator may be helpful is when $y_1^p < y_2^p$ and when $U_1(y_1^p) < U_1(y_2^p)$ or $U_2(y_2^p) < U_2(y_1^p)$ holds. Hence, in the remainder of this paper we will focus on this parameter constellation.[9]

Effectively, a type i for which $U_i(y_i^p) < U_i(y_j^p)$ $(j \neq i)$ prevents the principal from achieving her first best. This observation motivates the following definition. We say

---

[9]Since we disregarded stochastic direct mechanisms, the previous argument is not completely exhaustive in that, in the case $y_1^p > y_2^p$, the principal could possibly attain more by using stochastic mechanisms. For completeness, we will therefore return to the case $y_1^p > y_2^p$ and show for the non–commitment case that with a mediator a pooling contract is indeed generally optimal.

that the interests of type i are *incompatible* with those of the principal if and only if $U_i(y_i^p) < U_i(y_j^p)$ $(j \neq i)$. We will refer to such a type as *incompatible*.

Lemma 3 *If $y_1^p < y_2^p$, there exists at most one incompatible type.*

Since Proposition 1 indicates that the question of beneficial mediation is uninteresting if neither type is incompatible, we assume in the following that there exists an incompatible type. Given $y_1^p < y_2^p$, Lemma 3 shows that due to the monotonicity condition, there is at most one incompatible agent. We assume that this is type 2. This assumption is without loss of generality, because if the incompatible type is type 1, then we may "mirror" our problem by redefining the options as $y' = -y$ and exchange the roles of type 1 and 2.

Assumption 1 *The agent of type 2 is incompatible and $y_1^p < y_2^p$.*

Since only the incompatible type leads to a conflict, Assumption 1 implies that the parameter π measures the probability of conflict between the agent and the principal. In order to arrive at a more intuitive classification of the ex ante probability of conflict, we introduce two threshold levels $\pi_1 \geq 0$ and $\pi_2 \geq 0$. Let π_i be such that

$$\pi_i \equiv \arg \max_{\rho \in [0,1]} \{\rho | U_i(y(\rho)) = U_i(y_1^p)\}.$$

Note that $\pi_1 \leq \pi_2 < 1$, and that $\pi_i > 0$ if and only if $y_i^a > y_1^p$. Figure 1 illustrates the two thresholds for the case $y_1^p < y_1^a < y_2^a < y_2^p$ with $0 < \pi_1 < \pi_2 < 1$.[10]

Given π_1 and π_2, we use the following classification of the ex ante probability of conflict. We say that the ex ante probability of conflict is *small* if $\pi < \pi_1$ and that the ex ante probability of conflict is *large* if $\pi > \pi_2$. Note that if $y_1^a \leq y_1^p$ then there does not exist a small ex ante probability of conflict, because in this case $\pi_1 = 0$. Likewise, if $y_2^a \leq y_1^p$ then $\pi_2 = 0$ and any π represents a large probability of conflict.

Proposition 2 *Suppose Assumption 1 holds, then the optimal separation contract exhibits $y_1 < y_1^p$ and $y_2 < y_2^p$, and leaves agent 2's incentive constraint binding.*

If type 2 is incompatible, the principal's first best is not attainable, since it violates type 2's incentive constraint. Yet, for $y_1^p < y_2^p$ the requirements of the incentive constraints, that $y_1 \leq y_2$, are nevertheless aligned with the principal's preferences. In

---

[10]The other two possible cases are $y_1^a \leq y_1^p < y_2^a < y_2^p$ with $\pi_1 = 0 < \pi_2 < 1$ and $y_1^a < y_2^a \leq y_1^p < y_2^p$ with $\pi_1 = \pi_2 = 0$.

contrast to the case $y_1^p > y_2^p$, the principal may therefore prefer a separation contract to a pooling one. The optimal separation contract requires that y_1 and y_2 are smaller than y_1^p and y_2^p respectively. The choice of $y_2 < y_2^p$ is intuitive: Starting from the first best (y_1^p, y_2^p) —which violates type 2's incentive compatibility constraint (2)— lowering y_2 relaxes the constraint. At first sight it may be surprising that it is optimal to set y_1 below y_1^p, since a y_1 lower than y_1^p reduces the principal's utility from a truthful revelation of type 1. Yet, starting from the principal's first best this loss is only of the second order, since y_1^p is the optimal restructuring choice under type 1, i.e. $V_1'(y_1^p) = 0$. In contrast, a y_1 lower than y_1^p relaxes type 2's incentive constraint, which represents a first order gain.

Concerning our question of beneficial mediation, Proposition 2 reveals two important features of the optimal separation mechanism. First, the options prescribed by the optimal revelation contract are suboptimal ex post. Since the agent reveals himself perfectly by his choice of contract, the principal will have an ex post incentive to implement the options y_1^p and y_2^p rather than the options prescribed by the mechanism. The credibility of the principal's commitment is therefore crucial. In the cheap talk version of our implementation game there is no such commitment and the ex ante incentives to report truthfully are destroyed.

Second, the optimal information revealing mechanism commits the principal to an option $y_1 < y_1^p$. Yet, Lemma 2 established that there does not exist a belief for the principal that would lead to such a choice. Therefore, when the principal has no possibility to commit herself, she would never take this option. In the cheap talk version of the model the principal can therefore not achieve the outcome of the optimal separation mechanism. As this result is independent of whether the principal uses a mediator, it shows that we cannot expect the mediator to mitigate completely the limitations due to a lack of commitment of the principal.

5   Direct Communication

In the following we assume that the principal is unable to commit to a mechanism. This transforms the implementation problem into a game of cheap talk in which the principal cannot propose a menu from which the agent may pick his preferred option. Nevertheless, the principal may want to communicate with her agent in the hope that this leads to some revelation of information.

In this section we assume that communication must take place directly between prin-

cipal and agent. The direct communication game is as follows:[11]

1. The principal sets some message space M for the agent.

2. The agent announces a message $m \in M$.

3. The principal updates her beliefs.

4. The principal chooses an option y.

We apply the solution concept of Perfect Bayesian Equilibrium (PBE) to this game. Such an equilibrium specifies a message space M, an announcement strategy α_i for the agent, a belief $\rho = (\rho_1, \ldots, \rho_{|M|})$ of the principal, and an implementation strategy $y = (y_1, \ldots, y_{|M|})$. That is, if the agent sends the message $m \in M$, the principal's belief that the agent is of type 2 is ρ_m and induces her to implement restructuring option y_m. Since we are interested in the question whether the principal can do strictly better with a mediator than without, we will concentrate on the PBE that yields the principal the highest utility.[12]

Due to a generalized revelation principle proven in Bester and Strausz (2001), we may without loss of generality assume that the message space corresponds to the set of types, i.e. $M = \{1, 2\}$. This implies that the agent effectively announces some type i. Consequently, we may represent a strategy of agent i by some $\alpha_i \in [0, 1]$ which denotes the probability that the agent announces that he is of type 1. Moreover, there is no loss of generality in assuming that $\alpha_1 > 0$, $\alpha_2 < 1$, and $\alpha_1 \geq \alpha_2$.[13]

Thus we will look for a PBE with $M = \{1, 2\}$. The combination $(\alpha_1, \alpha_2, \rho_1, \rho_2, y_1, y_2)$ constitutes a Perfect Bayesian Equilibrium if it satisfies the following three conditions:

1. The agent's announcement strategy is optimal given the principal's implementation strategy, i.e.

$$\alpha_i U_i(y_1) + (1 - \alpha_i) U_i(y_2) = \max_\alpha \ \alpha U_i(y_1) + (1 - \alpha) U_i(y_2). \tag{3}$$

---

[11] We first restrict attention to a single stage of direct communication. See Section 7 for a discussion of this assumption.

[12] It is well known that in cheap talk games there always exists an uninformative "babbling" equilibrium yielding no information revelation.

[13] This implies that the agent tells the truth with a strict positive probability, but, in contrast to the standard revelation principle, it may be optimal for the principal to let some type lie with a positive probability. For more details see Bester and Strausz (2001).

2. The principal's belief is Bayes' consistent with the agent's strategy, whenever possible. This implies that

$$\rho_1 = \rho(\alpha_1, \alpha_2) \quad \text{and} \quad \rho_2 = \rho(1 - \alpha_1, 1 - \alpha_2),$$

with

$$\rho(x_1, x_2) \equiv \frac{x_2 \pi}{x_1(1 - \pi) + x_2 \pi}. \tag{4}$$

Note that since $\alpha_1 \in (0, 1]$ and $\alpha_2 \in [0, 1)$ both $\rho(\alpha_1, \alpha_2)$ and $\rho(1 - \alpha_1, 1 - \alpha_2)$ are well–defined.

3. The principal's implementation strategy is optimal given her belief p, i.e.

$$y_i = y(\rho_i).$$

In equilibrium, the agent's strategy combination (α_1, α_2) yields the principal the utility

$$V(\alpha_1, \alpha_2) = (1 - \pi)[\alpha_1 V_1(y(\rho(\alpha_1, \alpha_2))) + (1 - \alpha_1)V_1(y(\rho(1 - \alpha_1, 1 - \alpha_2)))]$$
$$+ \pi[\alpha_2 V_2(y(\rho(\alpha_1, \alpha_2))) + (1 - \alpha_2)V_2(y(\rho(1 - \alpha_1, 1 - \alpha_2)))].$$

The principal's utility is increasing in α_1 and decreasing in α_2.[14] This reflects the intuitive fact that more information is better for the principal. Since the principal's utility depends on the degree of information the agent reveals, it will be helpful to distinguish between the following five classes of equilibria:

1. A *full revelation* equilibrium in which the agent's type is perfectly revealed: $\alpha_1 = 1$, $\alpha_2 = 0$.

2. A *non–revelation* equilibrium in which the agent's announcement does not reveal anything: $\alpha_1 = \alpha_2$.

3. A *partial revelation* equilibrium in which the announcement of each agent reveals some, but not all information: $\alpha_1 < 1$, $\alpha_2 > 0$, and $\alpha_1 \neq \alpha_2$.

4. A *type 1 partially full revelation* equilibrium that leads to a full revelation of agent 1 with positive probability, but not of agent 2: $\alpha_1 < 1$ and $\alpha_2 = 0$.

---

[14]The envelope theorem yields $dV/d\alpha_1 = (1 - \pi)(V_1(y_1) - V_1(y_2)) \geq 0$ and $dV/d\alpha_2 = -\pi(V_2(y_2) - V_2(y_1)) \leq 0$. The sign follows due to $\alpha_1 \geq \alpha_2$, which implies $y_1 = y(\rho(\alpha_1, \alpha_2)) \leq y_2 = y(\rho(1 - \alpha_1, 1 - \alpha_2))$ in equilibrium.

5. A *type 2 partially full revelation* equilibrium that leads to a full revelation of agent 2 with positive probability, but not of agent 1: $\alpha_1 = 1$ and $\alpha_2 > 0$.

Under Assumption 1 a full revelation equilibrium does not exist. In such an equilibrium the principal chooses y_1^p and y_2^p, which leads agent 2 to pool with agent 1 rather than revealing himself truthfully. As is familiar from the literature on cheap talk, a non–revelation equilibrium always exists, but yields the principal less than any other class of equilibrium. Any of the remaining classes involves at least one message that reveals the agents only partially, which requires that both agents use this message with positive probability. The agent who uses also the other message is actively mixing over the two messages and must therefore be indifferent between the allocations which they induce. Due to the monotonicity assumption the two agents cannot be indifferent between two different allocations at the same time. Hence, a partial revelation equilibrium will not exist.

Now consider the two partially full revelation equilibria. A type i partially full revelation equilibrium implies $y_i = y_i^p$ and requires, first, that agent i is indifferent between y_i^p and the other option $y_j \neq y_i^p$ ($j \neq i$) while, second, agent j always prefers y_j so that he reveals himself truthfully. First consider the type 2 partially full revelation equilibrium. It implies $y_1^p < y_1 < y_2 = y_2^p$ and requires that agent 2 is indifferent between y_1 and y_2^p. However, under Assumption 1 agent 2 is incompatible, which implies that in the range $[y_1^p, y_2^p]$ the outcome y_2^p is his worst possible outcome. Therefore, he will strictly prefer any $y_1 \in [y_1^p, y_2^p)$ and an equilibrium of class 5 does not exist.

Hence, under Assumption 1 the only remaining candidate besides the non–revelation equilibrium is an equilibrium of class 4 which leads to a full revelation of type 1 with positive probability. This equilibrium implies $y_1 = y_1^p < y_2 < y_2^p$ and requires that agent 1 is indifferent between the allocation y_1^p and a different allocation y_2. This is only possible if $y_1^a > y_1^p$. Hence, if $\pi_1 = 0$ only the non–revelation equilibrium exists. On the other hand, if $\pi_1 > 0$, the concavity of the agent's utility function implies that there exists exactly one restructuring option $y_2 > y_1^p$ for which this indifference obtains. Namely, $y(\pi_1)$, as illustrated in Figure 1. The equilibrium therefore exists if there exists a mixing behavior of agent 1 which leads to the belief π_1 upon observing the message 2. Bayes' consistent updating implies that this is the case if and only if the ex ante probability of conflict is small, i.e., $\pi < \pi_1$. We therefore arrive at the following proposition.

Proposition 3 *Suppose Assumption 1 holds. Then the optimal Perfect Bayesian Equilibrium with direct communication exhibits the following structure:*

1. *If $\pi < \pi_1$, the optimal PBE is $(y_1, y_2) = (y_1^p, y(\pi_1))$ and agent 1 is perfectly revealed with probability $(\pi_1 - \pi)/((1 - \pi)\pi_1) > 0$.*

2. *If $\pi \geq \pi_1$, the optimal PBE is $(y_1, y_2) = (y(\pi), y(\pi))$ and no information is revealed in equilibrium.*

A direct comparison between Proposition 2 and Proposition 3 reveals that non–commitment not only makes it more difficult for the principal to induce information revelation, it may actually make it impossible. Only if the ex ante probability of a conflict of interest is small ($\pi < \pi_1$), is the principal able to extract information from the agent. But also in the informative equilibrium there remain three sources of inefficiencies as compared to the commitment case. First, the allocation y_1 is suboptimally high, as without commitment it is not possible for the principal to implement an option $y < y_1^p$. Second, there is a "stochastic misallocation", since agent 1 misrepresents his type with positive probability. Third, the allocation y_2 is suboptimally low as compared to the solution under full commitment. The latter two inefficiencies have the same origin: In order to induce agent 1 to mix, the principal must make him indifferent between the two equilibrium outcomes. Stochastic misallocation is therefore a necessary feature for information revelation under non–commitment and, in contrast to the full–commitment case, agent 1 rather than agent 2 is made indifferent in equilibrium.

An interesting interpretation of the solution is what may be called an "underrevelation principle": In equilibrium information revelation is only possible if the compatible type, who has no problem to reveal himself when the principal offers her two preferred options, underreveals himself. The imperfect revelation of type 1 provides cover for the incompatible type 2, making information revelation possible. In fact, inducing type 1 to provide such cover for agent 2 is the principal's main problem. She has to choose her option y_2 in such a way, that type 1 is indeed willing *not* to reveal himself completely. Since her choice y_2 has to be Bayesian incentive compatible, it limits the amount of information that can be revealed in equilibrium and restricts the set of parameter constellations for which the principal can induce information revelation.

6 Mediated Communication

In this section we allow the principal to employ a third party, the mediator, who may help with the communication between principal and agent.[15] The mediator's role is to

---

[15]In contrast to standard models of third–party delegation, the principal in our framework does not delegate the final implementation decision. It should be clear that this makes the role of the third party
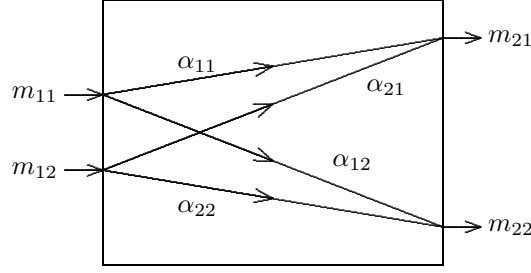
Figure 2: A communication rule

communicate first with the agent and then with the principal. Since the principal employs the mediator, we assume that she designs the exact rules of communication. A general communication rule prescribes the following. First, it specifies a message space M_1 from which the agent has to send a message to the mediator. Second, it specifies a message space M_2 from which the mediator sends a message to the principal. Third, it specifies the probability with which the mediator sends a message $m_2 \in M_2$ when the agent sent the message $m_1 \in M_1$. A communication structure P may therefore be written as a tuple (M_1, M_2, α) where α maps M_1 into a probability distribution over M_2. Figure 2 illustrates a communication rule with two messages for the agent and two for the mediator.

The game between the principal and the agent when a mediator is available runs as follows:

1. The principal announces publicly the mediator's communication rule $P = (M_1, M_2, \alpha)$.

2. The agent sends a message $m_1 \in M_1$ to the mediator. The message is communicated in private such that the principal does not observe it.

3. The mediator sends the message $m_2 \in M_2$ according to the probability distribution $\alpha(m_1)$ to the principal.

4. The principal updates her beliefs and decides which project to implement.

Note that the principal's choice of a communication rule P at stage 1 induces a proper subgame as of stage 2. A Perfect Bayesian Equilibrium of this subgame describes for each type of agent an announcement strategy, which may be represented by a probability distribution over the set M_1, and an implementation strategy for the principal that describes

---

much weaker. In line with standard literature on delegation, we assume that there exist no possibilities of collusion.

which option $y \in \mathbb{R}$ the principal chooses given the mediator's message. In principle also the principal's strategy may involve randomization. Last, a Perfect Bayesian Equilibrium describes a belief function ρ for the principal, which represents the belief of the principal given that the mediator sent a message $m_2 \in M_2$. Similarly to the previous section, a Perfect Bayesian Equilibrium has to satisfy three requirements: 1) the agent's announcement strategy is optimal given the principal's implementation strategy; 2) the principal's belief is Bayes' consistent with the agent's strategy, whenever possible; and 3) the principal's implementation strategy is optimal with respect to her beliefs.

Importantly, the mediator's description coincides with his role in the literature on communication. The following lemma expresses an important result of this literature.

Lemma 4 *Without loss of generality the principal may restrict attention to communication rules for which the message of the agent is his type and the message to the principal is a recommendation about the option y. Moreover, the principal may restrict attention to communication rules that are (Bayesian) incentive compatible, i.e., induce the principal to follow the mediator's recommendation and induce the agent to report his type truthfully.*

Lemma 4 is a generalized version of the classical revelation principle.[16] It shows that one may assume without loss of generality that the optimal communication rule uses a message space $M_1 = \{1, 2\}$ for the agent and the message space $M_2 = \mathbb{R}$ for the mediator. That is, we may restrict attention to communication rules which give an intuitive role to the mediator and is consistent with standard observation of mediation in real–life: The mediator first gathers information during private consultations and then makes a public proposal.

Due to the revelation principle we have only to consider incentive compatible communication rules $P = (\{1, 2\}, \mathbb{R}, \alpha_1, \alpha_2)$ with α_i a probability measure over $\mathbb{R}$. To circumvent measure–theoretical considerations we restrict attention to the class of communication rules that randomize over a finite, but arbitrarily large number of recommendations in $\mathbb{R}$. That is, we consider communication rules of the form $P = (\{1, 2\}, R, \alpha_1, \alpha_2)$ with $R \subset \mathbb{R}$ finite and $\alpha_i \in \mathbb{R}_+^{|R|}$ and $\sum_{j=1}^{|R|} \alpha_{ij} = 1$ for $i = 1, 2$. Without further loss of generality we adopt the following ordering assumption

$$\alpha_{2(j+1)}\alpha_{1j} \geq \alpha_{1(j+1)}\alpha_{2j} \tag{5}$$

for all $j = 1, 2, \ldots, |R| - 1$.

---

[16]For details see Myerson (1985, 1986) and Forges (1986).

An incentive compatible communication rule entails two different forms of incentive compatibility. First, the recommendations must be incentive compatible in the sense that the principal has no strict incentive to diverge from the mediator's proposal. For a recommendation $r_j \in \mathbb{R}$ this obtains if

$$r_j = y(\rho(\alpha_{1j}, \alpha_{2j})) \tag{6}$$

where $\rho(\cdot)$ is given by (4) and ensures Bayes' consistent updating. We call a recommendation r_j for which equality (6) holds incentive compatible. Under Assumption 1 $y(\rho)$ is increasing and the ordering condition (5) implies that an incentive compatible communication rule exhibits $r_j \leq r_{j+1}$ for all $j = 1, 2, \ldots, |R| - 1$.

Second, the communication rule must be incentive compatible in the sense that the agent does not have a strict incentive to misreport his type. A communication rule P is *incentive compatible with respect to type 1* if

$$\sum_j \alpha_{1j} U_1(r_j) \geq \sum_j \alpha_{2j} U_1(r_j). \tag{7}$$

A communication rule P is *incentive compatible with respect to type 2* if

$$\sum_j \alpha_{2j} U_2(r_j) \geq \sum_j \alpha_{1j} U_2(r_j). \tag{8}$$

A communication rule P is incentive compatible if all its recommendations are incentive compatible and if it is incentive compatible with respect to both types. As is well known, the need for incentive compatibility puts restrictions on the set of implementable communication rules:

Lemma 5 *Suppose Assumption 1. If an incentive compatible communication rule P induces some revelation of information then*

1. it holds $\pi < \pi_2$.

2. if one type's incentive constraint holds with equality, the other one's is satisfied with strict inequality.

3. there must be recommendations $r_j \in R$ such that $y(\pi) < r_j < y_2^p$.

Lemma 5 shows in particular that if the ex ante probability of conflict is large ($\pi \geq \pi_2$), information revelation is impossible. In this case the principal does not benefit from the mediator. Given this result we proceed by showing that a mediator is indeed helpful if the probability of conflict is not large. We thereby focus first on *incentive compatible 2–proposal rules* for which the number of proposals to the principal

coincides with the number of types. An incentive compatible 2–proposal rule has the form $P = (\{1, 2\}, \{r_1, r_2\}, (\alpha_{11}, \alpha_{12}), (\alpha_{21}, \alpha_{22}))$ and is illustrated in Figure 2. The ordering assumption (5) implies $\alpha_{11} \geq \alpha_{21}$, i.e. the recommendation r_1 (weakly) indicates that the agent is of type 1, while the recommendation r_2 is more indicative of type 2. Moreover, using $\alpha_{i2} = 1 - \alpha_{i1}$, the incentive compatibility conditions with respect to the agents, (7) and (8), reduce to

$$U_1(r_1) \geq U_1(r_2) \quad \text{and} \quad U_2(r_2) \geq U_2(r_1), \tag{9}$$

respectively. Note that these constraints coincide with the incentive compatibility conditions (1) and (2) of the full commitment framework. The incentive compatibility conditions with respect to the principal are given by (6) for $j = 1, 2$. Given these incentive constraints an optimal incentive compatible 2–proposal rule is a solution to the following maximization problem:

$$\max_{\alpha_{11}, \alpha_{21}, r_1, r_2} V(P) \equiv (1 - \pi)\{\alpha_{11}V_1(r_1) + (1 - \alpha_{11})V_1(r_2)\}$$
$$+ \pi\{\alpha_{21}V_2(r_1) + (1 - \alpha_{21})V_2(r_2)\}$$
$$\text{s.t.} \qquad \text{(6) and (9)}.$$

In order to derive the optimal proposal rule it is helpful to introduce the following definition of informativeness.[17] Consider two incentive compatible proposal rules in which all recommendations except for two of them are identical, i.e. with identical $(\alpha_{1j}, \alpha_{2j})$ for all $j \neq k, l$. We say that a proposal rule that includes the pair (r'_k, r_l) is *more informative* than a proposal rule that includes the pair (r_k, r_l) if $|r'_k - r_l| > |r_k - r_l|$. This definition is motivated by incentive compatibility, since the distance between r_k and r_l is larger only if these recommendations are more discriminative between the two types.

Lemma 6 *The principal's utility is increasing in the informativeness of an incentive compatible recommendation pair* (r_k, r_l).

Lemma 6 shows that our notion of informativeness is consistent with the intuitive idea that more information is better for the principal. Yet, the result is not trivial, since increasing informativeness has both a positive and a negative effect. Clearly, a more informative recommendation enables the principal to tailor her options more accurately. A negative effect, however, is caused by the need for incentive compatibility. If a pair

---

[17]Note that this definition and the following Lemma 6 hold for *general* proposal rules, not only 2–proposal rules, and provides the basis for all the optimality results derived in this paper (in particular Propositions 4 and 5).

(r'_k, r_l) is more informative than a pair (r_k, r_l), then incentive compatibility requires that both types induce the recommendation r_k less often. That is, also the type of agent for which recommendation r_k is indicative. This is a negative effect. Lemma 6 shows that due to the concavity of V_i the positive effect outweighs the negative one.

Having established that more informative, incentive compatible recommendations are beneficial, we are able to derive the optimal 2–proposal rule in the case that the ex ante probability of conflict is not high.

Proposition 4 *Suppose Assumption 1 and $\pi < \pi_2$. The optimal incentive compatible 2–proposal rule is $(y_1, y_2) = (y_1^p, y(\pi_2))$. The incentive constraint of agent 2 binds and agent 1 is perfectly revealed with probability $(\pi_2 - \pi)/((1 - \pi)\pi_2) > 0$.*

The optimal 2–proposal rule resembles the equilibrium of the game with direct communication. In both equilibria the incompatible agent is not fully revealed. More importantly, also the compatible agent 1 is not revealed completely, but pools with a positive probability with the incompatible type. Even though he is the compatible, unproblematic type from the principal's perspective, his type is underrevealed in order to provide cover for the incompatible agent so that this latter agent is never fully exposed.

The important difference between the two equilibria is that the degree of underrevelation of type 1 is less when the principal uses a mediator. This difference constitutes the beneficial effect of the mediator. He is able to provide cover for agent 2 more efficiently and thereby attain more informative allocations than the principal. Yet, type 2's incentive constraint restricts the mediator in inducing information revelation and the constraint is binding at the optimum.[18]

Since Lemma 5 shows that for $\pi > \pi_2$ mediators are not helpful to the principal, we arrive at our main result.[19]

Theorem 1 *Mediation is strictly beneficial to the principal if and only if the following conditions hold: (i) $y_1^p < y_2^p$, (ii) type 2 is incompatible, and (iii) $\pi \in (0, \pi_2)$.*

---

[18]Note the similarity with the optimal full–commitment contract.

[19]For completeness, we show in the appendix that for $y_1^p > y_2^p$, the mediator is unable to induce information revelation with a *general* communication rule. Hence, the equilibrium outcome coincides with the pooling equilibrium outcome, as already indicated by Proposition 1 (see also footnote 9). Moreover, if a mediator is unable to induce information revelation for $y_1^p > y_2^p$, then this necessarily also holds with direct communication, since a mediator may mimic any equilibrium of the direct communication game.
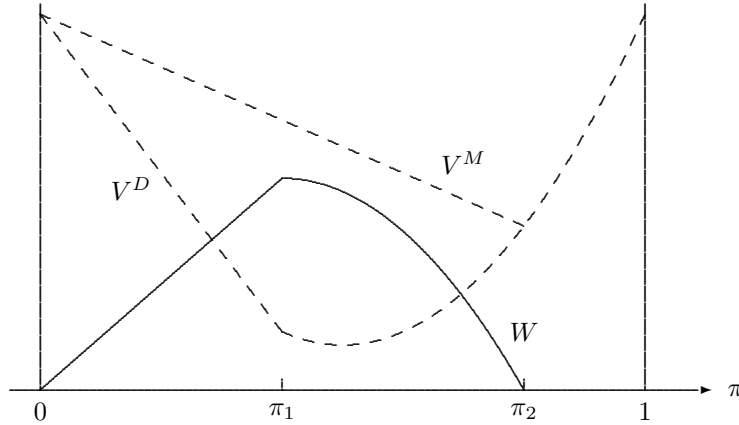
Figure 3: The value of mediation

We may explain the beneficial effect of the mediator by referring to the equilibrium requirement (3) of direct communication and the incentive constraints (9) of mediated communication. The principal's utility is increasing in the amount of information that is revealed. Due to her limited commitment, however, the principal is unable to induce full information revelation and can achieve at most a partial revelation of information by inducing an underrevelation of the compatible type. Without mediation underrevelation requires that the compatible type actively mixes over his messages. To induce such mixing, he must be kept indifferent between the two allocations. This is expressed by the equilibrium requirement (3). In contrast, with a mediator the agent's indifference is not required. Since the mediator performs the mixing, the compatible agent only has to prefer his mixing package over the mixing package of the other type. This requirement yields the incentive constraints (9), which are weaker than the equilibrium requirement (3).

The Theorem shows that the benefit of mediation depends on the ex ante probability of conflict π. Figure 3 illustrates this result graphically by drawing the principal's payoff associated with the optimal unmediated contract, V^D, and the optimal mediated contract, V^M. It demonstrates that, at low probabilities of conflict, both payoffs decrease, whereas they rise for larger values of π. The reason for this non–monotonicity is that the parameter π affects the amount of asymmetric information between the principal and agent in a non–monotonic way. For low values of π an increase raises the amount of asymmetric information, whereas for high values of π the degree of asymmetric information is reduced. At the extremes $\pi = 0$ and $\pi = 1$ there is no asymmetric information so that V^D and

V^M coincide. More importantly however, the Figure shows that, starting from $\pi = 0$, the decrease in the principal's payoff is stronger, if she does not use a mediator. This reflects that a principal is in a better position to deal with asymmetric information when she uses a mediator. In this case, a rise in the amount of asymmetric information hurts the principal less.

By introducing the difference $W \equiv V^M - V^D$ we may discuss the value of mediation to the principal. The figure illustrates that as a function of π the value W is hump–shaped and attains a unique maximum.[20] Considering that mediators are costly to employ in real life, this typical shape of W has an important empirical implication. With costly mediation, the interval of π in which mediation is beneficial shrinks from both sides. Hence, if mediators are costly and the probability of conflict is quite low, the principal does not resort to a mediator, but relies on direct communication. Consequently, one would see mediation only for intermediate levels of π, i.e. in situations in which the probability of conflict is neither too high nor too low. These two implications seem consistent with stylized facts about mediation and are testable empirically.

7   More General Forms of Communication

In the preceding sections we analyzed optimal communication under two restrictions. On the one hand, we allowed the principal and agent to communicate for only one round. On the other hand, we allowed the mediator to choose between only two recommendations to the principal. A priori it is however not clear whether these assumptions imply real restrictions or whether they can be made without loss of generality. This section addresses this question. We first turn to our restriction on mediation.

Allowing more than two recommendations leads to an artificial randomness of the communication rule. From standard theory of mechanism design it is well known that artificial randomness may relax incentive constraints and may therefore be part of an optimal mechanism even when players are risk averse. This result extends to our setting with limited commitment and prevents a simple characterization of the optimal contract. Instead, the following partial characterization of the optimal communication rule may be obtained.

---

[20]Straightforward calculations show that W is strictly concave on $[\pi_1, \pi_2]$. If $\pi_1 > 0$, it is linearly increasing on $[0, \pi_1]$. Consequently, W s quasi–concave in π and has a unique maximum.

Proposition 5 *Suppose the necessary and sufficient conditions of Theorem 1 for benefi-cial mediation are met. Then an optimal communication rule has the following properties:*

    (i) Agent 2's incentive constraint (8) binds.

    (ii) It holds $r_1 = y_1^p$.

    (iii) For all $j > 1$ it holds $r_j > \max\{y(\pi), y_2^a\}$.

Property (ii) implies that agent 1 is revealed with a probability strictly between zero and one. On the other hand, property (iii) implies that all recommendations except r_1 are more indicative of agent 2 than of agent 1.

Reflecting the fact that stochastic schemes may be optimal, we cannot exclude that the optimal communication rule uses more than two recommendations. However, a standard approach in implementation theory is to derive sufficient conditions on the risk attitudes of the parties that render stochastic mechanisms suboptimal. Naturally, these conditions regard the parties whose incentive constraints bind at the optimum. Since in the present model there exist incentive constraints with respect to the agent as well as to the principal, the conditions involve the risk attitudes of both parties. The following proposition shows that if the principal's utility function exhibits decreasing absolute risk aversion and the combined absolute risk aversion of principal 1 and type 2 concerning the option y is large enough the optimal rule does not involve artificial randomness.[21]

Proposition 6 *Given the conditions of Theorem 1, a sufficient condition for a 2–proposal rule to be optimal is that $V_i''' \geq 0$ for $i = 1, 2$ and for all $y \in (y_2^a, y_2^p)$ it holds that*

$$\frac{U_2''(y)}{U_2'(y)} + 2\frac{V_1''(y)}{V_1'(y)} \geq -\frac{V_1'(y)}{V_1(y_1^p) - V_1(y)}. \tag{10}$$

Under the conditions stated in Proposition 6 the optimal incentive compatible 2–proposal rule that we characterized in Proposition 4 is *generally* optimal.

We now turn to the restriction in the setting with direct communication, where we focused on one round of communication by the agent only. At first sight this seems innocuous, because only the agent has private information and there is nothing to learn from the principal.[22] However, Forges (1990) shows that even with one–sided asymmetric

---

[21]Note that the conditions are satisfied for any function U_2 if $V_1''' = 0$ and $V_2''' \geq 0$. E.g., a quadratic V_i.

[22]Indeed, if only the agent sends messages, Bester and Strausz (2001) demonstrate that there is no loss of generality in restricting attention to one round of communication and requiring that the agent sends only messages about his type.

information players may benefit from multiple rounds of communication with the following alternating structure. In odd–numbered rounds the agent reveals some information and in even–numbered rounds the principal and agent send messages simultaneously. Aumann and Hart (2003) explain that the simultaneous messages enable players to use joint lotteries to coordinate their behavior. More specifically, Krishna and Morgan (2004) demonstrate in the setting of Sobel and Crawford (1982) that with only two such alternating rounds of communication players can already reach a Pareto superior equilibrium outcome. The joint lotteries in Krishna and Morgan thereby partition the set of types in smaller subsets. Hence, their example is not applicable to our model, where there are only two types of agents. However, since one cannot guarantee that joint lotteries are not helpful for other reasons, we address the issue explicitly. The problem thereby is that there is not a general procedure to solve for the equilibrium outcomes.[23] Nevertheless, we may demonstrate that under Condition (10) mediation is strictly optimal even if we consider multiple stages of communication with jointly controlled lotteries. We start by noting that any outcome from a multi–stage communication can be implemented via a mediator.

Consider first communication with a finite number of stages. By Proposition 6 the unique optimal mediated mechanism induces only the two allocations y_1^* and y_2^*. Hence, if there exists an equilibrium with multi–stage direct communication that matches the payoffs associated with the optimal mediated mechanism, it may not lead to any outcome other than y_1^* and y_2^*. Now suppose such a multi–stage communication equilibrium exists, then by Theorem A of Aumann and Hart (2003) there exists a payoff–equivalent *canonical* multi–stage communication rule where in even–numbered periods a joint lottery is played and in odd–numbered periods the agent sends unilateral signals to the principal. Now consider the last communication stage, say l, of this equilibrium. If l is even, then the last stage is a joint lottery. However, a joint lottery by itself does not influence the principal's beliefs; it only serves as a substitute for a public random variable.[24] Therefore, since for a given belief the principal's response is unique, the joint lottery is inconsequential and there exists a payoff–equivalent canonical equilibrium with $l-1$ stages. It follows that if there exists an equilibrium with finitely many stages of direct communication that matches the payoffs associated with the optimal mediated mechanism, then there exists a payoff equivalent canonical equilibrium with k stages, where k is odd. Now if $k > 1$, there

---

[23]Aumann and Hart (2003) take an important step towards such a procedure, but their framework is only capable of handling fully specified settings. See also Krishna (2004) for an application of Aumann and Hart (2003) to specific settings of cheap talk.

[24]See, for example, Forges (1990).

exists a stage $k-1$ where the principal and the agent play a joint lottery. But independent of the outcome of this lottery the principal must after a subsequent signal of the agent in stage k either choose y_1^* or y_2^*. Due to her unique response function $y(.)$, the agent's signal in the final stage k must for either outcome of the joint lottery induce identical belief systems. Hence, the previous joint lottery at $k-1$ is inconsequential and may be dispensed with, and there exists an equivalent canonical equilibrium with only $k-2$ alternating stages. Repeating this argument shows that if there exists an equilibrium with multi–stage direct communication that matches the payoffs associated with the optimal mediated mechanism, then there exists one with only one round. Yet, by Theorem 1 this is not the case. It follows that there does not exist an equilibrium with multi–stage communication that is payoff equivalent to the optimal mediated mechanism.[25]

Consider now communication with an infinite number of stages. In this case the obvious advantage of a mediator is that he may implement the outcome in a single period rather than an infinite amount of time. Indeed, if one introduces a discount factor to reflect the cost of delay, then it is immediate that a mediator is strictly beneficial under the conditions of Theorem 1.

8 Concluding Remarks

In this paper we studied mediation in a model of cheap talk. As is well known, cheap talk is helpful only if there exists both a conflict and some shared interest between the two players. We introduced the ex ante probability π as a measure of this shared interest and showed that mediation enlarges the range for which information revelation can be induced in equilibrium. Moreover, for a fixed degree of shared interest, i.e. a fixed π, mediation increases the amount of information that can be revealed in equilibrium. These two effects may lead to a demand for mediators in situations of conflict.

We close this paper with a discussion about the generality and possible extensions of our framework:

Imperfect commitment and an "underrevelation" principle — We provided an intuition for our results by stressing the inability of commitment by the principal. Indeed, with full commitment a mediator is not helpful, since the principal can simply

---

[25]Effectively, our argument exploits the fact that multiple stages are only helpful if they induce additional allocations. However, the proof of Proposition 6 shows that there is a better mediation equilibrium with only two allocations.

commit herself to any behavior of the mediator. The inability to commit is the principal's central problem. It leads her to respond myopically to a supply of information, thus discouraging the agent to reveal himself. Under imperfect commitment partial revelation requires an underrevelation of the compatible type even though this type is by definition willing to reveal himself truthfully. An underrevelation of the compatible type is nevertheless required to provide cover for the other type so that the latter is never fully exposed.[26] This requirement restricts the potential to communicate, with or without mediator, and leads to inefficiencies, which are smaller when a mediator is available.

A mediator alleviates the principal's commitment problem to some degree. Yet, the enhanced commitment is rather subtle. It does not address the commitment problem directly, as also with a mediator the principal reacts myopically to information. Instead, the enhanced commitment is found in the way the principal can process information. With a mediator it is *as if* the principal is able to commit to a specific garbling of information before acting upon it. This she cannot do without a mediator, since to apply the correct garbling probabilities the agent's type must be known.

The mediator as an economic agent — According to Smith (1995) *"One of the hallmarks of mediation, and one of its important advantages, is mediation's generally private, confidential nature. Mediation's confidentiality may be one of the main reasons for its success in creating settlements. Parties are often unwilling to disclose confidential information about their view of the case to the opposing party during direct negotiation. Perhaps they intend to use the information for the first time at trial, or perhaps disclosure would be harmful to the party who possesses the information."* Following Smith and taking the mediator's trustworthiness for granted we showed that his services are beneficial.[27] Indeed, since the mediator has no stake in the game he has no incentive to diverge from the communication rule and sticking to it is incentive compatible. Yet, this is of course a rather limited treatment of the mediator as an economic agent and may become problematic if there exists, for example, a computational cost with following the optimal communication rule. More importantly, there exist collusive pressures once the mediator has obtained the agent's private information. In exchange for a small bribe the principal may ask the mediator to reveal more information than the communication rule prescribes.

Empirical observation indicates that the success of mediation depends indeed on the reputation and fairness of the mediator. In practice formal procedures of mediation are

---

[26]Bester and Strausz (2001) show that partial information revelation is a general feature of optimal mechanisms in contracting problems with imperfect commitment.

[27]See Brown and Ayres (1994, p.324–325) for additional references concerning the importance of private "caucusing" for the success of mediation.

structured to guarantee confidentiality. For instance, the formal mediation procedure of the London Court of International Arbitration (LCIA) states:

Article 10 Confidentiality and Privacy

10.1 All mediation sessions shall be private, and shall be attended only by the mediator, the parties and those individuals identified pursuant to Article 5.4.

10.2 The mediation process and all negotiations, and statements and documents prepared for the purposes of the mediation, shall be confidential and covered by "without prejudice" or negotiation privilege.

10.3 The mediation shall be confidential. Unless agreed among the parties, or required by law, neither the mediator nor the parties may disclose to any person any information regarding the mediation or any settlement terms, or the outcome of the mediation.

10.4 All documents or other information produced for or arising in relation to the mediation will be privileged and will not be admissible in evidence or otherwise discoverable in any litigation or arbitration in connection with the dispute referred to mediation, except for any documents or other information which would in any event be admissible or discoverable in any such litigation or arbitration.

10.5 There shall be no formal record or transcript of the mediation.

Yet, we want to emphasize that the mediator may be given strict incentives to follow the communication rule in a repeated version of our static model. More specifically, consider a dynamic model in which in each period a different principal and agent apply for the mediator's help and pay a fee for his services. In such a setting the recommendations are imperfect signals in the sense of Fudenberg, Levine, and Maskin (1994) about the mediator's action. We conjecture that a reputation equilibrium exists that sustains the truthful behavior of the mediator.

More general situations of conflict — By modeling the situation of conflict as a simple game of cheap talk with two types, we were able to derive the optimal mechanisms under mediation and non–mediation explicitly. In more complicated settings the analysis of optimal mechanisms becomes rather involved. Laffont and Tirole (1993, p.377) claim for instance that "*the lack of commitment in repeated adverse–selection situations leads to substantial difficulties for contract theory*". Yet, given our intuition we expect to obtain a similar beneficial role for a mediator for more complicated models of conflict. As soon as the optimal non–mediated contract involves partial information revelation, a mediator may improve outcomes, e.g. Bester and Strausz (2003). As shown in Bester and Strausz (2001) partial information revelation is a typical feature of mechanism design models with imperfect commitment (see also Hart and Tirole 1988 and Laffont and Tirole 1988,1990).

Mutually beneficial mediation — In this paper we assumed that the principal has all bargaining power. As a result she could dictate the use of a mediator without considering its effect on the agent. Indeed, comparing the respective equilibrium outcomes, the incompatible agent always prefers the equilibrium without mediator. Consequently, the beneficial effect of the mediator occurs partly at the expense of the agent. Although this setting may be applicable in some situations, in many settings the use of mediators requires the consent of both parties. A proper analysis of mutually beneficial mediation is however more complicated, since the agent's decision to accept or reject a mediator may be interpreted as a signal about the agent's type and therefore out–of–equilibrium beliefs will play a role.

To see this, suppose $y_1^a < y_1^p$ such that $\pi_1 = 0$. In this case, type 1 prefers the outcome with a mediator, while type 2 does not. Rejecting mediation may therefore be interpreted as revealing that the agent is of type 2, leading to a choice $y = y_2^p$ which type 2 finds worse than the mediation equilibrium outcome $y = y(\pi_2)$. Consequently, an equilibrium exists in which mediation occurs. This equilibrium depends on the out–of–equilibrium belief that the agent is of type 2 if mediation is rejected. Similarly, there exists an equilibrium in which mediation is rejected, which depends on the out–of–equilibrium belief that only the agent of type 2 accepts it.[28] This illustrates the additional problems that arise when the agent has a more active role than just sending messages.

Appendix

Proof of Lemma 1

The statement follows directly from

$$U_1(y_2) - U_1(y_1) = \int_{y_1}^{y_2} U_1'(y)dy < \int_{y_1}^{y_2} U_2'(y)dy = U_2(y_2) - U_2(y_1).$$

Proof of Lemma 2

First, note that $y(\rho)$ is implicitly defined by the first order condition

$$(1 - \rho)V_1'(y(\rho)) + \rho V_2'(y(\rho)) = 0. \tag{11}$$

Let $y < \min\{y_1^p, y_2^p\}$, then due to concavity of V_i it follows $V_i(y) < 0$. Therefore, there does not exist a $\rho \in [0, 1]$ such that (11) is satisfied for some $y(\rho) < \min\{y_1^p, y_2^p\}$. Likewise, let $y > \max\{y_1^p, y_2^p\}$, then $V_i(y) > 0$. Hence, there does not exist a $\rho \in [0, 1]$ such that (11) is satisfied for some $y(\rho) > \max\{y_1^p, y_2^p\}$.

---

[28]A refinement on out–of–equilibrium may select the equilibrium leading to mediation.

To prove the second statement differentiate (11) w.r.t. ρ to obtain $\partial y / \partial \rho = [V_1'(y) - V_2'(y)]/[(1-\rho)V_1''(y) + \rho V_2''(y)]$, where $y \in [\min\{y_1^p, y_2^p\}, \max\{y_1^p, y_2^p\}]$. Due to concavity of V_i the denominator is negative. If $y_1^p < y_2^p$ it follows that the numerator is negative as $V_1(y) \leq 0$ and $V_2(y) \geq 0$ with at least one strict inequality for all $\min y_1^p, y_2^p \leq y \leq \max y_1^p, y_2^p]$. Hence, $y'(\rho)$ is positive. If $y_1^p > y_2^p$, the numerator is positive and $y(\rho)$ is decreasing.

Proof of Proposition 1

Claim 2 trivial. To prove claim 1 we show that if $y_1^p > y_2^p$ then the optimal mechanism exhibits $y_1 = y_2$. First, we show that a direct mechanism with $y_1 > y_2$ is not incentive compatible. Obviously, at least one IC is binding at the optimum. If this is type 1 it follows from $0 = U_1(y_1) - U_1(y_2) = \int_{y_2}^{y_1} U_1'(y) \, dy < \int_{y_2}^{y_1} U_2'(y) \, dy = U_2(y_1) - U_2(y_2)$ that the mechanism is not incentive compatible for agent 2. Similarly, if agent 2's incentive constraint is binding, it follows from $0 = U_2(y_1) - U_2(y_2) = \int_{y_2}^{y_1} U_2'(y) \, dy > \int_{y_2}^{y_1} U_1'(y) \, dy = U_1(y_1) - U_1(y_2)$ that the mechanism is not incentive compatible for type 1.

Now suppose $y_1 < y_2$ and compare the principal's utility from this mechanism with that from the optimal pooling mechanism $\bar{y} = y(\pi) \in (y_2^p, y_1^p)$:

$$\Delta V \equiv (1-\pi)[V_1(y_1) - V_1(\bar{y})] + \pi[V_2(y_2) - V_2(\bar{y})] = \int_{\bar{y}}^{y_1} (1-\pi)V_1'(y) \, dy + \int_{\bar{y}}^{y_2} \pi V_2'(y) \, dy$$

$$< \int_{\bar{y}}^{y_1} (1-\pi)V_1'(\bar{y}) \, dy + \int_{\bar{y}}^{y_2} \pi V_2'(\bar{y}) \, dy < \int_{\bar{y}}^{y_1} [(1-\pi)V_1'(\bar{y}) + \pi V_2'(\bar{y})] \, dy = 0.$$

The first inequality follows from the concavity of V_i and the second because $V_2'(\bar{y}) < 0$ and $y_1 < y_2$ imply $\int_{y_2}^{y_1} V_2'(\bar{y}) dy > 0$. The final equality follows from the first order condition determining $y(\pi)$.

Proof of Lemma 3

Suppose agent 1 is incompatible and $y_1^p < y_2^p$, it then follows from

$$0 > U_1(y_2^p) - U_1(y_1^p) = \int_{y_1^p}^{y_2^p} u_1'(y) dy > \int_{y_1^p}^{y_2^p} u_2'(y) dy = U_2(y_2^p) - U_2(y_1^p)$$

that agent 2 is compatible.

Proof of Proposition 2

Note first that the optimal separation mechanism exhibits $y_1^* < y_2^*$. Next, note that at least one agent's incentive constraint must be binding at the optimum. Lemma 1 then implies that the other agent's IC has slack. Suppose agent i's IC binds and define the function $y_2^i(y)$ for the range $y_1^* < y_i^a$ implicitly by $U_i(y_1) = U_i(y_2^i(y))$. Hence, y_1^*

maximizes $(1-\pi)V_1(y)+\pi V_2(y_2^i(y))$ and satisfies the first order condition $(1-\pi)V_1'(y_1^*) = -\pi V_2'(y_2^i(y_1^*))\partial y_2^i(y_1^*)/\partial y$. Since $\partial y_2^i(y)/\partial y < 0$, it follows that y_1^* and y_2^* are such that $\text{Sign}(V_1'(y_1^*)) = \text{Sign}(V_2'(y_2^*))$. Consequently, either $y_1^* < y_1^p$ and $y_2^* < y_2^p$ or $y_1^* > y_1^p$ and $y_2^* > y_2^p$. But since agent 2 is incompatible while agent 1 is compatible, there does not exist a $y > y_1^p$ such that $y_2^i(y) > y_2^p$. We therefore conclude that the optimal separating mechanism exhibits $y_1^* < y_1^p$ and $y_2^* < y_2^p$.

To show that the incentive constraint of agent 2 is binding, suppose by contradiction that agent 1's IC binds, i.e. (1) is satisfied with equality. Lemma 1 implies that (2) is strictly satisfied. Now consider a small raise in y_2^* such that inequality (2) remains satisfied. As established, an optimal contract satisfies $y_1^* < y_1^p < y_2^* < y_2^p$, which implies $V_2'(y_2^*) > 0$ and $U_1'(y_2^*) < 0$, the raise in y_2^* therefore increases the principal's utility, while rendering (1) satisfied with strict inequality. Hence, a binding incentive constraint of agent 2 is not optimal.

Proof of Proposition 3

We first show non–existence of equilibria of the types 1, 3, and 5.

In a full revelation outcome necessarily $(\alpha_1, \alpha_2) = (1, 0)$. The principal's beliefs must therefore satisfy $\rho_1 = \rho(1, 0) = 0$ and $\rho_2 = \rho(0, 1) = 1$ and implement $y_1 = y(1) = y_1^p$ and $y_2 = y(0) = y_2^p$, which due to $U_2(y_2^p) < U_2(y_1^p)$ contradicts (3). A full revelation equilibrium does therefore not exist.

A partial revelation equilibrium does not exist, since in such an equilibrium $\alpha_i \in (0, 1)$ for both $i = 1, 2$ and $y_1 \neq y_2$. By (3) this would require $U_i(y_1) = U_i(y_2)$ for both $i = 1, 2$. Lemma 1 shows this is not possible.

Also a type 2 partially full revelation equilibrium does not exist. Such an equilibrium exhibits $\alpha_1 = 1$ and $0 < \alpha_2 < 1$. Consequently, $\rho_1 = \rho(1, \alpha_2) \in (0, 1)$ and $\rho_2 = \rho(0, 1 - \alpha_2) = 1$, which implies $y_1 \in (y_1^p, y_2^p)$ and $y_2 = y(0) = y_2^p$. Moreover, due to (3) $0 < \alpha_2 < 1$ requires $U_2(y_1) = U_2(y_2^p)$. However, such a $y_1 \in (y_1^p, y_2^p)$ does not exist, because $U_2(y_2^p) < U_2(y_1^p)$ and the concavity of U_2 implies $U_2(y_2^p) < U_2(y)$ for all $y \in [y_1^p, y_2^p)$.

Hence, only two equilibrium candidates are left. The non–revelation equilibrium, which always exists in the form $(\alpha_1, \alpha_2, \rho_1, \rho_2, y_1, y_2) = (\alpha, \alpha, \pi, \pi, y(\pi), y(\pi))$ with $\alpha \in (0, 1)$, and the partially full revelation equilibrium. Obviously, the latter yields the principal a higher payoff. However, a type 1 partially full revelation equilibrium exists if and only if $\pi < \pi_1$. This follows from the observation that in a type 1 partially full revelation equilibrium $0 < \alpha_1 < 1$ and $\alpha_2 = 0$. Consequently, $\rho_1 = \rho(\alpha_1, 0) = 0$ and $\rho_2 = \rho(1 - \alpha_1, 1) < 1$, which implies $y_1 = y(0) = y_1^p$ and $y_2 \in (y_1^p, y_2^p)$. Moreover, due to (3) $0 < \alpha_1 < 1$ requires $U_1(y_2) = U_1(y_1^p)$ and thus $y_2 = y(\pi_1)$. This in turn requires

$\rho_2 = \rho(1 - \alpha_1, 1) = \pi_1$ and hence, by (4), $\alpha_1 = (\pi_1 - \pi)/[(1 - \pi)\pi_1]$. However, α_1 must be non–negative and, in order to have some information revelation, must differ from $\alpha_2 = 0$. Therefore, an informative equilibrium requires $\pi < \pi_1$. For its existence it remains to be checked that $\alpha_2 = 0$ satisfies the incentive constraint (3) of agent 2; this follows from Lemma 1.

Proof of Lemma 5

If an incentive compatible communication rule $(\{1, 2\}, R, \alpha_1, \alpha_2)$ induces information revelation then there is some $r_j \in R$ with $\alpha_{1j} \neq \alpha_{2j}$. Since $\sum \alpha_{1j} = \sum \alpha_{2j} = 1$, there exist $r_j, r_k \in R$ with $\alpha_{1j} > \alpha_{2j}$ and $\alpha_{1k} < \alpha_{2k}$. Incentive compatibility of a recommendation r_j implies $\text{Sign}(\alpha_{2j} - \alpha_{1j}) = \text{Sign}(r_j - y(\pi))$, i.e. r_j and r_k satisfy $r_k > y(\pi) > r_j$.

If $\pi \geq \pi_2$ it follows $U_2(y(\pi)) \leq U_2(y_1^p)$. Consequently, due to the concavity of U_2, $\alpha_{2j} > \alpha_{1j}$ implies $U_2(r_j) < U_2(y(\pi))$, while $\alpha_{2j} < \alpha_{1j}$ implies $U_2(r_j) > U_2(y(\pi))$. It therefore follows that $\sum_j (\alpha_{2j} - \alpha_{1j})U_2(r_j) = \sum_j \max\{\alpha_{2j} - \alpha_{1j}, 0\}U_2(r_j) - \sum_j \max\{\alpha_{1j} - \alpha_{2j}, 0\}U_2(r_j) < \sum_j \max\{\alpha_{2j} - \alpha_{1j}, 0\}U_2(y(\pi)) - \sum_j \max\{\alpha_{1j} - \alpha_{2j}, 0\}U_2(y(\pi)) = 0$, contradicting the incentive compatibility condition (8). Therefore if an incentive compatible rule P induces information revelation, then necessarily $\pi < \pi_2$.

To prove the second statement note first that for $i = 1, 2$:

$$\sum_{j=1}^{|R|} (\alpha_{1j} - \alpha_{2j})U_i(r_j) = \sum_{j=1}^{|R|-1} \left(\sum_{k=1}^{j} \alpha_{1k} - \sum_{k=1}^{j} \alpha_{2k} \right) [U_i(r_j) - U_i(r_{j+1})].$$

Due to (5), one has $\sum_{k=1}^{j} \alpha_{1k} - \sum_{k=1}^{j} \alpha_{2k} > 0$ for all $j = 1, \ldots, |R| - 1$. Hence, by the monotonicity assumption and recalling that $r_j \leq r_{j+1}$,

$$\sum_{j=1}^{|R|} (\alpha_{1j} - \alpha_{2j})U_1(r_j) = -\sum_{j=1}^{|R|-1} \left(\sum_{k=1}^{j} \alpha_{1k} - \sum_{k=1}^{j} \alpha_{2k} \right) \int_{r_j}^{r_{j+1}} U_1'(y)\,dy$$

$$> -\sum_{j=1}^{|R|-1} \left(\sum_{k=1}^{j} \alpha_{1k} - \sum_{k=1}^{j} \alpha_{2k} \right) \int_{r_j}^{r_{j+1}} U_2'(y)\,dy = \sum_{j=1}^{|R|} (\alpha_{1j} - \alpha_{2j})U_2(r_j).$$

Therefore, if (7) holds with equality, (8) is satisfied with strict inequality and vice versa.

For the third statement, note that if $\pi < \pi_2$ then for all $r_l < y(\pi)$ it holds $U_2(r_l) \geq U_2(y(\pi))$. Now suppose that for all $r_k > y(\pi)$ it holds $r_k = y_2^p$, then $U_2(r_k) < U_2(y_1^p)$ for all k such $\alpha_{2k} > \alpha_{1k}$. It follows $\sum_j (\alpha_{2j} - \alpha_{1j})U_2(r_j) = \sum_j \max\{\alpha_{2j} - \alpha_{1j}, 0\}U_2(r_j) - \sum_j \max\{\alpha_{1j} - \alpha_{2j}, 0\}U_2(r_j) < \sum_j \max\{\alpha_{2j} - \alpha_{1j}, 0\}U_2(y(\pi)) - \sum_j \max\{\alpha_{1j} - \alpha_{2j}, 0\}U_2(y(\pi)) = 0$, contradicting the incentive compatibility condition (8). Therefore if an incentive compatible rule P induces information revelation, then there exists an r_j such that $y(\pi) < r_j < y_2^p$.

Proof of Lemma 6

Consider a communication rule P with some recommendation r_k such that $r_k \in (y_1^p, y_2^p)$. Define, for $\delta > 0$ but small, the proposal $P^{kl}(\delta)$ as the following transformation of P:

$$
\begin{aligned}
\alpha_{1k}(\delta) &\equiv \alpha_{1k} - \delta\alpha_{1l} & \alpha_{2k}(\delta) &\equiv \alpha_{2k} - \delta\alpha_{2l} & r_k(\delta) &\equiv y(\rho(\alpha_{1k}(\delta), \alpha_{2k}(\delta))) \\
\alpha_{1l}(\delta) &\equiv \alpha_{1l} + \delta\alpha_{1l} & \alpha_{2l}(\delta) &\equiv \alpha_{2l} + \delta\alpha_{2l} & r_l(\delta) &= r_l \\
\alpha_{1j}(\delta) &\equiv \alpha_{1j} & \alpha_{2j}(\delta) &\equiv \alpha_{2j} & r_j(\delta) &= r_j \quad \text{for } j \neq k, l.^{29}
\end{aligned}
$$

The transformation is structured in such a way that if the recommendations of the original proposal P are incentive compatible then this also holds for the recommendations in $P^{kl}(\delta)$. Moreover, if δ rises, the pair $(r_k(\delta), r_l)$ becomes more informative, since $|r_k(\delta) - r_l|$ is increasing in δ due to $\text{Sign}(\partial r_k / \partial \delta) = \text{Sign}(\alpha_{1l}\alpha_{2k} - \alpha_{1k}\alpha_{2l})$. It suffices to show that $dV(P^{kl})/d\delta > 0$. To see this note first that, since rule P is incentive compatible it holds for any $r_j \in R$

$$
(1 - \rho_j)V_1'(r_j) + \rho_j V_2'(r_j) \quad = \quad 0, \tag{12}
$$

where $\rho_j = \rho(\alpha_{1j}, \alpha_{2j})$. Moreover the concavity of V_1 and V_2 imply that

$$
\text{Sign}(r_j - y) = \text{Sign}((1 - \rho_j)V_1'(y) + \rho_j V_2'(y)). \tag{13}
$$

Using (12) one obtains

$$
\begin{aligned}
\frac{dV(P^{kl}(\delta))}{d\delta} &= (1 - \pi)\alpha_{1l}[V_1(r_l) - V_1(r_k)] + \pi\alpha_{2l}[V_2(r_l) - V_2(r_k)] \\
&= (\pi\alpha_{2l} + (1 - \pi)\alpha_{1l}) \int_{r_k}^{r_l} (1 - \rho_l)V_1'(y) + \rho_l V_2'(y) \, dy > 0
\end{aligned}
$$

where the sign follows from (13).

Proof of Proposition 4

Since agent 2 is incompatible, full revelation is not possible so that at least one incentive constraint must be binding. By Lemma 1 at most one incentive constraint is binding when some information revelation occurs. Since $U_i(y_1^p) > U_i(y_2^p)$ for both i, a binding incentive constraint implies that in an equilibrium with some information revelation $r_1 < r_2 < y_2^p$ and thus $\alpha_{21} < \alpha_{11} < 1$.

Now suppose an incentive compatible communication rule P is such that agent 1's incentive constraint is binding, i.e. $U_1(r_1) = U_1(r_2)$ and $U_1'(r_2) < 0$. Feasibility requires moreover $y_1^a > y_1^p$. Consider the communication rule $P^{21}(\delta)$ as defined in the proof of Lemma 6. Note that for $\delta > 0$ small enough the communication rule $P^{21}(\delta)$ is feasible and prescribes the incentive compatible recommendations $r_1(\delta) = r_1$ and $r_2(\delta) > r_2$. We now

show that the communication rule $P^{21}(\delta)$ with $\delta > 0$ remains incentive compatible with respect to both agents. Recall that $P^{21}(0) = P$ is such that the incentive compatibility constraint of agent 2 is slack. Due to continuity the constraint remains slack for a communication rule $P^{21}(\delta)$ with $\delta > 0$ small enough. Note furthermore that for $\delta > 0$ one has $r_2(\delta) > r_2$, and since $U_1'(r_2) < 0$, one obtains $U_1(r_1) = U_1(r_2) > U_1(r_2(\delta))$. Therefore, if P is incentive compatible, then also $P^{21}(\delta)$ for $\delta > 0$ small enough. Since by Lemma 6 the principal's utility increases with $\delta > 0$, a communication rule $P = P^{21}(0)$ for which $U_1(r_1) = U_1(r_2)$ cannot be optimal.

Now suppose P is such that the incentive constraint of agent 2 is binding and $\alpha_{21} > 0$. Consider the communication rule $P^{12}(\delta)$ with $\delta > 0$. Using the same argument as above, $P^{12}(\delta)$ remains incentive compatible for $\delta > 0$ small enough and increases the principal's utility. Hence, a communication rule P with a binding incentive constraint of agent 2 and $\alpha_{21} > 0$ cannot be optimal. We therefore conclude that the optimal 2–proposal rule P is characterized by a binding incentive constraint of agent 2 and $\alpha_{21} = 0$. This implies $(y_1, y_2) = (y_1^p, y(\pi_2))$ which in turn implies $\rho(1 - \alpha_{11}, 1) = \pi_2$ and hence, by (4), $\alpha_{11} = (\pi_2 - \pi)/[(1 - \pi)\pi_2]$. However, α_{11} must be non–negative and, in order to have some information revelation, must differ from $\alpha_{21} = 0$. Therefore, an informative equilibrium requires $\pi < \pi_2$, which obviously requires $\pi_2 > 0$.

Proof of Theorem 1

We prove that for $y_1^p > y_2^p$ no information revelation can occur in equilibrium. The Theorem then follows directly from Lemma 5 and Proposition 4. Let $y_1^p > y_2^p$ and suppose an incentive compatible proposal rule exists that induces some information revelation, i.e. P^* is such that $\alpha_{1j} > \alpha_{2j}$ for at least some $j = 1, \ldots, |R|$. Due to (5), one has $\sum_{k=1}^{j} \alpha_{1k} - \sum_{k=1}^{j} \alpha_{2k} \geq 0$ for all $j = 1, \ldots, |R| - 1$ with at least one strict inequality. Due to Lemma 2 incentive compatibility of the recommendations require $r_j \geq r_{j+1}$ with at least one strict inequality. Hence, by using (12) and the monotonicity assumption, it follows

$$
\sum_{j=1}^{|R|} (\alpha_{1j} - \alpha_{2j}) U_1(r_j) = \sum_{j=1}^{|R|-1} \left(\sum_{k=1}^{j} \alpha_{1k} - \sum_{k=1}^{j} \alpha_{2k} \right) \int_{r_{j+1}}^{r_j} U_1'(y)\, dy
$$

$$
< \sum_{j=1}^{|R|-1} \left(\sum_{k=1}^{j} \alpha_{1k} - \sum_{k=1}^{j} \alpha_{2k} \right) \int_{r_{j+1}}^{r_j} U_2'(y)\, dy = \sum_{j=1}^{|R|} (\alpha_{1j} - \alpha_{2j}) U_2(r_j). \tag{14}
$$

From (14) it follows that if a proposal rule is incentive compatible w.r.t. type 1 it is not incentive compatible w.r.t. type 2 and vice versa. Therefore an incentive compatible proposal rule that induces some information revelation does not exist.

Proof of Proposition 5

Assume $\pi < \pi_2$, i.e. $y_2^a > y_1^p$. It follows from Theorem 1 that the optimal communication rule induces some amount of information revelation. Since full revelation is not possible some, and by Lemma 5 one, incentive constraint is binding at the optimum.

For some $r_k, r_l \in R$ with $r_k \in (y_1^p, y_2^p)$ consider the transformation P^{kl} as defined in the proof of Lemma 6. In order to evaluate the impact of the transformation on the incentive constraints, define the functions

$$
\begin{aligned}
f_i^{kl}(\delta) &\equiv \sum_{j=1}^{|R|}[\alpha_{1j}(\delta) - \alpha_{2j}(\delta)]U_i(r_j(\delta)) - \sum_{j=1}^{|R|}[\alpha_{1j} - \alpha_{2j}]U_i(r_j) \\
&= \delta(\alpha_{1l} - \alpha_{2l})[U_i(r_l) - U_i(r_k(\delta))] + (\alpha_{1k} - \alpha_{2k})[U_i(r_k(\delta)) - U_i(r_k)].
\end{aligned}
\tag{15}
$$

Now consider the derivative of $f_i^{kl}(\delta)$ evaluated at $\delta = 0$:

$$
\frac{df_i^{kl}(0)}{d\delta} = (\alpha_{1l} - \alpha_{2l})[U_i(r_l) - U_i(r_k)] + (\alpha_{1k} - \alpha_{2k})U_i'(r_k)r_k'(0),
\tag{16}
$$

where

$$
\text{Sign}\,(r_k'(0)) = \text{Sign}(\alpha_{1l}\alpha_{2k} - \alpha_{1k}\alpha_{2l}) = \text{Sign}(r_k - r_l).
\tag{17}
$$

For statement (i), suppose by contradiction that P is such that agent 1's IC binds which is only feasible if $y_1^a > y_1^p$. An incentive compatible communication rule P that induces information revelation contains a pair (r_k, r_l) such that $y_2^p > r_k > y(\pi) > r_l$, by statement 3 of Lemma 5. Incentive compatibility implies $\alpha_{1k} < \alpha_{2k}$, $\alpha_{1l} > \alpha_{2l}$, while (17) implies $r_k'(0) > 0$.

If $U_1(r_k) \leq U_1(r_l)$, it follows, due to $r_k > r_l$, that $r_k > y_1^a$. Consequently, $U_1'(r_k) < 0$. These properties imply $df_1^{kl}(0)/d\delta > 0$, which means that agent 1's IC (7) remains satisfied for small $\delta > 0$. Agent 2's IC is also satisfied since, for $\delta = 0$, it has slack. Hence, there exists a $\delta > 0$ for which the transformation $P^{kl}(\delta)$ is feasible. By Lemma 6 such a P is not optimal.

Now consider the case $U_1(r_k) > U_1(r_l)$, which due to $r_k > r_l$ implies $r_l < y_1^a$. If there exists an $\varepsilon > 0$ such that for all $\delta \in (0, \varepsilon)$ it holds that $f_1^{kl}(\delta) \geq 0$, then P is not optimal by the above argument. If such an ε does not exist, then $f_1^{kl}(\delta) < 0$ for δ sufficiently close to zero. But then there exists a $\hat{\delta} > 0$ such that $f_1^{kl}(\hat{\delta}) = 0$. This follows from continuity of $f_1^{kl}(\delta)$ and the fact that there exists a $\delta' > 0$ such that $U_1(r_k(\delta')) = U_1(r_l)$, which implies $f_1^{kl}(\delta') > 0$. The transformation $P^{kl}(\hat{\delta})$ is feasible because, due to $f_1^{kl}(\hat{\delta}) = 0$, agent 1's IC still holds in equality, which, by Lemma 5, implies that also agent 2's IC is satisfied. Since $\hat{\delta} > 0$, Lemma 6 implies that P is not optimal. It therefore cannot be optimal to have agent 1's IC bind. Hence, at the optimum agent 2's IC binds.

For the remainder of the proof consider an incentive compatible communication rule P that induces some revelation of information and for which agent 2's IC binds. For statement (ii), let now $r_k < y(\pi) < r_l$, which implies $\alpha_{1k} > \alpha_{2k}$, $\alpha_{1l} < \alpha_{2l}$, and, by (17), $r'_k(0) < 0$. Assume by contradiction that $r_k > y_1^p$.

If $U_2(r_k) \leq U_2(r_l)$, it follows, due to $r_k < r_l$, that $r_k < y_2^a$. Consequently, $U'_2(r_k) > 0$. These properties imply $df_2^{kl}(0)/d\delta < 0$, which means that agent 2's IC (8) is satisfied for small $\delta > 0$. For $\delta > 0$ but small also agent 1's IC remains satisfied, since, for $\delta = 0$, it has slack. Hence, there exists a $\delta > 0$ for which the transformation $P^{kl}(\delta)$ is feasible. Lemma 6 implies that P is not optimal.

Now consider the case $U_2(r_k) > U_2(r_l)$, which due to $r_k < r_l$ implies $r_l > y_2^a$. If there exists an $\varepsilon > 0$ such that for all $\delta \in (0, \varepsilon)$ it holds that $f_2^{kl}(\delta) \leq 0$, then P is not optimal by the above argument. If such an ε does not exist, then $f_2^{kl}(\delta) > 0$ for δ sufficiently close to zero. But then there exists a $\hat{\delta} > 0$ such that $f_2^{kl}(\hat{\delta}) = 0$. This follows from continuity of $f_2^{kl}(\delta)$ and the fact that there exists a $\delta' > 0$ such that $U_2(r_k(\delta')) = U_2(r_l)$, which implies $f_2^{kl}(\delta') < 0$. The transformation $P^{kl}(\hat{\delta})$ is feasible because agent 2's IC still holds in equality which, by Lemma 5, implies that agent 1's still has slack. Since $\hat{\delta} > 0$, Lemma 6 implies that P is not optimal. Thus it is never optimal to have an $r_k \in (y_1^p, y(\pi))$. Since there must be an $r_k < y(\pi)$, we conclude that an optimal incentive compatible recommendation rule exhibits $r_1 = y_1^p$ and $r_j \geq y(\pi)$ for all $j > 1$.

For statement (iii), consider an optimal P, i.e. P exhibits $r_1 = y_1^p$ and $r_j \geq y(\pi)$ for all $j > 1$. Now consider the transformation $P^{k1}(\delta)$ with $k > 1$. Since $\alpha_{21} = 0$, it follows

$$\frac{df_2^{k1}(0)}{d\delta} \;=\; \alpha_{11}[U_2(y_1^p) - U_2(r_k)] + (\alpha_{1k} - \alpha_{2k})U'_2(r_k)r'_k(0), \qquad (18)$$

and since $r_k > r_1 = y_1^p$ it holds $r'_k(0) > 0$. Now if $y(\pi) \leq y_2^a$ and P contains an $r_k \in [y(\pi), y_2^a]$, then $U'_2(r_k) \geq 0$ and $\alpha_{1k} \leq \alpha_{2k}$. Moreover, $U_2(y_1^p) < U_2(r_k)$ and it follows that (18) is negative. Hence, for some small $\delta > 0$ the transformation $P^{k1}(\delta)$ is feasible and yields the principal more. Consequently, P is not optimal. If $y_2^a < y(\pi)$ and P contains an $r_k = y(\pi)$, then $\alpha_{1k} = \alpha_{2k}$. Since $\pi < \pi_2$ implies $U_2(y(\pi)) > U_2(y_1^p)$, it follows that (18) is negative and P is suboptimal.

Proof of Proposition 6

We must prove that an incentive compatible communication rule with $|R| > 2$ and $r_j \neq r_k$ for all $k \neq j$ cannot be optimal. Suppose by contradiction that such a P is optimal, then by Proposition 5 it satisfies $r_1 = y_1^p$ and $\max\{y(\pi), y_2^a\} < r_2 < r_3 \leq y_2^p$. Moreover, since agent 2's incentive constraint binds at the optimum, it holds that $r_2 < y(\pi_2)$ and there must also exist an $r_3 > y(\pi_2)$. This implies $U_2(r_2) > U_2(y_1^p) > U_2(r_3)$.

Denote by $P(\delta_2, \delta_3)$ the proposal rule which results from a joint transformation $P^{21}(\delta_2)$, $P^{31}(\delta_3)$, as defined in the proof of Lemma 6. Write as $V(\delta_2, \delta_3)$ the principal's payoff associated with $P(\delta_2, \delta_3)$. Its partial derivative with respect to δ_j evaluated at $\delta_2 = \delta_3 = 0$ is

$$\frac{\partial V(0,0)}{\partial \delta_j} = (1 - \pi)\alpha_{11}[V_1(y_1^p) - V_1(r_j)].$$

The principal's marginal gain from a joint transformation with $\delta_3 = \delta_3(\delta_2) = -\beta\delta_2$ with $\beta > 0$ is therefore

$$
\begin{aligned}
\left.\frac{dV(\delta_2, \delta_3(\delta_2))}{d\delta_2}\right|_{\delta_2=0} &= \frac{\partial V(0,0)}{\partial \delta_2} - \frac{\partial V(0,0)}{\partial \delta_3}\beta \\
&= (1-\pi)\alpha_{11}[V_1(y_1^p) - V_1(r_2) - \beta(V_1(y_1^p) - V_1(r_3))]. \qquad (19)
\end{aligned}
$$

To evaluate the impact of a marginal change of $P(0,0)$ on the incentive constraint of agent 2 define

$$F(\delta_2, \delta_3) = f_2^{21}(\delta_2) + f_2^{31}(\delta_3).$$

Recalling (18) and using $\delta_3(\delta_2) = -\beta\delta_2$ the total derivative of $F(\delta_2, \delta_3(\delta_2))$ evaluated at $\delta_2 = 0$ is:

$$
\begin{aligned}
\frac{dF(0,0)}{d\delta_2} &= \frac{\partial F(0,0)}{\partial \delta_2} - \frac{\partial F(0,0)}{\partial \delta_3}\beta \\
&= \alpha_{11}[U_2(y_1^p) - U_2(r_2)] - (\alpha_{22} - \alpha_{12})U_2'(r_2)r_2'(0) \\
&\quad - \beta\left(\alpha_{11}[U_2(y_1^p) - U_2(r_3)] - (\alpha_{23} - \alpha_{13})U_2'(r_3)r_3'(0)\right) \\
&= \alpha_{11}\left\{[U_2(y_1^p) - U_2(r_2)] - U_2'(r_2)y'(\rho_2)\rho_2\frac{\rho_2 - \pi}{\pi}\right. \\
&\quad \left. - \beta\left([U_2(y_1^p) - U_2(r_3)] - U_2'(r_3)y'(\rho_3)\rho_3\frac{\rho_3 - \pi}{\pi}\right)\right\} \qquad (20)
\end{aligned}
$$

where the last equation follows from $r_j'(0) = y'(\rho_j)\rho_j'(0)$ and $\rho_j'(0)(\alpha_{2j} - \alpha_{1j})/\alpha_{11} = \rho_j(\rho_j - \pi)/\pi$ with

$$\rho_j(\delta_j) \equiv \rho(\alpha_{1j}(\delta_j), \alpha_{2j}(\delta_j)).$$

A marginal change of P satisfies incentive compatibility of agent 2 if (20) is negative, i.e. if

$$\beta \geq \frac{[U_2(y_1^p) - U_2(r_2)] - U_2'(r_2)y'(\rho_2)\rho_2\frac{\rho_2 - \pi}{\pi}}{[U_2(y_1^p) - U_2(r_3)] - U_2'(r_3)y'(\rho_3)\rho_3\frac{\rho_3 - \pi}{\pi}}.$$

By (19) a marginal change of P (weakly) increases the principal's utility if

$$\beta \leq \frac{V_1(y_1^p) - V_1(r_2)}{V_1(y_1^p) - V_1(r_3)}.$$

Thus, a marginal change of P which (weakly) increases the principal's utility while leaving the rule incentive compatible exists, if

$$\frac{V_1(y_1^p) - V_1(r_2)}{V_1(y_1^p) - V_1(r_3)} \geq \frac{[U_2(y_1^p) - U_2(r_2)] - U_2'(r_2)y'(\rho_2)\rho_2 \frac{\rho_2 - \pi}{\pi}}{[U_2(y_1^p) - U_2(r_3)] - U_2'(r_3)y'(\rho_3)\rho_3 \frac{\rho_3 - \pi}{\pi}} \tag{21}$$

where $r_j = y(\rho_j)$. Since $U_2(r_2) > U_2(y_1^p) > U_2(r_3)$ and $U'(r_j) < 0 < y'(\rho_j)$ and $\rho_3 > \rho_2 > \pi$, condition (21) is strictly satisfied for any π if and only if

$$D(\rho_2) \leq D(\rho_3) \tag{22}$$

with

$$D(\rho) \equiv \frac{V_1(y_1^p) - V_1(y(\rho))}{U_2'(y(\rho))y'(\rho)\rho^2}. \tag{23}$$

Therefore, if $D(\rho)$ is weakly increasing a proposal with $|R| > 2$ is not optimal. Straightforward calculations yield that the derivative of $D(\rho)$ is larger or equal to zero when

$$\frac{U_2''(y(\rho))}{U_2'(y(\rho))} + \frac{2y'(\rho) + \rho y''(\rho)}{\rho y'(\rho)^2} \geq -\frac{V_1'(y(\rho))}{\Delta V(y(\rho))} \tag{24}$$

with

$$\Delta V(y) \equiv V_1(y_1^p) - V_1(y).$$

The definition of $y(\rho)$ implies $2y'(\rho) + \rho y''(\rho) = (2V_1''(y(\rho)) - \rho y'(\rho)b(y(\rho)))y'(\rho)/a(y(\rho))$ and $\rho y'(\rho) = V_1'(y(\rho))/a(y(\rho))$ with

$$
\begin{aligned}
a(y) &\equiv (1 - \rho)V_1''(y) + \rho V_2''(y) < 0, \\
b(y) &\equiv (1 - \rho)V_1'''(y) + \rho V_2'''(y).
\end{aligned}
$$

Hence, dropping the dependence on ρ we may rewrite (24) as

$$\frac{U_2''(y)}{U_2'(y)} + \frac{2V_1''(y)}{V_1'(y)} - \frac{b(y)}{a(y)} \geq -\frac{V_1'(y)}{\Delta V(y)}. \tag{25}$$

Now if $V_1'''(y) \geq 0$ and $V_2'''(y) \geq 0$ then $b(y) \geq 0$ and (25) is satisfied if

$$U_2''(y)/U_2'(y) + 2V_1''(y)/V_1'(y) \geq -V_1'(y)/\Delta V(y).$$

References

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

Baker, G., R. Gibbons, and K. Murphy (1999), "Informal Authority in Organizations", *Journal of Law, Economics, and Organization 15*, 56-73.

Bester, H. and R. Strausz (2001), "Imperfect Commitment and the Revelation Principle: the Single Agent Case", *Econometrica* 69, 1077-1098

Bester, H. and R. Strausz (2003), "Contracting with Imperfect Commitment and Noisy Communication", mimeo Free University of Berlin.

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

Crawford, V. and J. Sobel (1982), "Strategic Information Transmission", *Econometrica 50*, 1431-51.

Dewatripont, M. (1986), *On the Theory of commitment, with applications to the labor market*, Ph.D. dissertation, Harvard University.

Forges, F. (1986), "An Approach to Communication Equilibria", *Econometrica 54*, 1375-85.

Forges, F. (1988), "Can Sunspots Replace a Mediator?", *Journal of Mathematical Economics 17*, 347-68.

Forges, F. (1990), "Equilibria with Communication in a Job Market Example", *Quarterly Journal of Economics 105*, 375-98.

Fudenberg, D., D. Levine, and E. Maskin (1994), "The Folk Theorem with Imperfect Public Information", *Econometrica*, 997-1039.

Hart, O. and J. Tirole (1988), "Contract renegotiation and Coasian dynamics", *Review of Economic Studies 55*, 509-40.

Krishna, V., and Morgan, J. (2004), "The Art of Conversation", *forthcoming in Journal of Economic Theory*.

Laffont, J.J. and J. Tirole (1988), "The Dynamics of Incentive Contracts", *Econometrica 56*, 1153-1175.

Laffont, J.J. and J. Tirole (1990), "Adverse selection and renegotiation in procurement", *Review of Economic Studies 57*, 597-625.

Laffont, J.J. and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*, MIT Press.

Matthews, S. and A. Postlewaite (1989), "Pre-play Communication in Two-Person Sealed-Bid Double Auctions", *Journal of Economic Theory; 48*, 238-63.

Mitusch, K. and R. Strausz (1999), "Mediators and Mechanism Design: Why Firms Hire Consultants", DP 1999/15, FU-Berlin.

Myerson, R. (1985), "Bayesian equilibrium and incentive compatibility: An introduction" in: L. Hurwicz, D. Schmeidler and H. Sonnenschein, eds., *Social goals and social organization* (Cambridge University Press, Cambridge) 229-259.

Myerson, R. (1986), "Multistage Games with Communication", *Econometrica 54*, 323-58.

Smith, R. (1995), *ADR and Financial Institutions* (West Group) (see also http://www.robertmsmith.com/mediation.html).