On the benefits of mediation in contracting problems with limited commitment

Martin Pollrich*

- Preliminary and Incomplete Version -

April 12, 2013

Abstract

This paper demonstrates that the restriction to one-shot communication mechanisms in environments where the mechanism designer cannot fully commit to the outcome induced by the mechanism is overly restrictive. I show that an indirect communication mechanism yields strictly larger payoffs, compared to the best one-shot communication mechanism. One-shot direct mechanisms imply randomization by some agents type, which can be better controlled using a mediated mechanism. The latter allows for breaking indifference conditions and increases the flexibility of the mechanism designer. These results suggest a re-examination of existing literature on related problems, where the restriction to one-shot face-to-face communication is made without justification by a revelation principle.

JEL classification: D82, D86, D02, C72

Keywords: Contract Theory; Imperfect Commitment; Communication; Optimal Contracts

*Humboldt-University at Berlin, *martin.pollrich@hu-berlin.de*
The author is grateful to Roland Strausz, Helmut Bester and Matthias Lang for helpful advice and suggestions, and to seminar participants at the Free University Berlin and the Humboldt-University Berlin for helpful comments. Financial support by Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

1 Introduction

This paper analyzes the benefits of indirect communication in contracting problems with adverse selection and limited commitment. I consider a principal-agent framework in which the principal is imperfectly informed about the agents type. Furthermore the principal may use costly audits to learn the agents type ex-post and impose punishments on the on the agent. However the principal cannot contractually commit to performing audits. Whenever the optimal contract under the restriction of using only *one-shot communication mechanisms* induces audits with strictly positive probability, I show that there exists an *indirect communication mechanism* that yields strictly larger payoffs to the principal.

The usual focus in contract theory lies on one-shot communication mechanisms. This is typically justified by the revelation principle, which states that the range of implementable outcomes is simply the set of outcomes that give no incentive to the agent to misreport his type. The revelation principle however fails, when the principal cannot credibly commit to any outcome of the mechanism. In particular, it fails whenever the principal, as the designer of the mechanism, must take an action that is unverifiable so that it cannot be part of the mechanism.[1]

Applying the revelation principle to a contracting problem with adverse selection allows to simplify the communication game played between the principal and the agent along several dimensions. First of all no complex communication protocols, such as multi-stage communication or indirect communication, need to be implemented. Second, it is sufficient to have the agent send messages that belong to the set of ex-ante types. Third, it is without loss of generality to look at mechanisms, where the agent is induced to reveal his type truthfully.

With limited commitment and hence the revelation principle failing, the picture changes drastically. Bester and Strausz (2001) show that when restricting to one-shot communication mechanisms, it is still without loss of generality to restrict the message space to the set of types. However, honest revelation of the agents type may not be desirable anymore, in particular agents may randomize between revealing their true type and reporting to be of different types. Addressing the issue of communication in general, Bester and Strausz (2007) consider noisy communication between the principal and the agent. With the help of a numerical example they demonstrate that their general communication device may lead to strictly larger payoffs than the best one-shot communication mechanism.

General communication devices are well known from game theory, e.g. Myerson (1982) and Forges (1986). A particular example to see how communication affects implementable

[1] There may be other reasons for the revelation principle to fail, e.g. limited commitment in long-term relationships, renegotiation of ex-post inefficiencies or several mechanism designers competing against each other. These issues are not addressed in this paper.

outcomes can be found in the literature on cheap talk. Optimal one-shot face-to-face communication is studied in Crawford and Sobel (1982). Multiple rounds of communication can lead to Pareto improvements, as was shown in Krishna and Morgan (2004). More recently, Goltsman et al. (2009) demonstrate that impartial mediation leads to further Pareto improvements.[2]

In the domain of contract theory no sharp boundary between the different modes of communication can be drawn. In fact, there is not much work going beyond one-shot face-to-face communication. One exception, noted already before, is Bester and Strausz (2007) who look at general communication devices in an abstract framework. However, their focus lies on the analytical benefits, rather than a comparison of achievable payoffs.

The framework that is used throughout in this paper is a simplified version of the model of regulating a monopoly introduced by Baron and Besanko (1984). Using the neutral language of principal and agent, in this paper the agent is better informed about his cost of production than the principal. However, the latter has the possibility to verify the agent's realized cost ex-post and may impose penalty payments, depending on the audit findings. The authors investigate optimal audit contracts under the assumption of full commitment, that is the principal can commit to transfers, quantities, penalty schemes and audit probabilities. Optimal contracts in their setting are incentive compatible and typically entail audits with strictly positive probabilities. This raises the commitment issue, as the costly audit never leads to collection of a penalty and therefore is not optimal from an ex-post perspective. This point is taken on in Khalil (1997), who considers a simplified two-type model of the Baron & Besanko framework where the assumption of full commitment is relaxed. The principal now cannot commit to costly audits. Methodologically, Khalil restricts the analysis to direct, one-shot communication mechanisms, that is the agent can choose to report one of the two available types but do so non-truthfully. The optimal contract may now entail lying by the efficient type and random audits by the principal whenever the agent reports to be inefficient.

In a first step I justify the optimality of the Khalil contract by applying the results of Bester and Strausz (2001). The main part of the paper then lies in demonstrating that whenever the optimal contract induces audits with strictly positive probability, there exists a mediated mechanism that leads to strictly larger payoffs to the principal. In this sense, the restriction to one-shot face-to-face communication mechanisms is not without loss of generality.

When the optimal one-shot mechanism induces audits the agent's efficient type uses a mixed strategy. Essentially the contract leads to an inspection game, played between the principal and the agent. With some probability the agent misreports his type and the principal

[2] Similar comparisons can be made in multiplayer games. Add some literature here!

audits with positive probability after the agent reported to be inefficient. In this game, the randomization of one player is chosen such, that it makes the other just indifferent between her available actions. Using mediated communication allows for breaking these indifference condition for the agent. The randomization of reports is now done by the mediator and not by the agent himself. Instead, the agent needs only be kept indifferent between two particular randomizations, but not between the single elements over which the randomization is conducted. This generates a new degree of freedom for the principal in designing the optimal mechanism. With one-shot face-to-face communication a particular pair of transfer-quantity bundles can only be implemented in a unique way as a mixed equilibrium, whereas with mediation there are several ways to achieve this.

This result is likely to hold more generally. As Bester and Strausz (2001) have shown, the optimal one-shot direct mechanism typically entails randomization by the agent. This requires the agent to be indifferent between the respective options. With mediation however this indifference can be broken which increases the principal's flexibility in designing an optimal mechanism.

The remainder of this paper is organized as follows: Section 2 presents the model and reviews some of the results of the full commitment case. The following section 3 considers the contracting problem making use of only one-shot communication mechanisms. Section 4 contrasts this with mediated contracts and presents the major findings of this paper. Section 5 concludes.

2 Model

Consider the following principal-agent framework: A principal hires an agent to carry out the production of some good. Upon producing a total quantity $q \in [0, \infty)$, the agent incurs a cost $c(q) = \theta q$, where θ is an efficiency parameter privately known to the agent. We assume $\theta \in \Theta := \{\theta_l, \theta_h\}$ with $0 < \theta_l < \theta_h$. The principal does not observe θ, she assesses prior probability π on the cost parameter being θ_l. A quantity q of the good is worth $V(q)$ to the principal, where $V(0) = 0$ and to ensure strictly positive bounded output levels $V'(0) = \infty$ as well as $V'(\infty) = 0$. Efficient output levels are hence given by

$$V'(q_i^*) = \theta_i, \quad i = l, h. \tag{1}$$

As output is publicly observable the principal may write contracts conditioning on output. In addition we assume the principal possesses an audit technology, which allows to perfectly learn the agents cost after production took place. For instance the principal may check the accounts of the agent. Upon performing an audit, the principal incurs a cost of $c > 0$.[3] The audit reveals the cost the agent incurred from producing the specified output q and thus the agents true cost parameter θ. Based on the result of the audit, the principal may impose an extra payment on the agent, whose extent is exogenously fixed at the level $P > c$.[4] This payment can be interpreted as a fine that has to be paid to the principal under certain conditions and may be enforced by a third party. Therefore its extent cannot be altered by the principal but is always set by this third party when the conditions for demanding the payment are met. I assume P to be finite.[5] The principal cannot commit to performing an audit, but can commit to circumstances that may lead to a fine payment. Thus the contractual variables are the quantity q, the transfer t and a function $\mathcal{P}(\cdot)$ specifying which audit result may lead to the payment of P.

3 One-Shot Communication & Direct Mechanisms

In this chapter we follow the standard approach to tackle a contracting problem of the described kind. We assume that the principal can set up a one-shot communication mechanism, that is she proposes a message set M, from which the agent has to select some element $m \in M$ to send to the principal. M can be an arbitrary set, for analytical tractability let M be a metric space endowed with the Borel σ-algebra $\mathcal{M}$. With the message set M the principal commits to a decision function $d : M \to \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{P}$, where $\mathbb{P} = \{\mathcal{P} | \mathcal{P} : \Theta \to \{0, P\}\}$. A typical function value $d(m)$ consists of a transfer $t(m)$, a quantity $q(m)$ and a penalty function $\mathcal{P}(\cdot|m)$. We use the following interpretation of this structure: the principal has committed herself to the decision function $d(\cdot)$ and the agent can enforce decision $d(m)$ by sending message m. A *one-shot mechanism* or contract $\Gamma = \langle M, d \rangle$ specifies a message set M in combination with a decision function $d(\cdot)$.

A one-shot mechanism Γ induces the following game between the principal and the agent: First, the agent chooses some message $m \in M$. The agents choice determines the specified decision $d(m)$, i.e. a transfer $t(m)$, a quantity $q(m)$ and a penalty function $\mathcal{P}(\cdot|m)$. Furthermore the principal uses the agent's message to update her belief about the agent's type and she chooses whether to perform an audit. Given Γ this constitutes a game between the principal and the agent for which we shall use Perfect Bayesian Equilibrium as the relevant solution concept. In designing this game optimally the principal is therefore constrained to the PBE of the game induced by Γ.

[3] With $c = 0$ the problem is trivial, since the commitment issue becomes irrelevant.

[4] Reference to Matthias...

[5] It is well known that potentially infinite fine payments lead to first best outcomes, see for instance Nalebuff and Scharfstein (1987).

The tedious task of finding the optimal one-shot mechanism is substantially simplified by applying the revelation principle as introduced in Bester and Strausz (2001): The payoff of any optimal contract can also be achieved by a direct mechanism where each agents type uses the respective message with strictly positive probability. In other words we can restrict the analysis of finding the optimal one-shot communication mechanism to the message set $M = \Theta$ and PBE where each type θ_i sends message θ_i with strictly positive probability. This allows us to state the principal's problem of finding the optimal one-shot communication mechanism as follows: She chooses $\{q_h, q_l, t_h, t_l, P_{hh}, P_{hl}, P_{ll}, P_{lh}, \alpha_h, \alpha_l, \rho_h, \rho_l\}$ in order to maximize

$$\sum_i \pi_i \sum_k \rho_{ik} \left(V(q_k) - t_k + \alpha_k \big(P_{ki} - c\big)\right) \tag{2}$$

subject to the agent's participation constraints

$$t_h - \theta_h q_h - \alpha_h P_{hh} \geq 0, \tag{PC_h^d}$$
$$t_l - \theta_l q_l - \alpha_l P_{ll} \geq 0, \tag{PC_l^d}$$

the agent's incentive constraints

$$t_h - \theta_h q_h - \alpha_h P_{hh} \geq t_l - \theta_h q_l - \alpha_l P_{lh}, \tag{IC_h^d}$$
$$t_l - \theta_l q_l - \alpha_l P_{ll} \geq t_h - \theta_l q_h - \alpha_h P_{hl}, \tag{IC_l^d}$$

the agents randomizing constraints

$$\big(t_h - \theta_h q_h - \alpha_h P_{hh} - (t_l - \theta_h q_l - \alpha_l P_{lh})\big) \cdot \rho_h \leq 0, \tag{RC_h^d}$$
$$\big(t_l - \theta_l q_l - \alpha_l P_{ll} - (t_h - \theta_l q_h - \alpha_h P_{hl})\big) \cdot \rho_l \leq 0, \tag{RC_l^d}$$

the principals no-commitment constraints

$$\alpha_h \in arg \max_{0 \leq \alpha \leq 1} \alpha\{p_h P_{hl} + (1 - p_h) P_{hh} - c\} \tag{CC_h^d}$$
$$\alpha_l \in arg \max_{0 \leq \alpha \leq 1} \alpha\{p_l P_{ll} + (1 - p_l) P_{lh} - c\} \tag{CC_l^d}$$

and the Bayesian constraints

$$p_h = \frac{\pi(1 - \rho_l)}{\pi(1 - \rho_l) + (1 - \pi)\rho_h}, \tag{BC_h^d}$$
$$p_l = \frac{\pi \rho_l}{\pi \rho_l + (1 - \pi)(1 - \rho_h)}. \tag{BC_l^d}$$

Finding the optimal mechanism in this restricted class is still not straightforward, since for instance we do not know which of the first four constraints will be binding. In the following we look at particular subgames, starting from the reporting stage, and classify the best payoffs that can be achieved in each class of subgame. We divide subgames into those that induce audits with strictly positive probability and those that lead to no audit at all.

3.1 No-audit contract

If the solution to the principal's problem entails $\alpha_l = \alpha_h = 0$ it must correspond to the classical Baron-Myerson contract. Setting $P_{ij} = 0$, constraints (CC_h^d) and (CC_l^d) confirm the principal's strategy $\alpha_l = \alpha_h = 0$. Therefore we can drop the last four constraints. The resulting problem, i.e. maximizing (2) with respect to (PC_h^d), (PC_l^d), (IC_h^d), (IC_l^d), (RC_h^d) and (RC_l^d) corresponds to finding the optimal contract when audits are not available. But this is the classical Baron-Myerson contract. This result is summarized in the following lemma.

Lemma 1. *When the optimal one-shot mechanism induces no audits, i.e. $\alpha_l = \alpha_h = 0$, it is given by*

$$q_l^b = q_l^* \tag{3}$$

$$V'(q_h^b) = \theta_h + \frac{\pi}{1-\pi}\Delta\theta > \theta_h \tag{4}$$

$$t_h^b = \theta_h q_h^b \tag{5}$$

$$t_l^b = \theta_l q_l^b + \Delta\theta q_h^b \tag{6}$$

The agent truthfully reveals his type, i.e. $\rho_h = \rho_l = 1$.

The maximal profit the principal can achieve when no audits are induced is given by

$$V^b = \pi\big(V(q_l^*) - \theta_l q_l^* - \Delta\theta q_h^b\big) + (1-\pi)\big(V(q_h^b) - \theta_h q_h^b\big). \tag{7}$$

The agent receives a positive rent if of the efficient type, whereas the inefficient type does not receive any rent. The quantity produced by the efficient type is at its first-best level. The inefficient type's quantity is lower than the first-best level. It is this downward-distortion that reduces the rent paid to the agent. In the Baron-Myerson contract the trade-off rent versus efficiency is optimally resolved.

3.2 Audit contracts

We now turn to audit contracts, i.e. contracts that induce audits with strictly positive probability. A contract that induces audits necessarily exhibits positive penalties. The role these penalties play is however unclear. In the traditional understanding, penalties P_{ij} are used to deter agents from unwanted behavior. This is exactly how punishments are used in the full commitment contract. There the threat of a punishment makes the option of misreporting the type less attractive, without having to change transfers and quantities to the true type. The lack of commitment prevents the principal from directly applying this idea, since a costly audit is not performed when knowing that it won't lead to any penalty payment. If the principal wants to use penalty payments to make misreporting less attractive, this requires some level of misreporting when she is not able to commit to her audit strategy. Positive penalties can also be used to reduce effective transfer payments directly. For this the principal sets $P_{ii} = P$ and hence has an incentive to audit the respective message in order to collect the penalty. We show in the following lemma that the latter is always suboptimal. Audit contracts have a simple structure in that only after message θ_h the principal audits and only the efficient type has to pay a penalty when audited after sending message θ_h.

Proposition 1. *Whenever the optimal contract induces audits, i.e. some $\alpha_i > 0$, we obtain the following:*

$$\alpha_h = \alpha \in (0,1),\ \alpha_l = 0,\ P_{lh} = P_{ll} = P_{hh} = 0,\ P_{hl} = P$$

The inefficient type never randomizes, i.e. $\rho_h = 1$ and the efficient firm randomizes with probability $\rho_l = \rho \in (0,1)$.

Proposition 1 allows us to substantially simplify problem $\mathcal{P}$, whenever audits are performed in equilibrium. The principal's problem is to maximize

$$(\pi \cdot \rho)\big(V(q_l) - t_l\big) + (\pi \cdot (1-\rho) + 1 - \pi)\big(V(q_h) - t_h\big)$$

subject to

$$t_h - \theta_h q_h \geq 0 \tag{8}$$
$$t_l - \theta_l q_l \geq 0 \tag{9}$$
$$t_l - \theta_l q_l = t_h - \theta_l q_h - \alpha P \tag{10}$$
$$c = p_h \cdot P \tag{11}$$
$$p_h = \frac{\pi(1-\rho)}{\pi(1-\rho) + 1 - \pi} \tag{12}$$

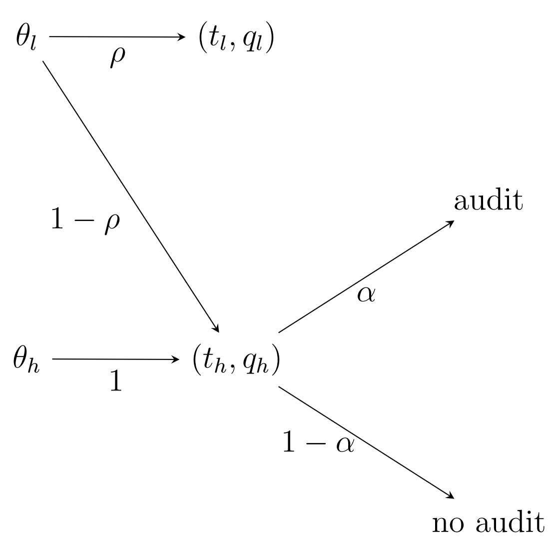

Figure 1: The audit contract

From (11) and (12) we can compute

$$\rho = \frac{\pi P - c}{\pi(P - c)}$$

The principals audit strategy α is chosen such that (10) holds with equality. The only remaining constraints are (8) and (9). Solving them for the transfers yields $t_l = \theta_l q_l$ and $t_h = \theta_h q_h$ and maximizing the principals objective results in the first best quantities $q_h = q_h^*$ and $q_l = q_l^*$.

The following Proposition summarizes

Proposition 2. *Suppose the optimal contract induces audits with strictly positive probability. Then it is given by $q_h = q_h^*$ and $q_l = q_l^*$ and the transfers are $t_h = \theta_h q_h^*$, resp. $t_l = \theta_l q_l^*$. The penalty function is the one given in Lemma 1. The principal audits only after message θ_h and this with probability $\alpha = (\Delta\theta\ q_h^*)/P$. The inefficient type always sends message θ_h, whereas the efficient type sends message θ_l with probability $\rho = (\pi P - c)/(\pi(P - c))$.*

The outcome of the audit contract is illustrated in figure 1. Note that the efficient type is indifferent between reporting θ_l and obtaining allocation t_l, q_l, and reporting θ_h and obtaining t_h, q_h but facing the risk of being audited with probability α. It is the audit probability α that makes the efficient type just indifferent. On the other hand, the principal is indifferent between auditing and not auditing when receiving the message θ_h. It is the mixing probability ρ of the efficient type that ensures this indifference. Effectively the principal and the agent play an inspection with asymmetric information and to achieve the mixed strategy equilibrium each players mixing probability must ensure indifference of the other player. In the next section I show how a mediator can be used to just break these indifference conditions and that this yields to strictly larger payoffs for the principal.

4 Mediated Communication

So far I made an important (simplifying) assumption: The principal is restricted to one-shot communication mechanisms. In this section I am going to prove that this assumption is not without loss of generality. To make this point, I study mediated communication between principal and agent. Assume the principal can make use of an impartial mediator in the following way: The agent is supposed to send some message $m \in M$ to the mediator. Given the message, a particular decision is taken as in the last section. Additionally the mediator gives a recommendation $r \in R$ to the principal. A crucial difference to the one-shot communication mechanism is the following: the principal does not observe the message sent by the agent, only the implemented decision and the recommendation received from the mediator. Based on this she may still update her belief and given this updated belief decide whether to audit. Compared to the direct mechanism from section 3 the principal is now not able to observe the actual message sent by the agent. If two distinct messages $m_1 \neq m_2$ lead to the same decision $d(m_1) = d(m_2)$ the principal is unable to tell the messages apart. Only together with the recommendation and the equilibrium reporting strategy of the agent she gets information to make a distinction. Mediated communication allows for more freedom in designing the information flow between agent and principal when playing the mechanism, in particular the principal is able to fine-tune the information received and thereby fine-tune her posterior probability assigned to the agents type. We will exploit this feature in designing a mediated communication mechanism that outperforms the best one-shot communication mechanism whenever it features pure randomization by the principal, i.e. whenever $\alpha \in (0, 1)$.

Consider the following mediated communication mechanism: The agent reports $\theta \in \Theta$ to the mediator. The message θ_l induces with probability β the pair t_l^a, q_l^a and the recommendation "no audit" to the principal, and with probability $1 - \beta$ the pair t_h^a, q_h^a and the recommendation "audit" to the principal. The message θ_h induces decision t_h^a, q_h^a with certainty and the recommendation "audit" is sent to the principal with probability ψ. This mechanism is illustrated in figure 2. Setting $\beta = \tau$ and $\psi = 1$ replicates the optimal one-shot mechanism. However the quantities from Proposition 2 can be implemented in several ways. The inefficiency of the audit mechanism is, that the efficient type produces the inefficient type's quantity with positive probability. With a mediated mechanism this inefficiency can be re-

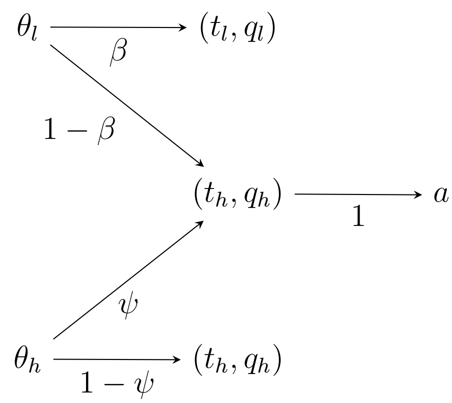

Figure 2: A mediated contract

duced, as will be shown in the following.
For this set $\psi = \alpha$. With this the efficient agents payoff from sending message θ_h to the mediator is the same as in the direct one-shot mechanism. The principal's posterior belief after recommendation "audit" is given by

$$\frac{\pi(1-\beta)}{\pi(1-\beta)+\alpha}. \tag{13}$$

Hence the principal follows this recommendation, whenever

$$\frac{\pi(1-\beta)}{\pi(1-\beta)+\alpha} \cdot P \geq c, \tag{14}$$

that is whenever $\beta \leq \big((P-c)\pi - c(1-\pi)\alpha\big)/\big((P-c)\pi\big)$. We are interested in raising the probability of the efficient type being assigned the efficient contract. Therefore we set $\beta = \big((P-c)\pi - c(1-\pi)\alpha\big)/\big((P-c)\pi\big)$. Clearly, the principal has no incentive to audit when not being recommended to do so. It remains to show that it is indeed optimal for each agent's type to truthfully report the type to the mediator. This is trivial for the inefficient agent. For the efficient agent it must hold that

$$\beta(t_l - \theta_l q_l^*) + (1-\beta)(t_h - \theta_h q_h^* - p) \geq t_h - \theta_h q_h^* - \alpha P. \tag{15}$$

Solving this constraint with equality for t_l yields

$$t_l = \theta_l q_l^* + \frac{1-\beta}{\beta}(P - \Delta\theta q_h^*). \tag{16}$$

Under this mediated communication mechanism the profit of the principal is

$$V^M = \phi\beta\big(V(q_l^*) - \theta_l q_l^* - \frac{1-\beta}{\beta}(P - \Delta\theta q_h^*)\big) + (1-\phi\beta)(V(q_h^*) - \theta_h q_h^*). \tag{17}$$

Recall the profit of the principal under the audit contract with pure randomization

$$V^A = \phi\rho\big(V(q_l^*) - \theta_l q_l^*\big) + (1-\phi\rho)(V(q_h^*) - \theta_h q_h^*). \tag{18}$$

The difference of the two profits is the following: In the mediated contract the efficient type is assigned the pair t_l, q_l with strictly larger probability, which by $V(q_l^*) - \theta_l q_l^* > V(q_h^*) - \theta_h q_h^*$ strictly increases profits. On the other hand, in the mediated mechanism the transfer t_l is larger to compensate the agent for frequent audits when being assigned the pair t_h, q_h. This reduces the principals profits. We are now going to show that the magnitude of the latter effect is always lower than the magnitude of the former effect. To see this, rewrite $V^M - V^A$ in the following way

$$\begin{aligned} V^M - V^A =& (\pi\beta - \pi\rho)\big(V(q_l^*) - \theta_l q_l^* - (V(q_h^*) - \theta_h q_h^*)\big) \\ & - \phi(1-\beta)(p - \Delta\theta q_h^*) \end{aligned}$$

Using the definitions of β and ρ we get

$$\begin{aligned} V^M - V^A =& \frac{(1-\pi)(1-\alpha)c}{P-c}\big(V(q_l^*) - \theta_l q_l^* - (V(q_h^*) - \theta_h q_h^*)\big) \\ & - \frac{(1-\pi)\alpha c}{P-c}(P - \Delta\theta q_h^*) \\ =& \frac{(1-\pi)c}{P-c}\Big((1-\alpha)\big(V(q_l^*) - \theta_l q_l^* - (V(q_h^*) - \theta_h q_h^*)\big) - \alpha(P - \Delta\theta q_h^*)\Big) \end{aligned}$$

Now recall, that $\alpha = \frac{\Delta\theta q_h^*}{P}$ in this case. Therefore

$$\begin{aligned} V^M - V^A =& \frac{(1-\pi)c}{P-c} \cdot \frac{P - \Delta\theta q_h^*}{P}\big(V(q_l^*) - \theta_l q_l^* - (V(q_h^*) - \theta_h q_h^*) - \Delta\theta q_h^*\big) \\ =& \frac{(1-\pi)c}{P-c} \cdot \frac{P - \Delta\theta q_h^*}{P}\big(V(q_l^*) - \theta_l q_l^* - (V(q_h^*) - \theta_l q_h^*)\big) \\ >& 0. \end{aligned}$$

The last inequality follows from $q_l^* = arg\max_g V(q) - \theta_l q$.

Therefore, the mediated contract improves upon the direct contract and makes the latter suboptimal. We summarize this finding in the following Proposition.

Proposition 3. *Whenever the optimal one-shot mechanism features strict randomization by the principal, i.e. whenever $\alpha \in (0, 1)$, there exists a mediated mechanism that yields strictly larger profits to the principal.*

Proposition 3 is our main finding. It shows that whenever the optimal one-shot communication mechanism involves randomization in the agent's reporting strategy, then there exists a mediated mechanism that yields strictly larger profits to the principal. Note that since the agent does not receive a positive rent in the one-shot mechanism, the mediated mechanism therefore also increases welfare. Mediation is strictly beneficial, because it allows for more freedom in designing the communication environment. In the one-shot communication mechanism the principal and the agent essentially play an inspection game. As we know from the classical inspection game, one player's randomization makes the other player just indifferent between her pure strategies. In this sense the mixing probabilities are self-referential. Furthermore, the efficient type is only willing to randomize between sending message θ_l and θ_h in the one-shot mechanism when he is indifferent between the respective outcomes q_l, t_l and q_h, t_h. With mediation the randomization is done by the mediator and not by the agent. The agent has to be merely incentivized to reveal his true type. In our particular mechanism, this implies the efficient type to be indifferent between two different randomizations, but not between the single elements the mediator randomizes over. The agent would strictly prefer t_l, q_l to anything else, but his choice not between t_l, q_l and t_h, q_h but between two randomizations involving the two and audits. The principals audit probability plays only a minor role in guaranteeing the agents indifference. It is rather a combination of the transfer t_l, the probability β and the probability α of sending recommendation "audit" to the principal after message θ_h was sent by the agent. All the mentioned values loose their traditional interpretation from the one-shot mechanism. α is not the audit probability anymore. The probability with which an audit is performed in the mediated mechanism is $\pi(1 - \beta) + (1 - \pi)\alpha$. By using the equilibrium values for α and β we can show that this expression actually is smaller than α. The mediated contract therefore uses less audits than the optimal audit contract. As already discussed before, in the mediated contract the efficient agent produces under the contract t_l, q_l more often, that is produces more often the for that type efficient quantity. This is reflected in the fact that β is strictly larger than τ. In the audit contract τ is typically interpreted as compliance. This interpretation is misleading in the mediated contract, since the agent always complies in reporting his true type to the mediator. The mediator assigns the inapt quantity on behalf of the principal in order to resolve her commitment problem. Lastly it is straightforward to see, that the efficient agent only receives his outside option in expectation. This feature is common from mediated contracts, e.g. Rahman and Obara (2010) heavily rely on this property to get their results and also in Goltsman et al. (2009)

we find this feature.

5 Discusssion

We have shown in this paper how mediated contracts can improve upon direct contracts in a particular contracting setting with limited commitment. Mediation is beneficial whenever the optimal one-shot communication mechanism involves randomization. In the direct contract any randomization requires indifference, which has to be guaranteed by the design of the mechanism and the particular equilibrium that is played. Using a mediator allows to break these indifference conditions and thereby allows for more flexibility in the design of the optimal mechanism.

Although in this paper we consider only a particular contracting setting, our results allow for the following prospect in the domain of contract theory: In settings where the standard revelation principle is not applicable, e.g. settings with limited commitment, and where the optimal one-shot mechanism features some randomization, it is not without loss of generality to use one-shot mechanisms in the first place. Mediation is beneficial exactly when randomization is to be induced in equilibrium.

This has consequences for the evaluation of existing contributions in this field. For instance Laffont and Tirole (1990) investigate optimal contracts in a dynamic principal-agent framework when contracts may be renegotiated during the contracting relationship. They proof that for a large set of parameters the equilibrium features randomization by some agent's type. The analysis in this paper suggests a re-examination of the obtained optimal contract using mediation, which may lead to strictly better contracts. A similar result is obtained in Freixas et al. (1985).

The setting we work in extends the one introduced in Bester and Strausz (2007). We show here that not only communication must be noisy when contracting with imperfect commitment, but that it is also necessary to allow for recommendations from the communication device to the principal. It is not sufficient for the principals Bayesian updating to learn the implemented decision, an additional informative signal from the the mediator may also be required. Therefore the general starting point for the analysis of optimal contracts in settings with limited commitment should be the class of optimal coordination mechanisms in generalized principal agent problems, introduced by Myerson (1982). The principal here has a dual role, as the principal for the parts of the contract she can commit to, and as an agent for the parts of the contract she cannot commit to. The optimal mechanism is then one where the true agent reports his type to the mediator and the principal receives a recommendation as she is the only player who has an action to take, that is not yet determined by the contract.

The analysis in this paper leaves open the structure of the optimal (mediated) contract in the described setting. This, and an extension of our findings to more general environments are left for future research.

References

Baron, D.P. and D. Besanko, "Regulation, asymmetric information, and auditing," *RAND Journal of Economics*, 1984, pp. 447–470.

Bester, H. and R. Strausz, "Contracting with imperfect commitment and the revelation principle: the single agent case," *Econometrica*, 2001, *69* (4), 1077–1098.

_ **and** _ , "Contracting with imperfect commitment and noisy communication," *Journal of Economic Theory*, 2007, *136* (1), 236–259.

Crawford, V.P. and J. Sobel, "Strategic information transmission," *Econometrica*, 1982, pp. 1431–1451.

Forges, F., "An approach to communication equilibria," *Econometrica*, 1986, pp. 1375–1385.

Freixas, X., R. Guesnerie, and J. Tirole, "Planning under incomplete information and the ratchet effect," *Review of Economic Studies*, 1985, *52* (2), 173–191.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani, "Mediation, arbitration and negotiation," *Journal of Economic Theory*, 2009, *144* (4), 1397–1420.

Khalil, F., "Auditing without commitment," *RAND Journal of Economics*, 1997, pp. 629–640.

Krishna, V. and J. Morgan, "The art of conversation: eliciting information from experts through multi-stage communication," *Journal of Economic Theory*, 2004, *117* (2), 147–179.

Laffont, J.J. and J. Tirole, "Adverse selection and renegotiation in procurement," *Review of Economic Studies*, 1990, *57* (4), 597–625.

Myerson, R.B., "Optimal coordination mechanisms in generalized principal–agent problems," *Journal of Mathematical Economics*, 1982, *10* (1), 67–81.

Nalebuff, B. and D. Scharfstein, "Testing in models of asymmetric information," *Review of Economic Studies*, 1987, *54* (2), 265–277.

Rahman, D. and I. Obara, "Mediated partnerships," *Econometrica*, 2010, *78* (1), 285–308.

A Proofs

Proof of Proposition 1: The proof consists of several steps, we proceed by showing some Lemmas first.

Lemma A.1. *If the optimal contract entails* $\alpha_l = 0$ *and* $\alpha_h > 0$ *we have* $P_{hl} = P$*,* $P_{hh} = P_{ll} = P_{lh} = 0$*. The efficient type randomizes, i.e.* $\rho_l < 1$*.*

Proof. First of all, since we have $\alpha_l = 0$ we can set $P_{lh} = P_{ll} = 0$ w.l.o.g. Then clearly $\alpha_l = 0$ remains optimal for the principal. Also we cannot have $P_{hl} = P_{hh}$. To see this, note that $P_{hl} = P_{hh} = 0$ contradicts $\alpha_h > 0$. Having $P_{hl} = P_{hh} = P > c$ would imply $\alpha_h = 1$. The principal thus obtains

$$(\pi\rho_l + (1-\pi)(1-\rho_h))(V(q_l) - t_l) + (\pi(1-\rho_l) + (1-\pi)\rho_h)(V(q_h) - t_h + P - c). \quad (19)$$

Reducing t_h to $\tilde{t}_h = t_h - p$ and setting all penalties to zero it is now easy to see that the same reporting strategies by the agents remain feasible. The principals profit however increases to

$$(\pi\rho_l + (1-\pi)(1-\rho_h))(V(q_l) - t_l) + (\pi(1-\rho_l) + (1-\pi)\rho_h)(V(q_h) - t_h + P). \quad (20)$$

Hence, $P_{hl} = P_{hh} = P$ cannot be optimal.
Now assume $P_{hh} = P$ and consequently $P_{hl} = 0$. The principal's profit is

$$(\pi\rho_l+(1-\pi)(1-\rho_h))(V(q_l)-t_l)+(\pi(1-\rho_l)+(1-\pi)\rho_h)(V(q_h)-t_h+\alpha_h((1-p_h)P-c)). \quad (21)$$

(PC_l^d) cannot be binding, since

$$t_l - \theta_l q_l \overset{(IC_l^d)}{\geq} t_h - \theta_l q_h > t_h - \theta_h q_h \geq t_h - \theta_h q_h - \alpha_h P \overset{(PC_h^d)}{\geq} 0. \quad (22)$$

Reducing all transfers by $\alpha_h P$, while keeping all quantities at the same level and setting all penalties to zero allows for a PBE with strictly larger profits. To see this, note that both participation constraints are still valid, in particular (PC_l^d) still follows from (IC_l^d) and (PC_h^d). As the incentive constraints are unaffected the same reporting strategies for the agent are implementable. No audits are conducted. The equilibrium profit is

$$(\pi\rho_l + (1-\pi)(1-\rho_h))(V(q_l) - t_l) + (\pi(1-\rho_l) + (1-\pi)\rho_h)(V(q_h) - t_h + \alpha_h P), \qquad (23)$$

which is larger than the profit in (21). Hence we cannot have $P_{hh} = P$.
Consequently we must have $P_{hl} = P$ and all other penalties equal to zero. To justify $\alpha_h > 0$ it must then also hold, that $\rho_l < 1$, which completes the proof. □

Lemma A.2. *The optimal contract cannot entail $\alpha_l > 0$ and $\alpha_h = 0$.*

Proof. Suppose $\alpha_l > 0$ and $\alpha_h = 0$. Again, we can assume w.l.o.g. that $P_{hi} = 0$. Furthermore, as in the proof of Lemma A.1 we must have $P_{ll} \neq P_{lh}$. Additionally, (PC_h^d) must be binding, since (PC_l^d) is already implied by (PC_h^d) and (IC_l^d), and one participation constraint must be binding.
We now consider the two relevant cases, $P_{ll} = P$, resp. $P_{lh} = P$.
Case 1: $P_{ll} = P$. Consequently we must have $P_{lh} = 0$. If $\rho_h = 1$ we had $\alpha_l = 1$ and could increase the principal's profit by replacing t_l with $t_l - P$ and reducing all penalties to zero.[6]
On the other hand, if $\rho_h < 1$ we can assume w.l.o.g. that $\rho_l = 1$, otherwise interchanging the roles of h and l makes Lemma A.1 applicable. From binding (PC_h^d) we derive $t_h = \theta_h q_h$ and from (IC_l^d) we get $t_l \geq \theta_l q_l + \Delta\theta q_h + P$, such that for the principal's expected profit V holds

$$V \leq (\pi+(1-\pi)(1-\rho_h))(V(q_l)-\theta_l q_l-\Delta\theta q_h-P+\alpha_l(p_l P-c))+(1-\pi)\rho_h(V(q_h)-\theta_h q_h) \quad (24)$$

Since $\alpha_l p_l \leq 1$ this implies

$$V < (\pi + (1-\pi)(1-\rho_h))(V(q_l) - \theta_l q_l - \Delta\theta q_h) + (1-\pi)\rho_h(V(q_h) - \theta_h q_h) \qquad (25)$$

The contract $\{(t_l, q_l), (t_h, q_h)\}$ sustains an equilibrium with reporting strategies $\rho_l = 1$ and ρ_h as in the contract above. For this, all penalties are set to zero and no audits are performed. In particular this contract is direct and incentive compatible, it therefore yields lower profits to the principal than the Baron-Myerson contract. Thus the contract we started with is not optimal.
Case 2: $P_{lh} = P$. Consequently $P_{ll} = 0$. Since $\alpha_l > 0$ we must have $\rho_h < 1$ in this case.

[6] This change does not affect the participation constraints and also not (IC_l^d). In (IC_h^d) only the right-hand side is reduced, which leaves $\rho_h = 1$ optimal. In the new subgame-equilibrium no audit takes place and all randomizations are the same, only less transfer is paid to the efficient type - hence profits are increased.

Hence (IC_h^d) is binding and together with (PC_h^d) this implies $t_h = \theta_h q_h$ and $t_l = \theta_h q_l + \alpha_l P$. The principal's expected profit is

$$\begin{aligned}
&(\pi\rho_l + (1-\pi)(1-\rho_h))(V(q_l) - \theta_h q_l - \alpha_l P + \alpha_l(P(1-p_l) - c)\\
&\quad + (\pi(1-\rho_l) + (1-\pi)\rho_h)(V(q_h) - \theta_h q_h)\\
=&(\pi\rho_l + (1-\pi)(1-\rho_h))(V(q_l) - \theta_h q_l) + (\pi(1-\rho_l) + (1-\pi)\rho_h)(V(q_h) - \theta_h q_h) \qquad (26)\\
&\quad - (\pi\rho_l + (1-\pi)(1-\rho_h))(p_l P + c)\\
<&(\pi\rho_l + (1-\pi)(1-\rho_h))(V(q_l) - \theta_h q_l) + (\pi(1-\rho_l) + (1-\pi)\rho_h)(V(q_h) - \theta_h q_h)
\end{aligned}$$

Since both types are willing to accept each of the two pairs (t_l, q_l), resp. (t_h, q_h),[7] the principal is therefore better off offering the contract that maximizes $V(q_i) - \theta_h q_i$ to both types without performing an audit. Hence, the original contract was not optimal.

□

Lemma A.3. *The optimal contract cannot entail $\alpha_l > 0$ and $\alpha_h > 0$.*

Proof. As in the two previous Lemmas it is straightforward to proof that $P_{ih} \neq P_{il}$ must hold for $i = h, l$. This yields four cases: (1) $P_{hh} = P_{ll} = P$ and $P_{hl} = P_{lh} = 0$; (2) $P_{hh} = P_{lh} = P$ and $P_{hl} = P_{ll} = 0$; (3) $P_{hl} = P_{ll} = P$ and $P_{hh} = P_{lh} = 0$; as well as (4) $P_{hl} = P_{lh} = P$ and $P_{hh} = P_{ll} = 0$.
Cases (1) and (2) can be ruled out by replicating the respective part of the proof of Lemma A.1.
Consider case (3), i.e. $P_{hl} = P_{ll} = P$ and $P_{hh} = P_{lh} = 0$.
Finally consider case (4), i.e. $P_{hl} = P_{lh} = P$ and $P_{hh} = P_{ll} = 0$. For this to yield $\alpha_l > 0$ and $\alpha_h > 0$ it must hold that both $\rho_l < 1$ and $\rho_h < 1$. The principal's profit reads as

$$(\pi\rho_l + (1-\pi)(1-\rho_h))\big(V(q_l) - t_l + \alpha_l((1-p_l)P - c)\big) + (\pi(1-\rho_l) + (1-\pi)\rho_h)\big(V(q_h) - t_h + \alpha_h(p_h P - c)\big)$$

In an optimum, the first-order conditions for ρ_l and ρ_h need to be satisfied.[8] Those conditions are

$$\text{w.r.t. } \rho_l \qquad V(q_l) - t_l = V(q_h) - t_h + \alpha_h P + (\alpha_l - \alpha_h)c \qquad (27)$$

$$\text{w.r.t. } \rho_h \qquad V(q_h) - t_h = V(q_l) - t_l + \alpha_l P - (\alpha_l - \alpha_h)c \qquad (28)$$

Obviously these equations are incompatible. □

[7] Since $0 = t_h - \theta_h q_h < t_h - \theta_l q_h$ the efficient firm produces when only (t_h, q_h) is offered. As (PC_l^d) is satisfied, this also holds whenever only (t_l, q_l) is offered.

[8] Recall that $\rho_i > 0$ in any optimal contract, thus no corner solutions for the ρ_i are possible.

It remains to show that α and ρ both cannot equal one. 1. $\square$

Proof of Proposition 2. The proof can be found in Khalil (1997). $\square$

Proof of Proposition 3. Follows directly from the arguments given in the main text. $\square$

Proof of Proposition ??. The Baron-Myerson contract yields expected profits of

$$\pi\left(V(q_l^b) - \theta_l q_l^b - \Delta\theta q_h^b\right) + (1-\pi)\left(V(q_h^b) - \theta_h q_h^b\right) \tag{29}$$

Consider the following mediated mechanism: $q_l^m = q_l^b$ and $q_h^m = q_h^b$ as well es $t_h^m = t_h^b$. Furthermore let $\beta_h = 1$ and $\beta_l = \beta$, i.e. the after report θ_l the agent is assigned t_l^m, q_l^m with probability β and t_h^m, q_h^m with probability $1-\beta$. After report θ_l the principal is recommended to audit with probability one if and only if t_h^m, q_h^m is assigned. After report θ_h this recommendation is given with probability α. Eventually the principal audits with probability ψ when recommended to do so.
To confirm the latter it must hold that

$$\frac{\pi(1-\beta)}{\pi(1-\beta) + (1-\pi)\alpha} P = c. \tag{30}$$

Furthermore, the efficient agent reports truthfully if

$$\beta\big(t_l^m - \theta_l q_l^m\big) + (1-\beta)\big(t_h^m - \theta_l^m q_h^m - \psi P\big) \geq t_h^m - \theta_l^m q_h^m - \psi\alpha P$$

Taking this condition binding and solving it for t_l yields

$$t_l = \theta_l q_l + \Delta\theta q_h + \frac{1-\beta-\alpha}{\beta}\psi P \tag{31}$$

Solving (30) for α yields

$$\alpha = \frac{\pi(1-\beta)(P-c)}{(1-\pi)c}.$$

Plugging this into (31) yields

$$t_l = \theta_l q_l + \Delta\theta q_h + \frac{1-\beta-\alpha}{\beta}\psi P$$

$\square$