Mechanism Design with Collusive Supervision[1]

Gorkem Celik[2]

February 7, 2008

[1]This is a very much revised version of a chapter in my dissertation submitted to Northwestern University. I am grateful to my advisors Michael Whinston and Jeff Ely for their encouragement, support, and continuing patience with me. The current version of the paper benefited from insightful comments by Tolga Yuret. Lester Kwong and Kim Lehrer provided valuable research assitance. I also thank an associate editor and two anonymous referees for their useful remarks and numerous teachers, colleagues, and students for valuable discussions. I am indebted to Koç University for their hospitality during the final stages of this project. I acknowledge financial support from Robert Eisner Graduate Fellowship, CSIO at Northwestern, and SSHRC Canada.

[2]Department of Economics, The University of British Columbia, #997-1873 East Mall, Vancouver, BC, V6T 1Z1 Canada. e-mail: celik@interchange.ubc.ca , phone: (604) 822 4814, fax: (604) 822 5915.

Abstract

We analyze an adverse selection environment with third party supervision. The *supervisor* is partly informed of the *agent*'s type. The supervisor and the agent collude while interacting with the *principal*. Contracting with the agent directly and ignoring the presence of the supervisor constitutes the no-supervision benchmark. We show that delegating to the supervisor reduces the principal's payoff compared to the no-supervision benchmark under a standard condition on the distribution of the agent's types. In contrast, if the principal contracts with both the agent and the supervisor, there exists a mechanism that improves the principal's payoff over the no-supervision payoff.

JEL Classification: D82; C72; L51

Keywords: Collusion, supervision, delegation, mechanism design

1 Introduction

The individual with the best information on the costs of an economic activity is often the "agent" who incurs these costs. One of the findings of the adverse selection literature is that a "principal" who wants to infer this information has to leave an information rent to the productive agent. In many circumstances, however, the agent is not the unique source for information on the production technology. The existence of an informed third party may improve the principal's payoff, by reducing the information rent he must sacrifice.[1] At the same time, the introduction of this "supervisor" creates the potential for collusive behavior against the principal's will. If the supervisor is completely corrupted by the agent, then the supervisor - agent pair behaves like a single player. In such a case, from the principal's perspective, contracting with the supervisor - agent coalition is no different from contracting with the agent in the absence of the supervisor.

An example fitting this discussion is a benevolent government's regulation of a firm with unknown cost. In this environment, the regulator can be thought of as the supervisor, whose close interaction with the firm provides her with better information on the cost. Another consequence of this close interaction is the regulator's vulnerability to capture by the firm: The regulator may end up as an advocate who protects the interests of the firm, rather than as an informant for the government. Other examples include the involvement of an auditor who reports the conduct of management to stockholders and an employee who reports the performance of a coworker to management.

A necessary condition for the principal to benefit from the supervisor's existence is an inefficiency in the performance of the supervisor - agent coalition. In the examples above, the need for supervision materializes as a response to an informational asymmetry between the principal and the agent. As such, it is natural to think that the supervisor may also be less informed than the agent. Once we introduce this possibility, supervision may matter.

[1] We use masculine pronouns for the principal and the agent, and feminine pronouns for the supervisor.

The contribution of this paper is the demonstration of how the principal can manipulate the interaction between these asymmetrically informed parties to support a coalitional inefficiency that serves his own interests.[2]

The most general organizational design for the principal is contracting with both the supervisor and the agent through a grand contract. A special case of this design would be the principal contracting with the supervisor and delegating to her the authority to contract with the agent. Delegation restricts the principal's ability to create direct incentives for the agent. Nevertheless, he can influence the supervisor's interaction with the agent to create indirect incentives. One commonly observed theme in the literature on multi-agent contracts is an organizational equivalence principle: From the principal's perspective, delegation performs as well as any other grand contract.[3] However, in our setup, this indirect influence scheme does not always fulfill the task. We show that delegating to the supervisor is dominated not only by the optimal grand contract, but also by not having access to the supervisor at all, as long as a condition on the distribution of the cost levels is satisfied. This condition corresponds to the condition which would ensure that the "monotonicity of the output levels" constraint is slack in the absence of supervision.

When delegation fails, it is still possible for the principal to benefit from the supervisory information. However, it is vital to keep the principal's communication with the agent open for beneficial supervision. Through this communication channel, the principal can provide the agent with an outside option to colluding with the supervisor. The agent can "blow the whistle"

[2]There is one strand of the earlier literature on collusion that sustains such a coalitional inefficiency by adopting exogenous transactional imperfections between the colluding parties. In this paper we do not make any such assumption regarding the transaction technology. In what follows, the only potential source for a collusion failure is the asymmetric information between the colluding parties. See Tirole [33] for a review of this *transaction cost* strand of the literature.

[3]See Melumad, Mookherjee, and Reichelstein [25] among others.

on the supervisor and contract directly with the principal.

Our model allows for three possible production cost levels for the agent. The supervisor can tell when the highest cost level is realized. However, she cannot distinguish the other two cost levels.[4] Modeling supervisory information as a non-trivial partition of the cost levels generates an environment where the supervisor is partly informed about the agent's cost. Obviously this is not the only way to model a partly informed supervisor. If the source for the supervisor's information is thought to be a signal correlated with the agent's cost, then our model corresponds to a particular - measure zero - specification, where the signals have disjoint supports on the cost space. One by-product of this information structure is the nested form of the information of different players. The supervisor knows less than the agent.[5] Therefore, collusion between these two players can be studied as a one sided asymmetric information problem, avoiding the informed principal considerations.[6] Another advantage of our information structure is a reduction in the number of incentive constraints that need to be accounted for when solving the collusion problem. This second point is particularly helpful in identifying the optimal collusion proof information rent levels.

It should be noted that the objective of the paper is not to provide a blanket statement that delegation is dominated regardless of the circumstances. Even in the context of our stylized model, when the monotonicity constraint is binding in the absence of supervision, delegation is an improvement over not having access to the supervisor. In this case, delegation may even

[4] The highest cost level may be the result of adopting an earlier generation of production technologies and the supervisor may be capable of identifying the relevant generation. The author thanks Jean Tirole for suggesting this interpretation of the connected partition structure.

[5] Another way of achieving a nested information structure is assuming that the agent has perfect information on the supervisor's signal. A paper by Faure-Grimaud, Laffont, and Martimort [9], which is discussed in the subsequent paragraphs, follows this latter approach.

[6] Quesada [30] studies collusion initiated by an informed party under one sided informational asymmetry.

be the principal's optimal response to collusion. Instead, our aim here is to demonstrate that delegation is not *always* the remedy to collusion. This is due to the fact that some forms of coalitional inefficiencies (specifically inefficiencies requiring an *overproduction* with respect to what is optimal for the coalition) are not possible to sustain under delegation. Moreover, when delegation fails, this does not imply there is no hope of making use of the supervisory information. We show that a centralized contract, where the principal directly communicates with the agent as well as with the supervisor, does indeed sustain the necessary coalitional inefficiencies that improve the principal's rent extraction over no-supervision.

In what follows, we model collusion as a "side contract" between the colluding parties. This methodology is due to two papers by Laffont and Martimort [18,20], where they identify the optimal outcome that is available for a principal contracting with two agents colluding under asymmetric information.[7] More recently, Che and Kim [6] consider a collusion setup, which is quite general in terms of the number of colluders, the distribution of types, and the production technology. They also allow for collusion to take place between a strict subset of the agents rather than being pervasive. They show that the *second best* payoff is attainable and, therefore, collusion does not inflict any harm on the principal.[8] However, due to the restrictions they impose on the correlation of information of the colluding parties, their result does not apply to the environment studied here.

A substantial portion of the research on collusion suggests that delegation is the optimal

[7] Another paper that formalizes collusion as a side contract is written by Caillaud and Jehiel [3] in the context of auctions with externalities. For models predating this approach see Kofman and Lawarree [16,17], and Felli [11]. Laffont and Martimort [18,20] study collusion between agents producing perfect complements. See Severinov [31] for collusion between agents producing substitutable goods.

[8] In related papers, Pavlov [29] and Che and Kim [7] establish conditions for the implementability of the second best payoff for an independent private values auction environment where bidders collude prior to participating in the auction.

organizational response to collusion. Baliga and Sjostrom [2], and Laffont and Martimort [19] establish the optimality of delegation in models of collusion between two productive agents in the contexts of moral hazard and adverse selection, respectively.[9] In contrast, Mookherjee and Tsumagari [27] demonstrate suboptimality of delegation when agents collude prior to making their decisions to participate in the mechanism. Mookherjee [26] provides a survey of this literature on delegation.

Another paper by Faure-Grimaud, Laffont, and Martimort [9] studies delegation in an informed supervisor - productive agent setup. In their setup, delegation is the optimal strategy for the principal under the possibility of supervisor - agent collusion.[10] This paper employs a connected partition structure to model the supervisor's information as opposed to the signals with full support that Faure-Grimaud, Laffont, and Martimort [9] utilize. These alternative information structures create two different environments. Benefitting from the supervisor's existence requires different types of coalitional inefficiencies in these different environments. In Section 5 of our paper, we discuss how delegation successfully creates one form of inefficiency but fails to sustain the other form.

The rest of the paper is organized as follows: In Section 2, we introduce the general model and outline the outcomes that are implementable i) under delegation to the supervisor and ii) under grand contracts, where the principal communicates with both players to eliminate the scope for collusion. In Section 3, we show why delegation fails as a response to the threat of collusion under a standard condition on the distribution of costs. In Section 4, we identify the optimal collusion proof outcome that achieves beneficial supervision when delegation fails

[9] To be precise, Baliga and Sjostrom [2] show that delegation is optimal for a wide range of parameters, but not all. Laffont and Martimort [19] show that delegation completely negates the adverse effects of collusion.

[10] Unlike us, Faure-Grimaud, Laffont, and Martimort [9] allow for risk aversion of the supervisor. Their optimal delegation mechanism attains the second best outcome when the supervisor is risk neutral. However, under risk aversion, collusion is harmful.

to do so. In Section 5, we show that the sufficient condition for the failure of delegation is also a necessary condition. When this condition is violated, delegation is an improvement over no-supervision. We conclude in Section 6. We relegate the proofs to the Appendix.

2 The Model

2.1 The No-Supervision Benchmark

The agent (A) is the player who bears the costs of production. His utility function is $t - \theta x$, where t is the transfer he receives from the principal, θ is the unit cost of production, and x is the output level. The variable θ takes on values from the set $\Theta = \left\{\underline{\theta}, \widetilde{\theta}, \overline{\theta}\right\}$, where $0 < \underline{\theta} < \widetilde{\theta} < \overline{\theta}$. The probability that the production cost is θ is denoted by $f(\theta)$. The distribution of the cost is common knowledge among the players. The agent also knows the realization of θ. Therefore, we refer to the variable θ as the type of A.

The principal (P) is the residual claimant of the output. For an output level x and a transfer level t, his payoff is $W(x) - t$. The function $W(\cdot)$ is twice continuously differentiable and satisfies the following standard conditions: $W'(x) > 0, W''(x) < 0$ for all x, and $\lim_{x\to 0} W'(x) = \infty$, $\lim_{x\to\infty} W'(x) = 0$. To induce A to participate in the productive activity, rather than not participating and receiving zero utility, P can commit to a contract. A contract is a collection of an arbitrary message space M, and two functions defined on M that specify:

i) the non-negative output level, $x : M \to \mathbb{R}_+$,

ii) the level of transfer to A, $t : M \to \mathbb{R}$.

Let x_θ be the output level when the realized type of the agent is θ. It follows from the standard treatment of this problem that an output profile $\{x_\theta\}_{\theta\in\Theta}$ is implementable through a contract if and only if it is weakly decreasing, i.e. $x_{\underline{\theta}} \geq x_{\widetilde{\theta}} \geq x_{\overline{\theta}}$. Moreover, the agent's lowest utility levels that are compatible with this implementation are revealed by the binding **IR** constraint of the highest cost type $\overline{\theta}$ and the binding upward adjacent **IC** constraints for the other types:

$$\begin{aligned} V^o_{\overline{\theta}}(\{x_\theta\}_{\theta\in\Theta}) &= 0, \\ V^o_{\widetilde{\theta}}(\{x_\theta\}_{\theta\in\Theta}) &= \left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}, \\ V^o_{\underline{\theta}}(\{x_\theta\}_{\theta\in\Theta}) &= \left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}. \end{aligned} \tag{1}$$

These utility levels reveal the information rent P is supposed to leave for A to induce A's truthful revelation of his type through his response to the contract. Accordingly, the optimal implementable set of output levels for P are determined by

$$\begin{aligned} \{x^{ns}_\theta\}_{\theta\in\Theta} &\in \arg\max_{\{x_\theta\}_{\theta\in\Theta}} \sum f(\theta)\left[W(x_\theta)-\theta x_\theta - V^o_\theta\left(\{x_\theta\}_{\theta\in\Theta}\right)\right] \\ \text{s.t. } x_{\underline{\theta}} &\geq x_{\widetilde{\theta}} \geq x_{\overline{\theta}}, \end{aligned} \tag{2}$$

where the superscript ns stands for no-supervision.

The assumption $\lim_{x\to 0} W'(x)=\infty$ implies that there is always some level of production, i.e. $x^{ns}_\theta > 0$ for all θ. Since the optimal information rent levels, identified by functions $V^o_\theta(\cdot)$, are increasing in the output levels of all types other than type $\underline{\theta}$, the optimal output levels for these types are distorted downward from their respective "first best" levels: The solution to the no-supervision problem induces "underproduction" with respect to the total welfare maximizing output levels. This underproduction phenomenon is a common feature of mechanism design problems provided that the productive agent's outside option yields the same utility level regardless of his type. Later in the paper, we will see that a type dependent reservation utility for the agent can reverse the direction of the inefficiency. This reversal is the key to beneficial supervision.

As is typical in similar design problems, requiring monotonicity of the output levels does not impose a restriction on the output level of the most efficient type in problem (2). Therefore constraint $x_{\underline{\theta}} \geq x_{\widetilde{\theta}}$ can be ignored. We refer to the remaining constraint $x_{\widetilde{\theta}} \geq x_{\overline{\theta}}$ as the *no-supervision monotonicity constraint.* When we ignore the monotonicity constraint and derive the first order conditions for problem (2), we notice that this constraint is slack if and only if

$$f\left(\widetilde{\theta}\right)\left(\overline{\theta}-\widetilde{\theta}\right) > f\left(\overline{\theta}\right) f\left(\underline{\theta}\right)\left(\widetilde{\theta}-\underline{\theta}\right). \tag{3}$$

When discussing delegation as an organizational form, we will see that its performance depends on whether this condition is satisfied.

2.2 Supervision

In this section, we introduce the supervisor (S) as an additional player in the principal - agent interaction. S does not have any direct interest in production. She is risk neutral and her payoff is determined only by the monetary payments she receives. The value of S to the design problem is in the information she possesses. S is able to tell whether A is the least efficient type, $\overline{\theta}$, or not. However, she is not able to distinguish types $\widetilde{\theta}$ and $\underline{\theta}$. Accordingly, S's information structure can be represented as the partition $\left\{\left\{\underline{\theta},\widetilde{\theta}\right\},\left\{\overline{\theta}\right\}\right\}$ of A's type space Θ. Notice that S observes a finer partition than does P, but a coarser partition than does A.

We define a **grand contract** as a collection of two arbitrary message spaces, M_S and M_A, as well as three functions defined on the product of these spaces. M_S and M_A consist of the messages that S and A can send to the principal respectively. The three functions specify:

i) the output level, $x : M_S \times M_A \rightarrow \mathbb{R}_+$,

ii) the transfer to A, $t : M_S \times M_A \rightarrow \mathbb{R}$,

iii) the wage of S, $w : M_S \times M_A \rightarrow \mathbb{R}$.

We now consider the output and payoff levels that are generated through a grand contract. As in the no-supervision benchmark, x_θ is defined as the output level of an agent with cost θ. Now we reserve the letter V to denote the *coalitional information rent*. That is, V_θ is the sum of the utility levels of A and S, whenever the agent's type is θ. Finally, U_θ is the utility level of the agent. We refer to $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ as an **outcome**. Notice that an outcome identifies the output level for every state of nature and the allocation of the rent created through the production of this output.

If any of the players rejects the grand contract, the game ends with zero production and no monetary transfer to the players. In other words, the opportunity cost of accepting the grand contract is 0 for both S and A. In order to ensure the acceptance of the grand contract by all agent types, P must leave them with a non-negative utility level:

$$U_\theta \geq 0 \text{ for all } \theta \in \Theta. \tag{4}$$

When A's type is θ, S's ex-post surplus is $V_\theta - U_\theta$. For S to participate in the mechanism that implements the outcome $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$, her interim expected surplus must be non-negative regardless of the partition cell she observes:

$$V_{\overline{\theta}} - U_{\overline{\theta}} \geq 0, \tag{5}$$

$$f\left(\widetilde{\theta}\right)\left(V_{\widetilde{\theta}} - U_{\widetilde{\theta}}\right) + f\left(\underline{\theta}\right)\left(V_{\underline{\theta}} - U_{\underline{\theta}}\right) \geq 0. \tag{6}$$

Under the interim participation constraints,[11] S can make a negative surplus in certain states of nature. Nevertheless she is willing to participate in the grand contract at the interim stage, since she cannot distinguish those states from those with positive surplus.[12]

Collusion Free Supervision

If there were no possibility of collusion in our environment, that is if S were not capable of offering a side contract to A, then her participation constraints depicted by (5) and (6) would be the only constraints governing S's behavior. By setting $V_\theta = U_\theta$ for all θ, P ensures S's participation in the mechanism. Since S's surplus is zero regardless of her report to P, S has no incentive to misreport the partition cell she observes. In other words, P is able to capture S's information for free. Since S's report reveals the information set she observes, A

[11] After observing the partition cell $\left\{\underline{\theta}, \widetilde{\theta}\right\}$, the conditional probability of facing a type $\theta \in \left\{\underline{\theta}, \widetilde{\theta}\right\}$ agent is $\frac{f(\theta)}{f(\underline{\theta})+f\left(\widetilde{\theta}\right)}$ for S. In the statement of the expected surplus, we rescale S's surplus by multiplying it by $f\left(\underline{\theta}\right)+f\left(\widetilde{\theta}\right)$.

[12] After establishing the relevance of supervision under collusion, we show that beneficial supervision can be sustained even under the stronger *ex-post participation constraints* for the supervisor.

can misreport his type only as some other type within the same information set. Accordingly, P can implement an output profile $\{x_\theta\}_{\theta\in\Theta}$ with the utility levels

$$\begin{aligned} V_{\overline{\theta}} &= U_{\overline{\theta}} = 0, \\ V_{\widetilde{\theta}} &= U_{\widetilde{\theta}} = 0, \\ V_{\underline{\theta}} &= U_{\underline{\theta}} = \left(\widetilde{\theta} - \underline{\theta}\right) x_{\widetilde{\theta}}, \end{aligned} \tag{7}$$

as long as $x_{\underline{\theta}} \geq x_{\widetilde{\theta}}$.

As before, the information rent to be paid to type $\overline{\theta}$ is zero. However, the existence of the supervisor gives P the opportunity to reduce the rent levels for the other types. To see this, note that it is not possible for type $\widetilde{\theta}$ to misreport his type as $\overline{\theta}$ when the supervisor truthfully reveals her information. Therefore, P is able to leave zero rent to both types $\widetilde{\theta}$ and $\overline{\theta}$. However, since type $\underline{\theta}$ still has the option of imitating type $\widetilde{\theta}$, he must be compensated with a positive information rent. Also notice that the monotonicity constraints are weaker with the introduction of the supervisor. For implementability under collusion free supervision, the output profile must be monotonic within each cell of S's information partition. That is, the outcome should respect the constraint $x_{\underline{\theta}} \geq x_{\widetilde{\theta}}$. However, there is no monotonicity requirement regarding output levels in separate partition cells.

With the possibility of collusion, the designer of the mechanism must account not only for the individual misreportings by the parties taking part in the contract but also for the collective manipulations of these individuals. In this case, the collusion free outcome constructed above is not implementable. This outcome leaves both A and S with zero rent when A's type is $\widetilde{\theta}$. However, A would have acquired a strictly positive rent by misreporting his type as $\overline{\theta}$. Such a misreport would require S's cooperation. Anticipating this potential for collusion, S can increase her expected surplus by offering a side contract asking for a *bribe* from A for playing along with the misreport.

Collusion and Supervision

To formalize the process of collusion, we assume that S offers a **side contract** to A after the acceptance of the grand contract by both parties. A side contract is a collection of a message space for A, M'_A, and two functions defined on M'_A, which specify:[13]

i) the messages that will be sent to P, $m : M'_A \to M_S \times M_A$,

ii) the bribe that A will pay to S, $b : M'_A \to \mathbb{R}$.

If A accepts the side contract, his message to the side contract, which is an element of M'_A, determines how S and A will respond to the grand contract.[14] If the side contract is rejected, both players are free to send any message they like.[15]

Each player is concerned with the total monetary transfer she or he receives or pays, but not with the source or destination of these transfers. As such, S's surplus is expressed as the sum of the wage and the bribe, $w + b$. Now that P and A are making monetary payments to this additional player, their utility functions must be amended to capture the effect of such transfers. Accordingly, we rewrite the utility functions of these players as $W(x) - t - w$, and $t - b - \theta x$ respectively.

Following Faure-Grimaud, Laffont, and Martimort [9], we study two organizational responses to collusion potential. The first organizational form we discuss is *delegation*. Under delegation, there is no direct communication or monetary transfer between P and A. Instead, P contracts with S only and delegates her the authority to contract with A through a side

[13] Alternatively, we could have defined the side contract to include a message space for S as well. Since S's type is known by A, this would not make a substantial difference in the analysis.

[14] In this paper, we ignore the issue of the enforcement of the contracts and assume that the side contract is binding as well as the grand contract. For treatments that relax the enforceability assumption see Martimort [23], Abdulkadiroglu and Chung [1], and Khalil and Lawarree [14].

[15] As is common in the earlier literature, we assume that S and A send their messages non-cooperatively whenever the side contract is rejected. For an example of an alternative treatment, where the proposer of the side contract commits to a certain message whenever the side contract is rejected, see Quesada [30].

contract. If S and A fail to agree on a contract, there is no production. In other words, P forgoes all direct interaction with A other than providing him with an exit option when A is not satisfied with S's offer.[16]

The second organizational response to collusion is *collusion proof implementation.* In this case, P contracts with both S and A directly through a grand contract. He designs the grand contract so that there is no scope for collusion between S and A. Therefore, S finds it optimal to offer a side contract that mimics the non-cooperative equilibrium of the game induced by the grand contract.

The main difference between these two organizational forms is the outside option they provide to A in the case where he rejects the side contract. Under delegation, A receives zero utility if collusion fails. However, when a collusion proof grand contract is in effect, it may ensure that A has a non-trivial outside option, providing him with a positive reservation utility. Moreover, the level of the reservation utility for A can change with his type. This last point will be crucial in establishing the relevance of collusion proof supervision.

Implementation

Under either organizational form, once a grand contract is accepted, the supervisor's side contract selection problem is quite similar to the principal's problem in the no-supervision benchmark. The outcomes available to S are determined by A's incentives. Among these available outcomes, S chooses the one that maximizes her expected surplus. Therefore, for an outcome to be implementable for P, it has to constitute a solution to the surplus maximization problem of S.

Suppose P wants to implement the outcome $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$. If A's type is θ, this outcome would provide S with the ex-post surplus of $V_\theta - U_\theta$. However, S can design a side contract

[16] Formally, delegation means offering a grand contract such that $M_A = \{\text{"}work\text{"}, \text{"}exit\text{"}\}$ and $t(m_s, m_a) = 0$ for all m_s and m_a. Moreover $x(m_s, \text{"}exit\text{"}) = 0$ and $w(m_s, \text{"}exit\text{"})$ is sufficiently small for all m_s. See Kim, Lawarree, and Shin [15] for a model where provision of the exit option is a choice variable for the principal.

that leads to the misreporting of A's type. We denote this misreport by $\hat{c}(\theta)$, where $\hat{c}: \Theta \rightarrow \Theta$. Under this misreport, the coalitional information rent is $V_{\hat{c}(\theta)} + (\hat{c}(\theta) - \theta)\, x_{\hat{c}(\theta)}$. The side contract also determines each colluding player's share of the information rent through the specification of the bribe. Let $\hat{u}(\theta)$ be A's share of the rent, where $\hat{u}: \Theta \rightarrow R$. We refer to $\{\hat{c}(\theta), \hat{u}(\theta)\}_{\theta\in\Theta}$ as a manipulation of the outcome $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$. Under a generic manipulation, S's ex-post surplus is $V_{\hat{c}(\theta)} + (\hat{c}(\theta) - \theta)\, x_{\hat{c}(\theta)} - \hat{u}(\theta)$. In order for $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ to be an implementable outcome, the expected value of $V_\theta - U_\theta$ must be weakly larger than the expected value of $V_{\hat{c}(\theta)} + (\hat{c}(\theta) - \theta)\, x_{\hat{c}(\theta)} - \hat{u}(\theta)$ under any manipulation available to S at the side contracting stage. Since S observes the relevant partition cell at the time that she offers the side contract, the expectation must be taken separately for each partition cell.

The remaining task is identifying the manipulations that are available to S at the side contracting stage. The first restriction on availability results from the fact that S cannot distinguish the different types within the same partition cell. This does not imply any restriction when A's type is $\overline{\theta}$. However, when S observes that A's type is either $\widetilde{\theta}$ or $\underline{\theta}$, she must provide A the incentive not to imitate the other type in the same partition cell. Accordingly, for $\{\hat{c}(\theta), \hat{u}(\theta)\}_{\theta\in\Theta}$ to be an available manipulation for S, the "agent's collusion stage incentive compatibility constraints" must be satisfied:

$$\mathbf{AIC}\left(\theta'|\theta\right) : \hat{u}(\theta) \geq \hat{u}\left(\theta'\right) + \left(\theta' - \theta\right) x_{\hat{c}(\theta')} \text{ for all } \theta, \theta' \in \left\{\widetilde{\theta}, \underline{\theta}\right\}.$$

The second restriction on the available manipulations is a consequence of the outside option of collusion for A. The outside option depends on the form of organizational structure P would follow. We start with the restriction under delegation. The outside option under this organizational form is the shut down of production, which provides A with zero reservation utility regardless of his realized type. In addition to the **AIC** constraints, for $\{\hat{c}(\theta), \hat{u}(\theta)\}_{\theta\in\Theta}$ to be an available manipulation for S, the "agent's individual rationality constraints under delegation" must be satisfied:

$$\mathbf{d}-\mathbf{AIR}\left(\theta\right):\hat{u}\left(\theta\right)\geq 0 \text{ for all } \theta\in\Theta.$$

Now we can state the conditions that reflect the feasibility of the outcome $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ under delegation:

$$\begin{aligned}\left\{\overline{\theta},U_{\overline{\theta}}\right\} \in & \arg\max_{\left\{\hat{c}\left(\overline{\theta}\right),\hat{u}\left(\overline{\theta}\right)\right\}}\left[V_{\hat{c}\left(\overline{\theta}\right)}+\left(\hat{c}\left(\overline{\theta}\right)-\overline{\theta}\right)x_{\hat{c}\left(\overline{\theta}\right)}-\hat{u}\left(\overline{\theta}\right)\right] \\ & \text{s.t. } \mathbf{d}-\mathbf{AIR}\left(\overline{\theta}\right).\end{aligned} \tag{8}$$

$$\begin{aligned}\left\{\theta,U_{\theta}\right\}_{\theta\in\left\{\widetilde{\theta},\underline{\theta}\right\}} \in & \arg\max_{\left\{\hat{c}(\theta),\hat{u}(\theta)\right\}_{\theta\in\left\{\widetilde{\theta},\underline{\theta}\right\}}}\sum_{\theta\in\left\{\widetilde{\theta},\underline{\theta}\right\}}f\left(\theta\right)\left[V_{\hat{c}(\theta)}+\left(\hat{c}\left(\theta\right)-\theta\right)x_{\hat{c}(\theta)}-\hat{u}\left(\theta\right)\right] \\ & \text{s.t. } \mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right),\mathbf{AIC}\left(\widetilde{\theta}|\underline{\theta}\right),\mathbf{d}-\mathbf{AIR}\left(\underline{\theta}\right),\text{ and } \mathbf{d}-\mathbf{AIR}\left(\widetilde{\theta}\right).\end{aligned} \tag{9}$$

We refer to conditions (8) and (9) as the **delegation feasibility conditions**.[17] Our terminology here follows that of Holmstrom and Myerson [12] who define the property of "incentive feasibility," to refer to the set of outcomes such that no type of agent is willing to misreport his type. Similarly, delegation feasibility requires that S is unwilling to offer a side contract that misreports A's type.

Together with the participation constraints, delegation feasibility conditions describe the set of implementable outcomes under delegation.

Definition 1 $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ *is a* ***delegation proof*** *outcome if it satisfies the participation constraints (4), (5), (6), and the delegation feasibility conditions (8), (9).*

As an alternative to delegation, P can follow a collusion proof design and provide A with U_θ as the outside option to collusion. In this case, any manipulation that is available to S through an acceptable side contract should leave an agent of type θ with at least the utility level U_θ.

[17] Following the notation introduced in the previous footnote, a *direct* delegation grand contract includes message sets $M_A = \{\text{"}work\text{"}, \text{"}exit\text{"}\}$ and $M_S = \Theta$. A direct contract such that $x(\theta, \text{"}work\text{"}) = x_\theta$ and $w(\theta, \text{"}work\text{"}) = V_\theta + \theta x_\theta$ is sufficient to induce a delegation feasible outcome.

These constraints on manipulations are represented by the "agent's collusion stage individual rationality constraints:"

$$\mathbf{AIR}\left(\theta\right):\hat{u}\left(\theta\right)\geq U_{\theta}\text{ for all }\theta\in\Theta.$$

Finally, we present the "**collusion feasibility conditions**" which guarantee that S is unable to find a side contract that would improve her expected surplus:

$$\begin{aligned}\left\{\overline{\theta},U_{\overline{\theta}}\right\} &\in \arg\max_{\left\{\hat{c}\left(\overline{\theta}\right),\hat{u}\left(\overline{\theta}\right)\right\}}\left[V_{\hat{c}\left(\overline{\theta}\right)}+\left(\hat{c}\left(\overline{\theta}\right)-\overline{\theta}\right)x_{\hat{c}\left(\overline{\theta}\right)}-\hat{u}\left(\overline{\theta}\right)\right] \\ &\quad\text{s.t. }\mathbf{AIR}\left(\overline{\theta}\right).\end{aligned}\tag{10}$$

$$\begin{aligned}\left\{\theta,U_{\theta}\right\}_{\theta\in\left\{\widetilde{\theta},\underline{\theta}\right\}} &\in \arg\max_{\left\{\hat{c}(\theta),\hat{u}(\theta)\right\}_{\theta\in\left\{\widetilde{\theta},\underline{\theta}\right\}}}\sum_{\theta\in\left\{\widetilde{\theta},\underline{\theta}\right\}}f\left(\theta\right)\left[V_{\hat{c}(\theta)}+\left(\hat{c}\left(\theta\right)-\theta\right)x_{\hat{c}(\theta)}-\hat{u}\left(\theta\right)\right] \\ &\quad\text{s.t. }\mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right),\mathbf{AIC}\left(\widetilde{\theta}|\underline{\theta}\right),\mathbf{AIR}\left(\underline{\theta}\right),\text{ and }\mathbf{AIR}\left(\widetilde{\theta}\right).\end{aligned}\tag{11}$$

Notice that collusion feasibility is a weaker concept than delegation feasibility, since solutions to (8) and (9) are also solutions to (10) and (11), respectively. This is not surprising since delegation implies a loss of control for the principal compared to the centralized collusion proof design.

To characterize the set of implementable outcomes under collusion proof implementation, we merge the participation constraints with the feasibility conditions above.

Definition 2 $\{x_{\theta},V_{\theta},U_{\theta}\}_{\theta\in\Theta}$ *is a* ***collusion proof*** *outcome if it satisfies the participation constraints (4), (5), (6), and the collusion feasibility conditions (10), (11).*

Since we have defined V_{θ} as the sum of the information rents for S and A, the expected payoff for P is given as

$$\sum_{\theta\in\Theta}f\left(\theta\right)\left[W\left(x_{\theta}\right)-\theta x_{\theta}-V_{\theta}\right].\tag{12}$$

This is the same objective function for P as in the no-supervision problem (2), provided that the functions labeled $V_{\theta}^{o}\left(\cdot\right)$ are replaced with the corresponding V_{θ}'s. P's task is finding an implementable (delegation proof or collusion proof depending on the organizational form assumed) outcome that maximizes this expected payoff. Note that P does not have any intrinsic preference on the distribution of the information rent V_θ between S and A, i.e., $\{U_\theta\}_{\theta\in\Theta}$ does not enter into his objective function. Nevertheless the distribution of the rent (the values for $\{U_\theta\}_{\theta\in\Theta}$) is relevant to P to ensure the implementability of an outcome.

In the sections that follow, we analyze the strengths and the weaknesses of delegation and collusion proof implementation from the principal's perspective.

3 Failure of Delegation

The performance of delegation as an organizational form is of interest to us for two reasons. First, there may be economic environments where the principal is forced to follow the delegation path due to a variety of exogenous factors including communication and information processing costs. We would like to know if this constraint inflicts any cost on the principal. Moreover, the analysis of how delegation performs in our setup will be helpful in contrasting the result we derive in the following section with similar results that suggest that supervision is beneficial even under the possibility of collusion.

Faure-Grimaud and Martimort [10] study delegation in a setup with an identical type and information structure as here. Unlike this paper, they allow for the risk aversion of the supervisor but assume that the highest cost agent is inefficient to the extent that it is never optimal for him to produce, i.e. $x_{\overline{\theta}}$ is set to zero. The first part of our discussion will build on an extension of their analysis for positive values of $x_{\overline{\theta}}$.

As we mentioned earlier, the defining feature of delegation is the shut down of production whenever S and A do not agree on a side contract. This implies that the outside option of a side contract for A is zero utility, as is the outside option of a no-supervision contract. Therefore, the delegation proof utility profile for A should mimic the construction of the function $V_\theta^o(\cdot)$

for both information sets of S. That is, the utility of A is determined by the binding $\mathbf{d}-\mathbf{AIR}$ constraints for cost levels $\overline{\theta}$ and $\widetilde{\theta}$, and by the binding $\mathbf{AIC}\left(\widetilde{\theta}|\underline{\theta}\right)$ constraint for cost level $\underline{\theta}$, i.e. $U_{\overline{\theta}}=U_{\widetilde{\theta}}=0$ and $U_{\underline{\theta}}=\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}$. A delegation proof outcome should also leave the appropriate rent to S so that she does not find it profitable to offer a side contract misreporting A's type to P.

Proposition 1 *Suppose the outcome* $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ *is delegation proof and the output profile* $\{x_\theta\}_{\theta\in\Theta}$ *is monotonic, i.e.* $x_{\overline{\theta}}\leq x_{\widetilde{\theta}}\leq x_{\underline{\theta}}$. *Then,*

$$\begin{aligned}
V_{\overline{\theta}} &\geq 0, \\
V_{\widetilde{\theta}} &\geq \left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\frac{f\left(\underline{\theta}\right)}{f\left(\widetilde{\theta}\right)}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\widetilde{\theta}}-x_{\overline{\theta}}\right), \qquad (13)\\
V_{\underline{\theta}} &\geq \left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}+\frac{f\left(\underline{\theta}\right)}{f\left(\widetilde{\theta}\right)}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\widetilde{\theta}}-x_{\overline{\theta}}\right).
\end{aligned}$$

The inequalities in (13) establish lower bounds for delegation proof coalitional rent levels implementing a monotonic output profile. Notice that the lower bounds for $V_{\widetilde{\theta}}$ and $V_{\underline{\theta}}$ are larger than the no-supervision information rents $V^o_{\widetilde{\theta}}(\cdot)$ and $V^o_{\underline{\theta}}(\cdot)$ calculated for the same profile of output levels. To understand these larger rents, suppose P tries to implement an outcome leaving $V^o_{\widetilde{\theta}}(\cdot)$ to the coalition whenever the agent's type is $\widetilde{\theta}$. In this case, S would offer a side contract instructing this type of agent to misreport his type as $\overline{\theta}$. Assuming $V_{\overline{\theta}}$ is set optimally at 0, the misreporting of the type $\widetilde{\theta}$ agent is not beneficial for S in this state of nature. The benefit of the misreport is accrued whenever the agent's type is $\underline{\theta}$. Now that type $\widetilde{\theta}$ is producing the lower output level $x_{\overline{\theta}}$ instead of $x_{\widetilde{\theta}}$, S can reduce the information rent she is supposed to leave to the more efficient type $\underline{\theta}$ from $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}$ to $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\overline{\theta}}$. The additional term $\frac{f(\underline{\theta})}{f(\widetilde{\theta})}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\widetilde{\theta}}-x_{\overline{\theta}}\right)$ on the lower bounds of $V_{\widetilde{\theta}}$ and $V_{\underline{\theta}}$ captures this rent which is left to S for not misreporting type $\widetilde{\theta}$ in order to reduce the information rent of type $\underline{\theta}$. Faure-Grimaud and Martimort [10] label this additional rent to the supervisor as the principal's *agency cost of* *intermediation.*[18]

The increasing coalitional rent under delegation is also reminiscent of the "double marginalization" of the information rents when there is an uninformed third party intermediating between the principal and the agent. Since the third party is not informed, she must leave an information rent to A. When this information rent is taken into account, the third party behaves as though she is a productive agent, whose production cost is equal to the "virtual" cost rather than the original cost A incurs. To substantiate this discussion, we consider a variant of our model, where S observes the same information as P. When P delegates to this uninformed supervisor, the coalitional information rent levels still respect the inequality

$$V_{\widetilde{\theta}} - V_{\overline{\theta}} \geq \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} + \frac{f\left(\underline{\theta}\right)}{f\left(\widetilde{\theta}\right)} \left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\widetilde{\theta}} - x_{\overline{\theta}}\right) \tag{14}$$

as before. However, since the supervisor's participation decision is made in the interim stage, without any information on A's type, the double rent, measured by the term $\frac{f(\underline{\theta})}{f(\widetilde{\theta})} \left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\widetilde{\theta}} - x_{\overline{\theta}}\right)$, can be taxed away by setting $V_{\overline{\theta}}$ to be negative. This observation is also an implication of a result by Melumad, Mookherjee, and Reichelstein [25], which is derived in a setup, where multiple agents contribute to production and hold private information. Their result shows that when the agents' types are *independently* distributed, the agents are risk neutral, and their participation decisions are made in the interim stage, then the principal can obtain the "collusion free" second best payoff by delegating to one of these agents.[19]

When types are *correlated*, as is the case when the supervisor's partial information is nested in the agent's private information, the result by Melumad, Mookherjee, and Reichelstein [25]

[18] Unlike in the study of Faure-Grimaud and Martimort [10], the lower bounds stated in (13) are not necessarily attainable for all parameterizations. In fact, there is no coalitional information rent profile to implement strictly monotonic output levels if the parameters satisfy $f\left(\underline{\theta}\right) \left(\widetilde{\theta} - \underline{\theta}\right) > f\left(\widetilde{\theta}\right) \left(\overline{\theta} - \widetilde{\theta}\right)$.

[19] Another condition for achieving the second best payoff is the observability of the agent specific production contributions. This requirement is trivially satisfied in the current setup since there is only one productive agent.

is no longer valid for examining the performance of delegation. If the supervisor's information takes the structure $\left\{\left\{\underline{\theta}, \widetilde{\theta}\right\}, \{\overline{\theta}\}\right\}$, her participation constraints stipulate that $V_{\overline{\theta}}$ is non-negative. Since S can misreport the agent as type $\overline{\theta}$ even after observing $\left\{\underline{\theta}, \widetilde{\theta}\right\}$, she must be compensated not to misreport. Accordingly, the coalitional rent must be positive even when A is the least efficient type within the partition cell $\left\{\underline{\theta}, \widetilde{\theta}\right\}$. This brings our model closer to a model constructed by McAfee and McMillan [24] to study delegation of an uninformed and limitedly liable supervisor (the *middle principal* in their terminology). McAfee and McMillan [24] demonstrate that the double rent for this supervisor cannot be taxed away due to the limited liability constraints. In the current model, there is no exogenous limited liability constraint for the supervisor. Nevertheless, the supervisor can still secure a non-negative surplus for herself by misreporting an agent as type $\overline{\theta}$.[20]

It is immediate from the proposition above that delegating to the supervisor performs (strictly) worse than the no-supervision case in implementing (strictly) *monotonic* output levels. However, as we have discussed earlier, it is also possible to implement *non-monotonic* output levels when a supervisor is present. In order to give a full comparison of these two setups, we have to examine the implementation of non-monotonic output levels as well. We relegate this analysis to the Appendix and report our findings in the proposition that follows.

Proposition 2 ***(Failure of Delegation)*** *Suppose the monotonicity constraint is slack for the no-supervision problem, i.e. (3) holds and therefore $x^{ns}_{\overline{\theta}} < x^{ns}_{\widetilde{\theta}}$. Then P's expected payoff from any delegation proof outcome is strictly lower than his optimal no-supervision expected payoff.*

[20] Similarly, Mookherjee and Tsumagari [27] demonstrate that delegation to one of the two productive agents (with independent types) is costly when these agents make their participation decisions after colluding and learning each other's type. The agents can refuse to participate if the mechanism leaves them with negative rent. The timing in our paper does not permit the players to collude on their participation decisions. However, the colluding parties can guarantee non-negative coalitional rents by collectively reporting the production cost as the highest possible.

Before ending our discussion of this section, we should note the significance of the slackness of the no-supervision monotonicity constraint in the derivation of Proposition 2. We observe in Section 5 that delegation is indeed an improvement over no-supervision if this monotonicity constraint is relevant. Nevertheless, the current result presents a class of parameterizations, identified by condition (3), where delegation to a partly informed supervisor is dominated by no-supervision from the principal's viewpoint. In the following section, we show that there is a collusion proof outcome that dominates no-supervision for all parameterizations.

4 Collusion Proof Supervision

4.1 Relevance of Collusion Proof Supervision

So far we have proved that delegating to S is not the right organizational form for the principal, if a standard assumption ensuring that the monotonicity constraint does not matter for the no-supervision problem. As we have mentioned before, delegation is only a special case of implementation with supervision. In order to say more regarding the value of supervision, we need to consider collusion proof implementation, which yields a larger set of implementable outcomes. In this section we pursue this task.

The advantage of collusion proof supervision stems from P's ability to provide A with positive reservation utility when side contracting with S. To see the effect of a positive reservation utility, consider the lower bounds on the delegation proof coalitional information rent levels depicted in (13). By increasing the reservation utility for the most efficient type $\underline{\theta}$, P can decrease these lower bounds. This reservation utility is denoted as $u_{\underline{\theta}}$. For values of $u_{\underline{\theta}}$ smaller than $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\overline{\theta}}$, there is no change in the lower bounds as in (13). By offering a side contract which misreports type $\widetilde{\theta}$ as type $\overline{\theta}$, S can reduce the information rent she is supposed to leave the most efficient type $\underline{\theta}$ from $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}$ to $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\overline{\theta}}$. As explained in the previous section, to prevent such a misreport, P must leave S with a double rent of the amount $\frac{f(\underline{\theta})}{f(\widetilde{\theta})}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\widetilde{\theta}}-x_{\overline{\theta}}\right)$

over the no-supervision information rent. However, when $u_{\underline{\theta}}$ is set between $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\overline{\theta}}$ and $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}$, misreporting type $\widetilde{\theta}$ as type $\overline{\theta}$ cannot reduce the information rent of type $\underline{\theta}$ lower than $u_{\underline{\theta}}$. Accordingly, leaving S with a double rent at the amount $\frac{f(\underline{\theta})}{f(\widetilde{\theta})}\left[\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}-u_{\underline{\theta}}\right]$ is sufficient to preclude this misreport. Finally, when $u_{\underline{\theta}}$ is $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}$, misreporting type $\widetilde{\theta}$ as type $\overline{\theta}$ does not change type $\underline{\theta}$'s information rent. In this case, the lower bounds on the coalitional rent levels $V_{\widetilde{\theta}}$ and $V_{\underline{\theta}}$ are equal to the no-supervision information rents $V^o_{\widetilde{\theta}}(\cdot)$ and $V^o_{\underline{\theta}}(\cdot)$ calculated for the same profile of output levels.

Another way of demonstrating that P can implement the no-supervision information rent levels in a collusion proof way is described here. For any output and information rent profile $\{x_\theta, V_\theta\}_{\theta\in\Theta}$ implementable under no-supervision, we can find a collusion proof outcome that replicates this same pair by setting $U_\theta = V_\theta$ for all θ. Such an implementation can be regarded as ignoring the existence of S and contracting only with A. However, the more appealing question here is whether the principal can benefit from the potentially weaker constraints associated with supervision. In other words, we would like to know whether supervision makes it possible to induce an output and information rent profile $\{x_\theta, V_\theta\}_{\theta\in\Theta}$ that is *not* no-supervision implementable and that provides a payoff to P that is higher than his no-supervision optimal payoff. The answer to this question determines the "relevance" of supervision in collusive environments.

For supervision to be relevant, P must be able to ensure that the S - A coalition fails to behave like a single composite player who maximizes the coalitional gains by means of a side contract. More precisely, a *necessary* condition for relevance is the existence of a collusion proof outcome that violates some of the binding constraints of the no-supervision problem. Recall that when the monotonicity constraint is slack, the only binding constraints of the no-supervision problem are the **IR** constraint of type $\overline{\theta}$ and the upward adjacent **IC** constraints of types $\widetilde{\theta}$ and $\underline{\theta}$. The following proposition shows that the no-supervision **IC** constraint regarding type $\underline{\theta}$ is indeed violated by the optimal collusion proof coalitional information rents implementing a monotonic output profile.

Proposition 3 ***(Relevance)****Any monotonic output profile $x_{\overline{\theta}} \leq x_{\widetilde{\theta}} \leq x_{\underline{\theta}}$ is collusion proof. The lowest coalitional information rent levels and the corresponding utility levels for A that make the collusion proof implementation of a monotonic output profile possible are given as*

$$\begin{aligned} V_{\overline{\theta}} &= 0, & U_{\overline{\theta}} &= 0, \\ V_{\widetilde{\theta}} &= \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}, & U_{\widetilde{\theta}} &= 0, \\ V_{\underline{\theta}} &= \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} + \left(\widetilde{\theta} - \underline{\theta}\right) x_{\widetilde{\theta}} - \tfrac{f(\widetilde{\theta})}{f(\underline{\theta})}\Pi, & U_{\underline{\theta}} &= \left(\widetilde{\theta} - \underline{\theta}\right) x_{\widetilde{\theta}} + \Pi, \end{aligned} \tag{15}$$

where $\Pi = \min\left\{\left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}, \left(\widetilde{\theta} - \underline{\theta}\right)\left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right)\right\}$.

An examination of the coalitional information rent levels in (15) indicate that the collusion proof outcome, or the class of outcomes (since each monotonic output profile gives a different outcome), above induces an improvement over no-supervision implementation. To see this, suppose the profile being implemented is the optimal no-supervision output profile $\{x_{\theta}^{ns}\}_{\theta\in\Theta}$. Notice that the coalitional information rent levels $V_{\overline{\theta}}$ and $V_{\widetilde{\theta}}$ in (15) are the same as the corresponding no-supervision information rent levels $V_{\overline{\theta}}^{o}(\cdot)$ and $V_{\widetilde{\theta}}^{o}(\cdot)$ evaluated for $\{x_{\theta}^{ns}\}_{\theta\in\Theta}$. However, $V_{\underline{\theta}}$, the coalitional rent for the state of the world where A is type $\underline{\theta}$, is strictly lower than its no-supervision counterpart $V_{\underline{\theta}}^{o}(\cdot)$. This is because the optimal no-supervision outcome induces positive output levels for all types $\left(x_{\overline{\theta}}^{ns} > 0\right)$ and the monotonicity constraint is slack for the most efficient type $\left(x_{\underline{\theta}}^{ns} > x_{\widetilde{\theta}}^{ns}\right)$.

The proof of Proposition 3, presented in the Appendix, demonstrates that the information rent levels in (15) are implementable with a monotonic output profile given Π is a non-negative number smaller than or equal to both $\left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}$ and $\left(\widetilde{\theta} - \underline{\theta}\right)\left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right)$. For such values of Π, P can reduce the coalitional information rent from the no-supervision information rent levels for type $\underline{\theta}$ by $\frac{f(\widetilde{\theta})}{f(\underline{\theta})}\Pi$ by raising the reservation utility of type $\underline{\theta}$ by Π. Notice though that the resulting outcome is not coalitionally efficient: S and A could increase their total utility by misreporting type $\underline{\theta}$ as type $\widetilde{\theta}$. This misreporting increases S's surplus by $\Pi + \frac{f(\widetilde{\theta})}{f(\underline{\theta})}\Pi$ and decreases A's utility by Π. If S was certain of A being type $\underline{\theta}$, she would be willing to pay a bribe of Π in order to persuade A to misreport his type as $\widetilde{\theta}$. The problem for S is that she cannot differentiate between types $\underline{\theta}$ and $\widetilde{\theta}$. Therefore, the latter type can always imitate the former to get the same bribe. When weighted by the corresponding probabilities of these types, the expected coalitional gain from misreporting type $\underline{\theta}$ is cancelled by the bribe that S must pay A to realize the gain.

The argument above suggests that inducing a higher reservation utility for the type $\underline{\theta}$ agent increases the payoff of the principal. The increase in the reservation utility of type $\underline{\theta}$ that P can sustain is determined by the following two factors. First, as stipulated by her participation constraint (6), S's share of the information rent cannot fall below zero. Moreover, the agent of type $\widetilde{\theta}$ must not find it profitable to imitate type $\underline{\theta}$ (as stipulated by $\mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right)$). The value of Π stated in Proposition 3 reflects these two upper bounds on A's reservation utility.

When S knows that the type of A is either $\widetilde{\theta}$ or $\underline{\theta}$, the relevant constraints for her maximization are $\mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right)$ and $\mathbf{AIR}\left(\underline{\theta}\right)$, the incentive compatibility constraint of the relatively inefficient type and the individual rationality constraint of the efficient type. This is contrary to what we would expect under uniform reservation utilities. When considering S's maximization, A's type not only signals his cost parameter, but also his outside option if he refuses S's collusion offer. By increasing type $\underline{\theta}$'s reservation utility, the principal provides a "countervailing incentive" for A in the collusion problem.[21,22]

We can substitute in the coalitional information rent levels in (15) to find the optimal

[21] See Lewis and Sappington [21], Maggi and Rodriguez-Clare [22], and Julien [13] for discussions of how exogenous type specific reservation utilities create countervailing incentives. In our model, as in models of common agency and renegotiation, the type specific outside options are endogenously created.

[22] To our knowledge, these countervailing incentives for colluding agents are first discussed by Caillaud and Jehiel [3]. An endogenous incentive reversal of the colluding agents is also utilized by Mookherjee and Tsumagari [27], Pavlov [29], Che and Kim [7], and Celik [5] in the contexts of collusion between productive agents, collusion between bidders and collusion involving an uninformed insurer.

collusion proof output levels within the class of monotonic profiles:

$$\max_{\{x_\theta\}_{\theta\in\Theta}} \sum f(\theta)\left[W(x_\theta) - \theta x_\theta - V_\theta\right] \tag{16}$$

$$\text{s.t. } x_{\underline{\theta}} \geq x_{\tilde{\theta}} \geq x_{\overline{\theta}} \text{ and (15).}$$

As long as the monotonicity constraint is slack for the no-supervision problem, the results we derive regarding the non-monotonic output levels in the Appendix imply that the solution to this problem is also the globally optimal collusion proof output profile. We state this result formally with the following proposition.

Proposition 4 *Suppose the monotonicity constraint is slack for the no-supervision problem, i.e. (3) holds and therefore $x_{\overline{\theta}}^{ns} < x_{\tilde{\theta}}^{ns}$. A solution to problem (16) is the optimal collusion proof output profile.*

4.2 Remarks on the Optimal Collusion Proof Outcome

- Comparison with Collusion Free Supervision

The optimal outcome identified in (15) is an improvement over no-supervision. However, a comparison with the coalitional information rents in (7) reveals that this outcome falls short of achieving the same expected payoff as the optimal collusion free outcome. This is in contrast to a result by Che and Kim [6] discussed in the introduction. Che and Kim [6] show that the collusion of risk neutral parties imposes no cost to the principal in a large class of circumstances. Their result allows for the correlation of information held by the colluding parties as long as a *pairwise identifiability* condition (which they label as $\mathbf{PI}'$) is satisfied by the joint distribution of types. When the supervisor's information is modeled as a partition of the agent's private information, this condition is violated. Therefore, Che and Kim's [6] result does not apply to our environment.[23]

[23] Che and Kim [6] also show that condition $\mathbf{PI}'$ is generic when there are more than two colluding parties, but

- Ex-post Participation Constraints for the Supervisor

The outcome defined by Proposition 3 can leave S with a negative surplus whenever A's type is $\underline{\theta}$. Under the interim participation constraints, this is not a hurdle for implementation. However, if S can walk out of the contract after A's type is revealed or if S is shielded by a limited liability requirement, then the participation constraint (6) must be replaced by the stronger ex-post participation constraints $V_\theta - U_\theta \geq 0$ for θ equals $\widetilde{\theta}$ and $\underline{\theta}$. Under these ex-post participation constraints, the information rent levels in (15) are still the lowest rent levels implementing monotonic output levels provided that the definition for Π is amended as $\Pi = \min \left\{ \frac{f(\underline{\theta})}{f(\underline{\theta})+f(\widetilde{\theta})} \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}, \left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right) \right\}$.

- Stochastic Contracts

In the definitions of both the grand contract and the side contract we have not allowed for lotteries, forcing either contract to be deterministic. At the grand contract level this does not impose a loss for the principal: The principal's direct payoff is concave and the agent's utility is linear in the output levels. Moreover, all players are risk neutral in monetary transfers. Therefore, stochastic grand contracts are dominated by the deterministic ones from the principal's perspective.[24] However, the possibility of stochastic side contracts increases the available collusion opportunities and, therefore, shrinks the implementable set of outcomes further.

Stochastic side contracts would allow S and A to agree on randomizing between different messages they send the grand contract. Capturing the possibility of stochastic side contracts requires amending the definition of the side contract by extending the range of the function

non-generic when there are two. Therefore, their result is not likely to apply to collusion between a supervisor and an agent under an alternative information structure.

[24] Strausz [32] presents an example of beneficial stochastic contracts when the agent's utility is not linear in output.

which determines the messages by S and A:

$$m : M'_A \rightarrow \Delta \left(M_S \times M_A \right).$$

By offering a stochastic side contract, S is able to induce manipulations randomizing between different misreports for a given type of A such that $\hat{c} : \Theta \rightarrow \Delta\Theta$. Extending the set of manipulations available for the colluding parties may jeopardize the implementability of the outcome identified by (15). To see this, suppose $\Pi = \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} < \left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right)$, implying the upward incentive constraint $\mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right)$ is slack. Even with stochastic side contracts, it is not possible for the coalition to achieve complete coalitional efficiency by misreporting type $\underline{\theta}$ as type $\widetilde{\theta}$ with probability one. However, S and A can increase the coalitional gain by implementing this misreport with a non-degenerate probability. When this probability is small enough, $\mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right)$ is satisfied even without increasing the rent left for type $\widetilde{\theta}$. Therefore S extracts the entire coalitional gain. Consequently, the outcome in (15) is not implementable. In contrast, when $\Pi = \left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right) \leq \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}$ and $\mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right)$ is binding, randomizations over misreports do not improve S's rent extraction. Any manipulation that assigns a positive probability to misreporting type $\underline{\theta}$ as $\widetilde{\theta}$ requires leaving a bribe to type $\widetilde{\theta}$ that would cancel any benefit S may receive from the misreporting. Therefore, if $\Pi = \left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right) \leq \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}$ then the information rent levels in (15) are collusion proof even when stochastic side contracts are available.[25]

- Distortions and Social Welfare

[25] In contrast to the deterministic side contracts, when stochastic side contracts are allowed, improvements over the no-supervision optimal outcome demand for a *non-marginal* increase in the reservation utility of the most efficient type $\underline{\theta}$. As long as $\left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right) \leq \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}$, the information rent levels in (15) are collusion proof under the stochastic side contracts for $\Pi = \left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right)$. However, these information rent levels are not collusion proof for a smaller value of Π. This non-marginal increase in the reservation utility is required to induce countervailing incentives for the agent by making his downward incentive constraint $\mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right)$ binding while colluding with the supervisor.

The relevance result established that P's payoff under collusion proof supervision is strictly higher than its no-supervision optimal level. Another question of interest is the overall effect of supervision on the *social welfare.* Since utility functions of all three players are quasilinear in money, the extent of output distortions is a good measure of welfare. Let $\{x_\theta^*\}_{\theta\in\Theta}$ be the solution to the maximization problem (16). Also define $\left\{x_\theta^{fb}\right\}_{\theta\in\Theta}$ as the profile of "first best" output levels, where $W'(x_\theta^{fb}) = \theta$. A comparison of $\{x_\theta^*\}_{\theta\in\Theta}$ with the no-supervision optimal output levels $\{x_\theta^{ns}\}_{\theta\in\Theta}$ is sufficient to identify the social welfare effects of supervision. To simplify the analysis, we consider parameters for which the monotonicity constraint in problem (16) is slack.[26]

We will identify the welfare effects of supervision under two polar cases. First, consider the case, where $\underline{\theta}$ is small enough so that $\Pi = \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}^* < \left(\widetilde{\theta} - \underline{\theta}\right)\left(x_{\underline{\theta}}^* - x_{\widetilde{\theta}}^*\right)$.[27] In this case, an increase in the output level of the least efficient type leads to an increase in the value of Π. Therefore, as can be seen from the information rent levels in (15), an increase in $x_{\overline{\theta}}$ is not as costly for P as it would have been under no-supervision. Accordingly, at the solution of problem (16), the output level for the least efficient type is distorted less than it would have been at the solution to the no-supervision problem ($x_{\overline{\theta}}^{ns} < x_{\overline{\theta}}^* < x_{\overline{\theta}}^{fb}$). Since the output levels for types $\underline{\theta}$ and $\widetilde{\theta}$ would be the same as their no-supervision levels, the optimal collusion proof outcome creates higher social welfare than does the optimal no-supervision outcome.

We can also construct a case where the welfare implication of supervision is reversed: Sup-

[26] The inequality below is sufficient for the monotonicity constraint to be slack:

$$f\left(\widetilde{\theta}\right)\left[f\left(\overline{\theta}\right) + f\left(\underline{\theta}\right)\right]\left(\overline{\theta} - \widetilde{\theta}\right) \geq f\left(\overline{\theta}\right)\left[f\left(\widetilde{\theta}\right) + f\left(\underline{\theta}\right)\right]\left(\widetilde{\theta} - \underline{\theta}\right).$$

This condition is stronger than condition (3), which ensures that the monotonicity constraint is slack for the no-supervision problem.

[27] The upper bound on $\underline{\theta}$ satisfying this inequality depends on the function $W(\cdot)$. However, since $\lim_{\underline{\theta}\to 0} x_{\underline{\theta}}^* = \infty$, there exists a value for $\underline{\theta}$ that satisfies the inequality.

pose $\overline{\theta}$ is large enough so that $\Pi = \left(\widetilde{\theta} - \underline{\theta}\right)\left(x^*_{\underline{\theta}} - x^*_{\widetilde{\theta}}\right) < \left(\overline{\theta} - \widetilde{\theta}\right) x^*_{\widetilde{\theta}}$.[28] In this case, the value of Π is increasing in the difference between the output levels of types $\underline{\theta}$ and $\widetilde{\theta}$. Accordingly, P finds it profitable to increase $x_{\underline{\theta}}$ and reduce $x_{\widetilde{\theta}}$ from their no-supervision levels ($x_{\widetilde{\theta}} < x^{ns}_{\widetilde{\theta}} < x^{fb}_{\widetilde{\theta}}$ and $x^*_{\underline{\theta}} > x^{ns}_{\underline{\theta}} = x^{fb}_{\underline{\theta}}$). This time, both output levels are distorted further from their first best levels relative to their no-supervision levels. Since the output level for type $\overline{\theta}$ is the same as its no-supervision level, we conclude that the optimal supervision outcome creates lower social welfare than does the optimal no-supervision outcome.[29,30]

- Accuracy of the Supervisory Information

Faure-Grimaud, Laffont, and Martimort [9] show that the principal's optimal collusion proof payoff approaches its no-supervision level when a supervisor's information approaches either perfect information of A's type or the complete lack of information. Since the information structure we adopt for the supervisor is between these two extremes by design, we cannot conduct a similar comparative statics analysis here. However, we show two exercises that provide some insight into the effects of the accuracy of S's information. First, consider the case where the cost levels of the two types that S cannot distinguish tend to each other, i.e. $\left(\widetilde{\theta} - \underline{\theta}\right) \to 0$. Notice that in the limit, our model converges to a setup with a supervisor who is perfectly informed of the cost level of the agent. In this case, the expected payoff of the principal approaches its optimal no-supervision level since $\Pi \to 0$. Similarly, as we reduce $f\left(\overline{\theta}\right)$, the model converges to uninformative supervision with types $\widetilde{\theta}$ and $\underline{\theta}$ only. As $f\left(\overline{\theta}\right)$

[28] Since $\lim_{\underline{\theta} \to \widetilde{\theta}} \left(x^*_{\underline{\theta}} - x^*_{\widetilde{\theta}}\right) = 0$, there exists a value for $\underline{\theta}$ that satisfies this inequality.

[29] The possibility that supervision may decrease social welfare is another differentiating aspect of the analysis of this paper. In models of exogenous transaction costs between colluders (surveyed by Tirole [33]), or in models using asymmetric information to justify the transaction cost approach (such as Faure-Grimaud, Laffont, and Martimort [8,9]), supervision always increases overall efficiency.

[30] For the intermediate values of $\underline{\theta}$, $\Pi = \left(\widetilde{\theta} - \underline{\theta}\right)\left(x^*_{\underline{\theta}} - x^*_{\widetilde{\theta}}\right) = \left(\overline{\theta} - \widetilde{\theta}\right) x^*_{\widetilde{\theta}}$, and the social welfare effect of supervision is ambiguous.

tends to zero, the optimal $x_{\overline{\theta}}$ that solves (16) tends to zero as well. Consequently, $\Pi \to 0$ and the expected payoff of the principal approaches its optimal no-supervision level.[31]

- A Model with Finitely Many Cost Levels

Celik [4] extends the analysis to a model where the agent has n possible cost levels and the supervisor's information is represented as an arbitrary connected partition of these cost levels. As long as the monotonicity constraints are slack for the no-supervision problem and the supervisor is partly informed[32] on the cost level, collusion proof implementation dominates no-supervision, which in turn dominates delegation to the supervisor. As in the three cost level model described here, the connected partition structure and the slackness of the monotonicity constraints are crucial in identifying that the upward incentive compatibility constraints are the only relevant constraints to be violated for beneficial supervision.

5 Binding Monotonicity Constraint

Suppose the monotonicity constraint is slack for the no-supervision problem. In this case, our results indicate that the optimal collusion proof outcome induces monotonic output levels. These output levels constitute a solution to the maximization problem (16). The optimal collusion proof outcome is strictly better than the no-supervision optimal outcome. However, the optimal collusion proof outcome is not available to P if he surrenders his control of the outside option of collusion by delegating to S. In fact all the delegation proof outcomes are strictly dominated by the no-supervision optimal outcome.

[31] Notice that as either $\left(\widetilde{\theta} - \underline{\theta}\right)$ or $f\left(\overline{\theta}\right)$ tend to zero, inequality (3) can be satisfied so that the solution to (16) gives the optimal collusion proof outcome for either exercise.

[32] More specifically, we need the supervisor's information structure to contain a partition cell, which excludes the highest cost level and includes more than one cost level.

Another question of interest is how these results change if the monotonicity constraint is binding for the no-supervision problem. We show that non-monotonic output levels are collusion proof whenever monotonicity is a binding constraint for the no-supervision problem. Moreover, delegation achieves the lowest coalitional information rent levels implementing a non-monotonic output profile.

Proposition 5 *Suppose the monotonicity constraint is binding for the no-supervision problem, i.e. (3) does not hold and therefore* $x_{\underline{\theta}}^{ns} = x_{\widetilde{\theta}}^{ns}$*. Then any non-monotonic output profile such that* $x_{\widetilde{\theta}} < x_{\overline{\theta}} \leq x_{\underline{\theta}}$ *is collusion proof. The lowest coalitional information rent levels and the corresponding utility levels for A that make the collusion proof implementation of such a non-monotonic output profile possible are given as*

$$\begin{aligned} &V_{\overline{\theta}} = 0, && U_{\overline{\theta}} = 0, \\ &V_{\widetilde{\theta}} = \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} - \frac{f(\underline{\theta})}{f(\widetilde{\theta})} \left(\widetilde{\theta} - \underline{\theta}\right) \left(x_{\overline{\theta}} - x_{\widetilde{\theta}}\right), && U_{\widetilde{\theta}} = 0, \\ &V_{\underline{\theta}} = \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} + \left(\widetilde{\theta} - \underline{\theta}\right) x_{\overline{\theta}}, && U_{\underline{\theta}} = \left(\widetilde{\theta} - \underline{\theta}\right) x_{\widetilde{\theta}}. \end{aligned} \tag{17}$$

Moreover, the resulting outcome is delegation proof.

To compare the delegation proof outcome above with the optimal no-supervision outcome, we utilize functions $V_{\theta}^{o}(\cdot)$ defined in (1). These functions yield the no-supervision information rent levels for the monotonic output profiles. However, they are also well defined for non-monotonic output profiles. In fact, the expected coalitional information rent level $f\left(\overline{\theta}\right) V_{\overline{\theta}} + f\left(\widetilde{\theta}\right) V_{\widetilde{\theta}} + f\left(\underline{\theta}\right) V_{\underline{\theta}}$ for the outcome above is equal to $f\left(\overline{\theta}\right) V_{\overline{\theta}}^{o}(\cdot) + f\left(\widetilde{\theta}\right) V_{\widetilde{\theta}}^{o}(\cdot) + f\left(\underline{\theta}\right) V_{\underline{\theta}}^{o}(\cdot)$ evaluated for the same non-monotonic output profile. This leaves P with the same objective function as in the no-supervision problem (2). Suppose the solution to the no-supervision problem is non-monotonic when the monotonicity constraint is ignored. Then delegation is an improvement over no-supervision since P is not constrained by monotonicity under delegation.

As is the case for the optimal collusion proof outcome derived in Proposition 3, the outcome above is coalitionally inefficient: A and S could have increased their joint rent by

$\frac{f(\underline{\theta})}{f(\widetilde{\theta})}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\overline{\theta}}-x_{\widetilde{\theta}}\right)$ if they misreported agent type $\widetilde{\theta}$ as $\overline{\theta}$. However, as a result of this misreport, type $\widetilde{\theta}$ agent would increase his production level to $x_{\overline{\theta}}$ from $x_{\widetilde{\theta}}$. The increased production level for this relatively high cost type would require increasing the information rent of the agent with type $\underline{\theta}$. This increase in information rent of type $\underline{\theta}$ would cancel out the benefit S would have received from misreporting type $\widetilde{\theta}$.

Suppose S observes that the type of the agent is either $\widetilde{\theta}$ or $\underline{\theta}$. The relevant constraints for S's side contract selection problem are the individual rationality constraint of the relatively inefficient type **AIR** $\left(\widetilde{\theta}\right)$, and the incentive compatibility constraint of the efficient type **AIC** $\left(\widetilde{\theta}|\underline{\theta}\right)$. Notice that this pattern of constraints is typical under the uniform reservation utilities.

Let us compare the coalitional inefficiencies arising under the two collusion proof outcomes we defined in Propositions 3 and 5. The former outcome, which induces a monotonic output profile, is depicted by (15). Here, the coalition fails to *reduce* the output level to $x_{\widetilde{\theta}}$ for type $\underline{\theta}$, even though this would have increased the coalitional rent. The inefficiency in question is "overproduction." This is contrary to the typical "underproduction" observed in design problems with uniform reservation utilities, such as the no-supervision problem we discussed as a benchmark. Sustaining overproduction for the coalition requires the principal's active manipulation of the outside option.

In contrast, the collusion proof outcome depicted by (17) is coalitionally inefficient because the coalition fails to *increase* the output level to $x_{\overline{\theta}}$ for type $\widetilde{\theta}$. No incentive reversal is required to generate this underproduction as an equilibrium phenomenon for the supervisor - agent interaction. Accordingly, there is no need for P to manipulate the outside option for collusion. For this reason, delegation is a successful tool in the implementation of this outcome.

The discussion above provides an opportunity to revisit a result by Faure-Grimaud, Laffont, and Martimort [9], which establishes delegation as the optimal organizational response to collusion. This result is derived for a model that is characterized by a unit production cost, which can assume two possible values (low cost or high cost), and a supervisory signal, which can also assume two values (low signal or high signal). The realization of the signal is positively correlated with the realization of the cost. The realized cost is observed by the productive agent. The realized signal is observed by both the agent and his supervisor. This environment induces four different states of nature, each involving a different cost - signal pair.

The optimal collusion proof outcome identified by Faure-Grimaud, Laffont, and Martimort [9] is an improvement over the optimal no-supervision outcome. In fact, this collusion proof outcome achieves the same expected payoff as the collusion-free supervision payoff when the supervisor is risk neutral. The source of this improvement is the principal's ability to implement different output and coalitional rent levels for two states with the same cost level. The output level for the high cost - low signal state is lower than the high cost - high signal state. Moreover, when the realized state is high cost - low signal, the supervisor is severely punished for the mismatch. In this case, it would be a coalitional improvement for the S - A pair to behave as though the state is high cost - high signal. For this improvement not to materialize, the side contract must fail to *increase* the output to its coalitionally efficient level. This underproduction feature of the required collusion failure is compatible with the performance of the side contract under delegation.

Before closing this section, we comment on an implication of Proposition 5 on the optimality of delegation. Suppose the no-supervision monotonicity constraint is binding. Then Proposition 5 identifies the least expensive way of implementing a *non-monotonic* output profile in a collusion proof manner. However, it should be noted that the proposition does not imply that the optimal collusion proof output profile must be non-monotonic. The solution identified in (15) with monotonic output levels can yield a higher expected payoff for the principal than any non-monotonic profile. Whether the optimal non-monotonic profile is superior to the optimal monotonic profile depends on the specification of P's direct utility from production $W(\cdot)$ as well as the values of the other parameters. Since we do not want to assume any further structure on $W(\cdot)$ we leave this question open.

6 Conclusion

In this paper, we provide a justification for third party supervision even when this third party can collude with the supervised agent. We model the supervisor's information as a connected partition of the agent's type space, and we model collusion as a side contract that is offered by the supervisor after the grand contract is announced by the principal. The outside option for this side contract is the non-cooperative play of the game that is induced by the grand contract. Therefore, the principal can affect the type dependent opportunity cost of collusion through his choice of the grand contract. We show that the principal can increase his payoff with the introduction of the supervisor. Although the supervisor - agent coalition would be better off by collectively misrepresenting certain states of nature, the principal can rule out such behavior with the appropriate manipulation of the outside option of collusion.

In our framework, delegation corresponds to a special class of grand contracts which are not responsive to the agent's reporting. Under delegation, the outside option for collusion is the shut down of production and, therefore, the principal loses his power to manipulate the type dependent opportunity cost of collusion. As long as the monotonicity constraint of the no-supervision implementation problem is slack, delegation performs worse than the absence of supervision for the principal. Nevertheless, if the no-supervision monotonicity constraint is binding, then delegation is valuable since it can sustain implementation of non-monotonic output levels.

7 Appendix

7.1 Implementing Non-monotonic Output Levels

In this part of the Appendix, we study the implementation of non-monotonic output levels when an informed supervisor is present. Under the no-supervision benchmark, the incentive constraints of the agent imply that the implementable output levels should be ordered as $x_{\overline{\theta}} \leq x_{\widetilde{\theta}} \leq x_{\underline{\theta}}$. When there is a supervisor who can distinguish type $\overline{\theta}$ from the other types, the only monotonicity constraint derived from the agent's incentives is $x_{\widetilde{\theta}} \leq x_{\underline{\theta}}$. This leaves us with two non-monotonic orderings of output levels which are possibly implementable: $x_{\widetilde{\theta}} \leq x_{\underline{\theta}} < x_{\overline{\theta}}$ and $x_{\widetilde{\theta}} < x_{\overline{\theta}} \leq x_{\underline{\theta}}$. With Lemma 1, we show that the former ordering contradicts the supervisor's surplus maximization. And with Lemma 2, we establish a lower bound on the expected coalitional information rent level achieving the latter ordering.

Lemma 1 *There is no collusion proof outcome inducing output levels such that $x_{\underline{\theta}} < x_{\overline{\theta}}$.*

Proof. Suppose outcome $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ is collusion proof and $x_{\underline{\theta}} < x_{\overline{\theta}}$. Then it satisfies the collusion feasibility conditions stated in (10) and (11). First, take the maximization problem in (10). Consider a manipulation where type $\overline{\theta}$ is misreported as type $\underline{\theta}$ such that $\hat{c}\left(\overline{\theta}\right) = \underline{\theta}$ and $\hat{u}\left(\overline{\theta}\right) = U_{\overline{\theta}}$. Since there is no change in the utility level of type $\overline{\theta}$, constraint $\mathbf{AIR}\left(\overline{\theta}\right)$ is satisfied. For this deviation not to improve the objective function it must be that

$$V_{\overline{\theta}} \geq V_{\underline{\theta}} - \left(\overline{\theta} - \underline{\theta}\right) x_{\underline{\theta}}. \tag{18}$$

Now take the maximization problem in (11). Consider a manipulation where type $\underline{\theta}$ is misreported as type $\overline{\theta}$ such that $\hat{c}\left(\widetilde{\theta}\right) = \widetilde{\theta}, \hat{u}\left(\widetilde{\theta}\right) = U_{\widetilde{\theta}}, \hat{c}\left(\underline{\theta}\right) = \overline{\theta}$, and $\hat{u}\left(\underline{\theta}\right) = U_{\underline{\theta}}$. Since there is no change in the utility levels of the agent types, this deviation satisfies **AIR** constraints. Moreover, **AIC** constraints are also satisfied since the output level of the efficient type is increased. For this deviation not to improve the objective function it must be that

$$V_{\underline{\theta}} \geq V_{\overline{\theta}} + \left(\overline{\theta} - \underline{\theta}\right) x_{\overline{\theta}}. \tag{19}$$

These two inequalities imply $x_{\underline{\theta}} \geq x_{\overline{\theta}}$, which is a contradiction to the supposition that $x_{\underline{\theta}} < x_{\overline{\theta}}$.

■

Lemma 2 *Suppose outcome $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ is collusion proof. If $x_{\widetilde{\theta}} < x_{\overline{\theta}}$ then*

$$f\left(\overline{\theta}\right)V_{\overline{\theta}}+f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}} \geq f\left(\overline{\theta}\right)V_{\overline{\theta}}^{o}\left(\{x_\theta\}_{\theta\in\Theta}\right)+f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}^{o}\left(\{x_\theta\}_{\theta\in\Theta}\right)+f\left(\underline{\theta}\right)V_{\underline{\theta}}^{o}\left(\{x_\theta\}_{\theta\in\Theta}\right), \tag{20}$$

where functions $V_\theta^o(\cdot)$ are defined by (1).

Proof. $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ satisfies the collusion feasibility conditions stated in (10) and (11). First, take the maximization problem in (10). Consider a manipulation where type $\overline{\theta}$ is misreported as type $\widetilde{\theta}$ such that $\hat{c}\left(\overline{\theta}\right)=\widetilde{\theta}$ and $\hat{u}\left(\overline{\theta}\right)=U_{\overline{\theta}}$. Since there is no change in the utility level of type $\overline{\theta}$, constraint **AIR** $\left(\overline{\theta}\right)$ is satisfied. For this deviation not to improve the objective function it must be that

$$V_{\overline{\theta}} \geq V_{\widetilde{\theta}} - \left(\overline{\theta}-\widetilde{\theta}\right)x_{\widetilde{\theta}}. \tag{21}$$

Now take the maximization problem in (11). Consider a manipulation where type $\widetilde{\theta}$ is misreported as type $\overline{\theta}$ such that $\hat{c}\left(\widetilde{\theta}\right)=\overline{\theta}$, $\hat{u}\left(\widetilde{\theta}\right)=U_{\widetilde{\theta}}$, $\hat{c}\left(\underline{\theta}\right)=\underline{\theta}$, and $\hat{u}\left(\underline{\theta}\right)=\max\left\{U_{\underline{\theta}}, U_{\widetilde{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\overline{\theta}}\right\}$. Note that there is no change in the utility level of type $\widetilde{\theta}$ and constraints **AIC** $\left(\widetilde{\theta}|\underline{\theta}\right)$ and **AIR** $\left(\underline{\theta}\right)$ are satisfied by the construction of $\hat{u}\left(\underline{\theta}\right)$. For this deviation not to improve the objective function it must be that

$$V_{\widetilde{\theta}} \geq V_{\overline{\theta}} + \left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}} - \frac{f\left(\underline{\theta}\right)}{f\left(\widetilde{\theta}\right)}\max\left\{U_{\widetilde{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\overline{\theta}}-U_{\underline{\theta}}, 0\right\}. \tag{22}$$

When the two inequalities above are merged,

$$f\left(\widetilde{\theta}\right)\left(\overline{\theta}-\widetilde{\theta}\right)\left(x_{\widetilde{\theta}}-x_{\overline{\theta}}\right) \geq -f\left(\underline{\theta}\right)\max\left\{U_{\widetilde{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\overline{\theta}}-U_{\underline{\theta}}, 0\right\}. \tag{23}$$

An implication of (23) is

$$U_{\widetilde{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\overline{\theta}}-U_{\underline{\theta}} > 0. \tag{24}$$

Otherwise, it follows from (23) that $x_{\widetilde{\theta}} \geq x_{\overline{\theta}}$.

Now consider another manipulation for (11) where types $\widetilde{\theta}$ and $\underline{\theta}$ are misreported as type $\overline{\theta}$ such that $\hat{c}\left(\widetilde{\theta}\right) = \hat{c}\left(\underline{\theta}\right) = \overline{\theta}, \hat{u}\left(\widetilde{\theta}\right) = U_{\widetilde{\theta}}$, and $\hat{u}\left(\underline{\theta}\right) = U_{\widetilde{\theta}} + \left(\widetilde{\theta} - \underline{\theta}\right) x_{\overline{\theta}}$. Since there is pooling of the two possible types, **AIC** constraints are trivially satisfied. Moreover, constraint **AIR** $\left(\underline{\theta}\right)$ follows from (24). For this deviation not to improve the objective function it must be that

$$f\left(\widetilde{\theta}\right)\left(V_{\widetilde{\theta}} - U_{\widetilde{\theta}}\right) + f\left(\underline{\theta}\right)\left(V_{\underline{\theta}} - U_{\underline{\theta}}\right) \geq f\left(\widetilde{\theta}\right)\left(V_{\overline{\theta}} + \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} - U_{\widetilde{\theta}}\right) + f\left(\underline{\theta}\right)\left(V_{\overline{\theta}} + \left(\overline{\theta} - \underline{\theta}\right) x_{\overline{\theta}} - U_{\widetilde{\theta}} - \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}\right). \tag{25}$$

Moving terms $U_{\widetilde{\theta}}$ and $U_{\underline{\theta}}$ to the left hand side of the inequality yields

$$f\left(\widetilde{\theta}\right) V_{\widetilde{\theta}} + f\left(\underline{\theta}\right) V_{\underline{\theta}} \geq f\left(\widetilde{\theta}\right)\left(V_{\overline{\theta}} + \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}\right) + f\left(\underline{\theta}\right)\left(V_{\overline{\theta}} + \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} + U_{\underline{\theta}} - U_{\widetilde{\theta}}\right). \tag{26}$$

Since constraint **AIC** $\left(\widetilde{\theta}|\underline{\theta}\right)$ is satisfied for the outcome $\{x_\theta, V_\theta, U_\theta\}_{\theta \in \Theta}$ we can rewrite this inequality as

$$f\left(\widetilde{\theta}\right) V_{\widetilde{\theta}} + f\left(\underline{\theta}\right) V_{\underline{\theta}} \geq f\left(\widetilde{\theta}\right)\left(V_{\overline{\theta}} + \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}\right) + f\left(\underline{\theta}\right)\left(V_{\overline{\theta}} + \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} + \left(\widetilde{\theta} - \underline{\theta}\right) x_{\widetilde{\theta}}\right). \tag{27}$$

Participation constraints for the agent (4) and the supervisor (5) imply that $V_{\overline{\theta}} \geq 0$, which completes the proof of the lemma. ■

7.2 Proof for Proposition 1

Participation constraints for the agent (4) and the supervisor (5) imply that $V_{\overline{\theta}} \geq 0$. From delegation feasibility, we also know that $\{x_\theta, V_\theta, U_\theta\}_{\theta \in \Theta}$ induces a solution to the maximization problem in (9). Now consider a manipulation where type $\widetilde{\theta}$ is misreported as type $\overline{\theta}$ and the information rent left for type $\underline{\theta}$ is reduced accordingly: $\hat{c}\left(\widetilde{\theta}\right) = \overline{\theta}, \hat{u}\left(\widetilde{\theta}\right) = U_{\widetilde{\theta}} = 0, \hat{c}\left(\underline{\theta}\right) = \underline{\theta}$, and $\hat{u}\left(\underline{\theta}\right) = \left(\widetilde{\theta} - \underline{\theta}\right) x_{\overline{\theta}}$. This deviation satisfies $\mathbf{d} - \mathbf{AIR}$ and **AIC** constraints. For this deviation not to improve the objective function in (9), it must be that

$$f\left(\widetilde{\theta}\right) V_{\widetilde{\theta}} + f\left(\underline{\theta}\right)\left(V_{\underline{\theta}} - \left(\widetilde{\theta} - \underline{\theta}\right) x_{\widetilde{\theta}}\right) \geq f\left(\widetilde{\theta}\right)\left(V_{\overline{\theta}} + \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}}\right) + f\left(\underline{\theta}\right)\left(V_{\underline{\theta}} - \left(\widetilde{\theta} - \underline{\theta}\right) x_{\overline{\theta}}\right). \tag{28}$$

After rearranging the inequality we get

$$V_{\widetilde{\theta}} \geq V_{\overline{\theta}} + \left(\overline{\theta} - \widetilde{\theta}\right) x_{\overline{\theta}} + \frac{f\left(\underline{\theta}\right)}{f\left(\widetilde{\theta}\right)}\left(\widetilde{\theta} - \underline{\theta}\right)\left(x_{\widetilde{\theta}} - x_{\overline{\theta}}\right). \tag{29}$$

Notice that inequality (29) is a stronger condition than the no-supervision **IC** constraint. This rules out type $\widetilde{\theta}$'s imitation of type $\overline{\theta}$, since the second term on the right hand side is positive for $x_{\widetilde{\theta}} > x_{\overline{\theta}}$. This difference between the no-supervision and the delegation constraints is crucial for establishing our result. A similar constraint to inequality (29) arises in Melumad, Mookherjee, and Reichelstein's [25] paper. However, in their model, the additional term $\frac{f(\underline{\theta})}{f(\widetilde{\theta})}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\widetilde{\theta}}-x_{\overline{\theta}}\right)$ can be taxed away from the "delegate" by setting $V_{\widetilde{\theta}}$ to be negative without causing any positive cost of delegation. In contrast, $V_{\widetilde{\theta}}$ cannot be a negative number in our model. This implies

$$V_{\widetilde{\theta}} \geq \left(\overline{\theta}-\widetilde{\theta}\right) x_{\overline{\theta}} + \frac{f\left(\underline{\theta}\right)}{f\left(\widetilde{\theta}\right)}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\widetilde{\theta}}-x_{\overline{\theta}}\right). \tag{30}$$

Now consider another manipulation where type $\underline{\theta}$ is misreported as type $\widetilde{\theta}$ such that $\hat{c}\left(\widetilde{\theta}\right) = \hat{c}\left(\underline{\theta}\right) = \widetilde{\theta}$, $\hat{u}\left(\widetilde{\theta}\right) = U_{\widetilde{\theta}} = 0$, and $\hat{u}\left(\underline{\theta}\right) = U_{\underline{\theta}} = \left(\widetilde{\theta}-\underline{\theta}\right) x_{\widetilde{\theta}}$. Since there is no change in the utility levels of the agent types, this deviation satisfies $\mathbf{d}-\mathbf{AIR}$ and **AIC** constraints. For this deviation not to improve the objective function in (9), it must be that

$$f\left(\widetilde{\theta}\right) V_{\widetilde{\theta}} + f\left(\underline{\theta}\right)\left(V_{\underline{\theta}} - \left(\widetilde{\theta}-\underline{\theta}\right) x_{\widetilde{\theta}}\right) \geq f\left(\widetilde{\theta}\right) V_{\widetilde{\theta}} + f\left(\underline{\theta}\right) V_{\widetilde{\theta}}. \tag{31}$$

After rearranging the inequality we get

$$V_{\underline{\theta}} \geq V_{\widetilde{\theta}} + \left(\widetilde{\theta}-\underline{\theta}\right) x_{\widetilde{\theta}}. \tag{32}$$

When we substitute in inequality (30), this yields

$$V_{\underline{\theta}} \geq \left(\overline{\theta}-\widetilde{\theta}\right) x_{\overline{\theta}} + \left(\widetilde{\theta}-\underline{\theta}\right) x_{\widetilde{\theta}} + \frac{f\left(\underline{\theta}\right)}{f\left(\widetilde{\theta}\right)}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\widetilde{\theta}}-x_{\overline{\theta}}\right), \tag{33}$$

completing the proof.

7.3 Proof for Proposition 2

Proposition 1 implies that any (strictly) monotonic output profile is delegation proof with (strictly) higher coalitional information rents than the no-supervision information rents. Moreover, when the no-supervision monotonicity constraint is slack, Lemma 2 implies that there is no collusion proof outcome with a non-monotonic output profile undominated by the no-supervision optimal outcome. The proof follows from the fact that any delegation proof outcome is also collusion proof.

7.4 Proof for Proposition 3

In the first part of the proof, we show that the information rent levels identified in (15) constitute lower bounds on collusion proof information rent levels implementing a monotonic output profile. In the second part, we show that this information rent profile is indeed collusion proof with any arbitrary monotonic output profile as long as Π is a non-negative number smaller than both $\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}$ and $\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\underline{\theta}}-x_{\widetilde{\theta}}\right)$.

Part 1

Participation constraints for the agent (4) and the supervisor (5) imply that $V_{\overline{\theta}}\geq 0$. Any collusion proof outcome $\{x_\theta, V_\theta, U_\theta\}_{\theta\in\Theta}$ is collusion feasible and therefore, induces a solution to problem (11). Now consider a manipulation where type $\widetilde{\theta}$ is misreported as type $\overline{\theta}$ such that $\hat{c}\left(\widetilde{\theta}\right)=\overline{\theta}, \hat{c}\left(\underline{\theta}\right)=\underline{\theta}$, $\hat{u}\left(\widetilde{\theta}\right)=U_{\widetilde{\theta}}$, and $\hat{u}\left(\underline{\theta}\right)=U_{\underline{\theta}}$. Since there is no change in the utility levels of either type, **AIR** constraints are satisfied. Moreover, **AIC** constraints are satisfied since this manipulation decreases the production level of the inefficient type further. For this deviation not to improve the objective function in (11), it must be that

$$f\left(\widetilde{\theta}\right)\left(V_{\widetilde{\theta}}-U_{\widetilde{\theta}}\right)+f\left(\underline{\theta}\right)\left(V_{\underline{\theta}}-U_{\underline{\theta}}\right)\geq f\left(\widetilde{\theta}\right)\left(V_{\overline{\theta}}+\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}-U_{\widetilde{\theta}}\right)+f\left(\underline{\theta}\right)\left(V_{\underline{\theta}}-U_{\underline{\theta}}\right). \quad (34)$$

After rearranging the inequality and using $V_{\overline{\theta}}\geq 0$, we get

$$V_{\widetilde{\theta}}\geq\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}. \quad (35)$$

Now consider another manipulation, where type $\underline{\theta}$ is pooled with type $\widetilde{\theta}$ at production level $x_{\widetilde{\theta}}$, such that $\hat{c}\left(\widetilde{\theta}\right)=\hat{c}\left(\underline{\theta}\right)=\widetilde{\theta}$, $\hat{u}\left(\widetilde{\theta}\right)=U_{\underline{\theta}}-\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}$, and $\hat{u}\left(\underline{\theta}\right)=U_{\underline{\theta}}$. Since this is a pooling contract, **AIC** constraints are trivially satisfied. For this deviation not to improve the objective function in (11), it must be that

$$f\left(\widetilde{\theta}\right)\left(V_{\widetilde{\theta}}-U_{\widetilde{\theta}}\right)+f\left(\underline{\theta}\right)\left(V_{\underline{\theta}}-U_{\underline{\theta}}\right)\geq\left[f\left(\widetilde{\theta}\right)+f\left(\underline{\theta}\right)\right]\left[V_{\widetilde{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}-U_{\underline{\theta}}\right]. \tag{36}$$

After rearranging the inequality we get

$$f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}}\geq\left[f\left(\widetilde{\theta}\right)+f\left(\underline{\theta}\right)\right]\left[\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}\right]-f\left(\widetilde{\theta}\right)\left(U_{\underline{\theta}}-U_{\widetilde{\theta}}\right). \tag{37}$$

We will derive two different bounds on the term $\left(U_{\underline{\theta}}-U_{\widetilde{\theta}}\right)$ on the right hand side of this inequality. The first of these is the constraint $\mathbf{AIC}\left(\underline{\theta}|\widetilde{\theta}\right)$: $U_{\underline{\theta}}-U_{\widetilde{\theta}}\leq\left(\overline{\theta}-\widetilde{\theta}\right)x_{\underline{\theta}}$. When we substitute this in (37), we get

$$f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}}\geq f\left(\widetilde{\theta}\right)\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+f\left(\underline{\theta}\right)\left[\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}\right]-f\left(\widetilde{\theta}\right)\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\underline{\theta}}-x_{\widetilde{\theta}}\right). \tag{38}$$

The second bound on $U_{\underline{\theta}}-U_{\widetilde{\theta}}$ is a consequence of the participation constraints $U_{\widetilde{\theta}}\geq 0$ and (6): $U_{\underline{\theta}}-U_{\widetilde{\theta}}\leq V_{\underline{\theta}}+\frac{f(\widetilde{\theta})}{f(\underline{\theta})}V_{\widetilde{\theta}}$. When we substitute this second bound in (37), we get

$$f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}}\geq\left[f\left(\widetilde{\theta}\right)+f\left(\underline{\theta}\right)\right]\left[\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}\right]-\frac{f\left(\widetilde{\theta}\right)}{f\left(\underline{\theta}\right)}\left[f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}}\right]. \tag{39}$$

After transferring the terms containing $V_{\widetilde{\theta}}$ and $V_{\underline{\theta}}$ to the left hand side, this yields

$$\frac{\left[f\left(\widetilde{\theta}\right)+f\left(\underline{\theta}\right)\right]}{f\left(\underline{\theta}\right)}\left[f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}}\right]\geq\left[f\left(\widetilde{\theta}\right)+f\left(\underline{\theta}\right)\right]\left[\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}\right]. \tag{40}$$

By dividing both sides of the inequality by $\frac{\left[f(\widetilde{\theta})+f(\underline{\theta})\right]}{f(\underline{\theta})}$, we get

$$f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}}\geq f\left(\underline{\theta}\right)\left[\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}\right]. \tag{41}$$

We merge (38) and (41) as

$$f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}}\geq f\left(\widetilde{\theta}\right)\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+f\left(\underline{\theta}\right)\left[\left(\overline{\theta}-\widetilde{\theta}\right)x_{\overline{\theta}}+\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}\right]-f\left(\widetilde{\theta}\right)\Pi, \tag{42}$$

where Π is defined in the proposition. This completes the construction of the lower bound on the collusion proof information rent levels implementing a monotonic output profile.

Part 2

The participation constraints are satisfied by the outcome identified in (15). Next, we check for collusion feasibility. We start by observing that **AIC** and **AIR** constraints are satisfied for problems (10) and (11). Note that **AIR** constraints are tautological and **AIC** constraints are satisfied since $\left(\widetilde{\theta}-\underline{\theta}\right)x_{\underline{\theta}} \geq U_{\underline{\theta}}-U_{\widetilde{\theta}} \geq \left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}$. For either (10) or (11) to fail, there must exist an alternative outcome satisfying the constraints and yielding a higher value of the objective function. A necessary condition for this is the existence of a type θ with a misreport $\hat{c}(\theta)$ that gives a higher coalitional rent:

$$V_{\hat{c}(\theta)}+(\hat{c}(\theta)-\theta)\,x_{\hat{c}(\theta)}>V_{\theta}. \tag{43}$$

The only possible coalitional improvement is misreporting type $\underline{\theta}$ as another type. The maximum value for $V_{\hat{c}(\underline{\theta})}+(\hat{c}(\underline{\theta})-\underline{\theta})\,x_{\hat{c}(\underline{\theta})}-V_{\underline{\theta}}$ under such a deviation is $\frac{f(\widetilde{\theta})}{f(\underline{\theta})}\Pi$ (achieved when $\hat{c}(\underline{\theta})=\widetilde{\theta}$). To implement this increase in the coalitional information rent, S incurs the cost of adjusting the agent's share of this rent. This share is determined by the **AIC** and **AIR** constraints. We concentrate on the rent for type $\widetilde{\theta}$. After the deviation, **AIC** $\left(\underline{\theta}|\widetilde{\theta}\right)$ and **AIR** $(\underline{\theta})$ together imply that

$$\hat{u}\left(\widetilde{\theta}\right) \geq \hat{u}(\underline{\theta})-\left(\widetilde{\theta}-\underline{\theta}\right)x_{\hat{c}(\underline{\theta})} \geq U_{\underline{\theta}}-\left(\widetilde{\theta}-\underline{\theta}\right)x_{\widetilde{\theta}}=\Pi. \tag{44}$$

Since the utility of type $\widetilde{\theta}$ is 0 prior to deviation, there is a net cost incurred by S of the amount Π whenever the type of the agent is $\widetilde{\theta}$. When weighted by the appropriate probabilities in the objective function in (11), we see that S's gain whenever A is type $\underline{\theta}$ is completely consumed by her loss whenever A is type $\widetilde{\theta}$. Therefore, there is no profitable deviation for S.

7.5 Proof for Proposition 4

It follows from Proposition 3 that the solution to (16) is optimal within the class of the collusion proof outcomes inducing monotonic output profiles. Moreover, when the no-supervision monotonicity constraint is slack, Lemma 2 implies that there is no collusion proof outcome with a non-monotonic output profile superior to the no-supervision optimal outcome, which is already dominated by the solution to (16).

7.6 Proof for Proposition 5

The expected coalitional information rent level $f\left(\overline{\theta}\right)V_{\overline{\theta}}+f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}+f\left(\underline{\theta}\right)V_{\underline{\theta}}$ for the outcome (17) is equal to $f\left(\overline{\theta}\right)V_{\overline{\theta}}^{o}\left(\cdot\right)+f\left(\widetilde{\theta}\right)V_{\widetilde{\theta}}^{o}\left(\cdot\right)+f\left(\underline{\theta}\right)V_{\underline{\theta}}^{o}\left(\cdot\right)$ evaluated for the same non-monotonic output profile. Recall that this is the lower bound on the expected coalitional information rent identified in Lemma 2. To complete the proof, it suffices to show delegation proofness of (17) for any arbitrary output profile satisfying $x_{\widetilde{\theta}}<x_{\overline{\theta}}\leq x_{\underline{\theta}}$. Establishing delegation proofness of (17) follows the same steps as establishing collusion proofness of (15) in the second part of the proof of Proposition 3.

Outcome (17) satisfies the participation constraints for both A and S. Next, we check for delegation feasibility. We start by observing that **AIC** and $\mathbf{d}-$**AIR** constraints are satisfied for problems in (8) and (9). For either conditions (8) or (9) to fail, there must exist an alternative outcome satisfying the constraints and yielding a higher value of the objective function. A necessary condition for this is the existence of a type θ with a misreport $\hat{c}\left(\theta\right)$ that gives a higher coalitional rent:

$$V_{\hat{c}(\theta)}+\left(\hat{c}\left(\theta\right)-\theta\right)x_{\hat{c}(\theta)}>V_{\theta}. \tag{45}$$

By construction of the coalitional rent levels in (17), there is no coalitional improvement from misreporting type $\underline{\theta}$. Moreover, it follows from the failure of condition (3) that $f\left(\widetilde{\theta}\right)\left(\overline{\theta}-\widetilde{\theta}\right)<f\left(\underline{\theta}\right)\left(\widetilde{\theta}-\underline{\theta}\right)$. This rules out the possibility of a coalitional improvement from misreporting type $\overline{\theta}$. The remaining possibility is a coalitional improvement from misreporting type $\widetilde{\theta}$. The maximum value for $V_{\hat{c}(\widetilde{\theta})}+\left(\hat{c}\left(\widetilde{\theta}\right)-\widetilde{\theta}\right)x_{\hat{c}(\widetilde{\theta})}-V_{\widetilde{\theta}}$ under such a manipulation is $\frac{f(\underline{\theta})}{f(\widetilde{\theta})}\left(\widetilde{\theta}-\underline{\theta}\right)\left(x_{\underline{\theta}}-x_{\widetilde{\theta}}\right)$ (achieved when $\hat{c}\left(\widetilde{\theta}\right)=\overline{\theta}$). To implement this increase in the coalitional information rent, S incurs the cost of adjusting A's share of this rent. A's share is determined by the $\mathbf{d}-$**AIC** and **AIR** constraints. We concentrate on the rent for type $\underline{\theta}$. After the deviation, **AIC** $\left(\widetilde{\theta}|\underline{\theta}\right)$ and

$\mathbf{d}-\mathbf{AIR}\left(\widetilde{\theta}\right)$ together imply that

$$\hat{u}\left(\underline{\theta}\right) \geq \hat{u}\left(\widetilde{\theta}\right) + \left(\widetilde{\theta} - \underline{\theta}\right) x_{\hat{c}\left(\widetilde{\theta}\right)} \geq \left(\widetilde{\theta} - \underline{\theta}\right) x_{\overline{\theta}}. \tag{46}$$

Since the utility of type $\underline{\theta}$ is $\left(\widetilde{\theta} - \underline{\theta}\right) x_{\widetilde{\theta}}$ prior to the deviation, the net cost of the manipulation for S whenever the type of the agent is $\underline{\theta}$ is $\left(\widetilde{\theta} - \underline{\theta}\right)\left(x_{\underline{\theta}} - x_{\widetilde{\theta}}\right)$. When weighted by the appropriate probabilities in the objective function in (9), we see that S's gain whenever A is type $\widetilde{\theta}$ is completely consumed by her loss whenever A is type $\underline{\theta}$. Therefore, there is no profitable manipulation available to S.

References

[1] A. Abdulkadiroglu, K. Chung, Auction design with tacit collusion, Columbia University working paper (2003).

[2] S. Baliga, T. Sjostrom, Decentralization and collusion, J. Econ. Theory 83 (1998), 196-232.

[3] B. Caillaud, P. Jehiel, Collusion in auctions with externalities, RAND J. Econ. 29 (1998), 680-702.

[4] G. Celik, Three essays on the informational aspects of untrustworthy experts, elusive agents, and corrupt supervisors, Ph.D. Thesis, Northwestern University, 2002.

[5] G. Celik, Counter marginalization of information rents: Implementing negatively correlated compensation schemes for colluding parties, The B.E. Journal of Theoretical Economics (Contributions) 8 (2008), Article 3.

[6] Y. Che, J. Kim, Robustly collusion-proof implementation, Econometrica 74 (2006), 1063-1107.

[7] Y. Che, J. Kim, Optimal collusion-proof auctions, Columbia University working paper (2007).

[8] A. Faure-Grimaud, J.J. Laffont, D. Martimort, A theory of supervision with endogenous transaction costs, Ann. Econ. Finance 1 (2000), 231-263.

[9] A. Faure-Grimaud, J.J. Laffont, D. Martimort, Collusion, delegation and supervision with soft information, The Rev. Econ. Stud. 70 (2003), 253-280.

[10] A. Faure-Grimaud, D. Martimort, On some agency costs of intermediated contracting, Econ. Letters 71 (2001), 75-82.

[11] L. Felli, Preventing collusion through discretion, LSE working paper (1996).

[12] B. Holmstrom, R. Myerson, Efficient and durable decision rules with incomplete information, Econometrica 51 (1983), 1799-1819.

[13] B. Jullien, Participation constraints in adverse selection models, J. Econ. Theory 93 (2000), 1-47.

[14] F. Khalil, J. Lawarree, Incentives for corruptible auditors in the absence of commitment, J. Ind. Econ. 15 (2006), 457-478.

[15] D. Kim, J. Lawarree, D. Shin, Exit option in hierarchical agency, Int. J. Ind. Organ. 22 (2004), 1265-1287.

[16] F. Kofman, J. Lawarree, Collusion in hierarchical agency, Econometrica 61 (1993), 629-656.

[17] F. Kofman, J. Lawarree, A prisoner's dilemma model of collusion deterrence, J. Public Econ. 59 (1996), 117-136.

[18] J.J. Laffont, D. Martimort, Collusion under asymmetric information, Econometrica 65 (1997), 875-911.

[19] J.J. Laffont, D. Martimort, Collusion and delegation, RAND J. Econ. 29 (1998), 280-305.

[20] J.J. Laffont, D. Martimort, Mechanism design with collusion and correlation, Econometrica 68 (2000), 309-342.

[21] T.R. Lewis, D.E.M. Sappington, Countervailing incentives in agency problems, J. Econ. Theory 49 (1989), 294-313.

[22] G. Maggi, A. Rodriguez-Clare, On countervailing incentives, J. Econ. Theory 66 (1995), 238-263.

[23] D. Martimort, The life cycle of regulatory agencies: Dynamic capture and transaction costs, Rev. Econ. Stud. 66 (1999), 929-947.

[24] P. McAfee, J. McMillan, Organizational diseconomies of scope, J. Econ. Manage. Strategy 4 (1995), 399-426.

[25] N. Melumad, D. Mookherjee, S. Reichelstein, Hierarchical decentralization of incentive contracts, RAND J. Econ. 26 (1995), 654-692.

[26] D. Mookherjee, Decentralization, hierarchies, and incentives: A mechanism design perspective, J. Econ. Lit. 44 (2006), 367-390.

[27] D. Mookherjee, M. Tsumagari, The organization of supplier networks: effects of delegation and intermediation, Econometrica 72 (2004), 1179-1219.

[28] R. Myerson, Mechanism design by an informed principal, Econometrica 51 (1983), 1767-1797.

[29] G. Pavlov, Colluding on participation decisions, Boston University working paper (2006).

[30] L. Quesada, Collusion as an informed principal problem, University of Wisconsin working paper (2004).

[31] S. Severinov, The value of information and optimal organization, RAND J. Econ. forthcoming.

[32] R. Strausz, Deterministic versus stochastic mechanisms in principal-agent models, J. Econ. Theory 127 (2006), 306-314.

[33] J. Tirole, Collusion and the theory of organizations, in: J.J. Laffont (Ed.), Advances in Economic Theory: Sixth World Congress, Vol. II, Cambridge Univ. Press, Cambridge, 1992.