MECHANISM DESIGN WITH COLLUSIVE SUPERVISION

by

Gorkem Celik
Department of Economics
University of British Columbia

August 2003

Discussion Paper No.: 03-06

DEPARTMENT OF ECONOMICS
THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, CANADA V6T 1Z1

http://www.econ.ubc.ca

Mechanism Design with Collusive Supervision*

Gorkem Celik†

First Version: March 2001
This Version: August 5, 2003

Abstract

We analyze an adverse selection environment with third party supervision. We assume that the *supervisor* and the *agent* can collude while interacting with the *principal*. As long as the supervisor is symmetrically informed with the agent, the former's existence does not improve the principal's rent extraction. This is due to the *coalitional efficiency* between the supervisor and the agent. However, asymmetric information between these two parties can cause a *collusion failure*, which undermines the coalitional efficiency. In that case, we show that the principal can increase his payoff, by manipulating the agent's opportunity cost for colluding with the supervisor. Delegating the authority to contract with the agent to the supervisor is not successful in enhancing the principal's payoff, since the principal loses the instrument to manipulate the opportunity cost of collusion under delegation. The increase in the principal's rent extraction does not necessarily imply an overall welfare improvement. Social welfare may decline with the introduction of the supervisor.

Key Words: Collusion, supervision, mechanism design

JEL Classification: D82, C72, L51

1 Introduction

The individual with the best information on the costs of an economic activity is often the "agent" who incurs these costs. It is one of the findings of the adverse selection literature that a "principal"

*This is a revised version of a chapter in my dissertation submitted to Northwestern University. I am grateful to my advisors Michael Whinston and Jeff Ely. Part of this work was completed when I was a visiting student at the University of Toulouse, where I benefited from valuable conversations with Bruno Jullien, Jean-Jacques Laffont, David Martimort, and Jean Tirole. I thank Antoine Faure-Grimaud, Guofu Tan, and Okan Yilankaya for helpful comments. I also thank participants at various seminars and conferences. Financial support from Robert Eisner Graduate Fellowship and CSIO at Northwestern is gratefully acknowledged.

†Department of Economics, The University of British Columbia, #997-1873 East Mall, Vancouver, BC, V6T 1Z1 Canada. e-mail: celik@interchange.ubc.ca

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 is sacrificing.[1] At the same time, the introduction of this "supervisor" also creates a potential for collusive behavior against the principal's will. If the supervisor is totally corrupted by the agent, then the supervisor - agent pair behaves like a single player. In that case, from the principal's perspective, contracting with the supervisor - agent coalition is not different than contracting with the agent, and ignoring the supervisor.

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

A necessary condition for the principal's making some use of 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. The contribution of this paper is showing how the principal can manipulate the supervisor - agent interaction to support a coalitional inefficiency that serves his own interests.

In the context of our analysis, asymmetric information is the only reason why the supervisor - agent coalition may fall short of perfect collusion. It is trivial that a supervisor without any information is useless. At the other extreme, if the supervisor is as informed as the agent, supervision is useless again, since an inefficiency at the coalition level cannot be supported. At the intermediate range, however, we identify a mechanism which makes supervision relevant to increase the principal's payoff.

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's contracting with the supervisor and delegating her the authority to contract with the agent. One commonly observed theme in the literature on multi-agent contracts is an organizational equivalence principle: Delegation performs as well as any other grand contract would.[2] Delegation restricts the principal's ability to create

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

[2] See Melumad, Mookherjee, and Reichelstein (1995) among others.

direct incentives for the agent. Nevertheless, he can influence the supervisor's interaction with the agent to create indirect incentives. However, in our setup, this indirect influence scheme does not fulfil the task. In the environment we consider, we establish a failure result for delegation: From the principal's perspective, delegating to the supervisor is at least as bad as not having a supervisor. In order to make use of supervision, keeping the communication channel open with the agent has vital importance. This failure result for delegation, together with the existence of a general mechanism that makes supervision relevant, imply that organizational equivalence does not carry on to our environment.

One final implication of the paper concerns social welfare. Although introduction of supervision increases the payoff for the principal, we show that it may decrease the overall efficiency. This is contrary to the predictions from many earlier models of supervision.

The paper is organized as follows: In the next section, we give an overview of the related literature. We introduce the general model in section 3 and characterize the set of outcomes that are implementable under collusion. This characterization result is essential to evaluate the performance of supervision under different assumptions. In section 4, we show why delegation fails as a response to the threat of collusion. In section 5, we prove that supervision is beneficial for the principal provided that the supervisor is less informed than the agent. In section 6, we discuss these results in connection with the existing literature. In section 7, we introduce a more structured model to identify the optimal implementable outcome under supervision. We employ the same model to examine implications regarding social welfare. We also conduct an exercise to see how performance of delegation reacts to a variation in the standard assumptions. Section 8 concludes. Section 9 is the appendix which includes the omitted proofs in the text.

2 A Review of The Literature

Collusion in the principal - supervisor - agent hierarchy has been studied by economic researchers for many years.[3] Tirole (1986) assumes the principal can offer the grand contract at the ex ante stage, before the type for the agent and information for the supervisor are realized. Although there is no inefficiency in the collusion stage, supervision is useful, since the principal can exploit the fact that the supervisor is uninformed when she is signing the contract.[4]

Laffont and Tirole (1991) replace this timing with the standard adverse selection paradigm, where

[3]For an extensive review of the collusion literature, see Tirole (1992).

[4]This is especially easy to see when the supervisor is risk-neutral. In that case, the principal would sell "the right to be the residual claimant" to the supervisor, at a price that equals the expected profit she would make. Also see Kofman and Lawarree (1993) for an example of how the principal can exploit the fact that the agent is uninformed when he is signing the contract.

the contract is signed after the parties are informed. To make supervision relevant, they employ an asymmetric transaction technology across the players. That is, they assume exogenous transaction costs for any transfer from the agent to the supervisor, whereas a transfer from the principal is frictionless. Due to its convenience, this exogenous transaction cost perspective has a wide acceptance among the researchers of collusive behavior.[5] In this paper, rather than assuming an exogenous imperfection between the supervisor and the agent, we derive a failure of collusion from the asymmetric information between these parties. We observe that the asymmetric information model has some implications, such as the possible decline in the overall efficiency with the introduction of supervision, that cannot be captured by the transaction cost models of collusion.

Our analysis in this paper is also related to the literature on multi-agent contract theory. Melumad, Mookherjee, and Reichelstein (1995) analyze the effect of the principal's delegation to a third party. They allow this party to take part in the production process. But they assume that she has no information about the type of the other productive agent. In this environment, they show an organizational equivalence: From the principal's perspective, delegating to the third party performs as well as contracting directly with both agents.[6] There is an immediate corollary to the organizational equivalence result: Since delegation induces explicit side contracting between the agents, the possibility of collusion is completely harmless for the principal.[7] In a recent paper, Mookherjee and Tsumagari (2003) show that the equivalence result fails when collusion is between two productive agents who contribute to the production of a single good. The poor performance of delegation in this last paper is due to the same reasons we will propose to explain its complete failure in our environment.[8]

We encounter a similar organizational equivalence result in the framework of moral hazard environments with collusion: Baliga and Sjostrom (1998) show that the principal can constrain the efficiency of the collusion through a careful division of the information rent among the colluding agents. They show that, as long as the appropriate agent is chosen to delegate, delegation does not impose any loss for the principal.

---

[5]See Laffont and Martimort (1999), Laffont and Meleu (2001) among others. And for earlier exceptions, see Felli (1996), Kofman and Lawarree (1996). The former uses informational asymmetry between the supervisor and the agent to show that the possibility of collusion is harmless given a certain colluding technology. The crucial assumption for this strong claim is the lack of full commitment at the collusion stage, rather than the informational asymmetry. The latter examines collusion between an agent and two supervisors that are sent sequentially to monitor the agent. The principal can benefit from *sometimes* informing the supervisor whether she is sent as the second monitor but not telling the agent whether the supervisor is informed.

[6]See Baron and Besanko (1994) for a similar result, and McAfee and McMillan (1995) for how this result breaks down under limited liability.

[7]See also Laffont and Martimort (1998) on this.

[8]For an up to date survey of the literature on delegation, see Mookherjee (2003).

Faure-Grimaud, Laffont, and Martimort (2002) (hereafter FLM) pursue a similar research question to ours in the context of a different environment. This environment is characterized by a unit production cost which may assume two possible values, and an informative supervisory signal which may also assume two possible values.[9] The equivalence result is valid for this environment as well. That is, delegating to the supervisor is an optimal organizational form to defy collusion.

In this paper, we diverge from the FLM environment by extending the number of possible cost levels (to an arbitrary natural number), and adopting a connected (monotone) partition of these costs as the information structure for the supervisor. This information structure is general in the sense that it can capture fully informed and completely uninformed supervisors as well as a variety of other circumstances in-between these extremes. Nevertheless, it is not a generalization of the FLM environment since it rules out signals that lead to posteriors with intersecting supports over the cost levels. This information structure provides us with an environment, where delegation completely fails as a response to the threat of collusion: Under standard assumptions regarding the distribution of costs, delegating to the supervisor performs at least as badly as not having a supervisor.[10] Therefore the relevance of supervision here depends on the performance of a more general class of mechanisms.

After establishing failure of delegation, we move on to identifying a mechanism that would sustain beneficial supervision in spite of the potential for collusion. In this mechanism, relevance of supervision is a result of the agent's type dependent outside options at the time that he is colluding with the supervisor. As is shown by Lewis and Sappington (1989), for some states of nature, an agent with a type dependent reservation utility will have an incentive to overstate his productivity, in order to increase his compensation for sacrificing the *outside option*.[11] Such an incentive is in the opposite direction from the original incentive to understate the productivity to increase the compensation for *production costs*. In our model, under standard assumptions, reversing the agent's incentives at the collusion stage is the only instrument for the principal to benefit from supervision. Failure of delegation results from its restriction over the use of this instrument.

As in the models of common agency and renegotiation,[12] the outside option at the collusion stage is endogenously determined in our model. We show that the principal can always support the appropriate outside options by designing the appropriate mechanism, provided that there is informational asymmetry

[9]For a similar approach to collusion between two productive agents, see Laffont and Martimort (1997, and 2000).

[10]The reason that our environment yields a dramatically different delegation result than that of the FLM environment will be discussed in section 6.

[11]Also see Maggi and Rodriguez-Clare (1995), Julien (2000) on the countervailing incentives.

[12]See Stole (1990) for common agency and Laffont and Tirole (1993) Ch. 10 for renegotiation. Also see Caillaud, Jullien and Picard (1996a, 1996b) on how a principal could gain from changing the reservation utilities of other players in a moral hazard setup.

between the supervisor and the agent.

Studying multi-agent contracts generally furnishes the researcher with the option of what participation constraints to adopt. The *interim* participation constraints provide the player with a non-negative *expected* utility conditional on the player's information. The *ex-post* participation constraints rule out any realization of strictly negative utility. We can think of assumptions that could justify either set of constraints. In many studies, the researcher commits to such an assumption and states results pertaining to the corresponding constraint set. In this paper, we do not restrict attention to a single participation concept. Instead, we prove the failure and irrelevance results under the weaker interim constraints and the relevance results under the stronger ex-post constraints. Therefore, each result is valid under either set of constraints.

3 The General Model

We will consider a setup with three players: the agent (A), the supervisor (S), and the principal (P).

The agent is the player who bears the costs of production. For a given output level x, A's disutility from production is $c_n x$, where n is the type for A. n can assume values from the set $\mathbb{N} = \{1, ..., N\}$. The unit cost is decreasing in A's type, i.e., $c_1 > c_2 > ... > c_N > 0$. The prior probability that A is type n is denoted by f_n. And, $F_n = \sum_{i \leq n} f_i$ is the cumulative distribution function associated with f_n. The prior distribution is common knowledge among the players. The agent also knows the realization of his type.

The agent receives a transfer, $t \in \Re$, from the principal and pays a bribe, $b \in \Re$, to the supervisor. A's utility as a function of his type, output, bribe and transfer is

$$t - b - c_n x.$$

Let $D = \{d_1, ..., d_L\}$ be a connected partition of set $\mathbb{N}$. Given A is type n, S observes d_l such that $n \in d_l$. We will refer to $l \in \mathbb{L} = \{1, ..., L\}$ as the type of S. With a minor abuse of notation, l will also denote the function $l : \mathbb{N} \to \mathbb{L}$, which maps types for A to types for S. Since D is a connected partition, we can assume $l(\cdot)$ is weakly increasing without loss of generality. We will also define $\underline{n}(l)$ and $\overline{n}(l)$ as the smallest and the largest elements of set d_l respectively.

An example that fits the connected partition assumption is as follows: Each element of D can be considered as a generation of production technologies, where each generation induces a number of different cost levels. Since A is the party incurring the production costs, he observes the relevant cost level. On the other hand, S is only capable of identifying the relevant generation. Under this interpretation of the model the connected partition specification amounts to assuming later generations to be superior to the earlier ones: Even the highest cost level associated with a later generation is lower than the lowest cost level induced by an earlier generation.

S is the player without any direct interest in the production. Her payoff is determined by the monetary payments she gets. Let w be the payment she receives from the principal. Then S's utility is:

$$w + b$$

P's only information about the types of A and S is the prior probability distribution. As the residual claimant of the production, he receives the direct benefit of $W(x)$, from the production of x units of output. The utility for P, as a function of output and transfer levels, is:

$$W(x) - t - w$$

We will assume that $W(\cdot)$ is a twice continuously differentiable function that satisfies some standard conditions: $W'(x) > 0, W''(x) < 0$, for all x, and $\lim_{x\to 0} W'(0) = \infty$, $\lim_{x\to\infty} W'(x) = 0$.

Note that the information profile for the players has a nested structure: A knows the information set that S observes. And P's only information is a probability distribution which is the common prior among all the players. Also note that every player is risk neutral in monetary transfers.

Here is the timing for the game:

T1: n, and therefore l are realized. n is observed by A. l is observed by S.

T2: P announces a **grand contract**. The grand contract is 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. And the functions specify:

i) the output level, $x : M_S \times M_A \to \Re_+$

ii) the transfer for A, $t : M_S \times M_A \to \Re$

iii) the wage for S, $w : M_S \times M_A \to \Re$

T3: S and A simultaneously decide whether to accept or reject the grand contract. If any of the players reject the grand contract, the game ends with zero production level and no monetary transfers to any player. In that case S and A receive zero utility. If they both accept, the game moves to the next stage.

T4: S offers a **side contract** to A. The side contract is a collection of a message space for A, M'_A, as well as two functions defined on M'_A, that specify:

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, $b : M'_A \to \Re$

T5: A decides whether to accept or reject the side contract. If A accepts the side contract, he also decides which element of M'_A to send to S.

T6: Both S and A send their messages to P. If the side contract is accepted by A at T5, these messages are determined by A's choice at T5.[13] If the side contract is rejected, then both players are

[13]In this paper, we ignore the issue of the enforcement of the contracts and assume that the side contract is binding as

free to send any message they want.

T7: The output level and transfers from P to other players are determined by the grand contract and the messages sent in the previous period. If the side contract is accepted, then A makes S the bribe transfer that is associated with his choice at T5.[14]

We will start with analyzing the game that follows the acceptance of the grand contract by both S and A. The conditions that ensure acceptance will be discussed later.

- **No Collusion:**

For now, we will assume that S cannot offer a side contract to A at T4. In that case, S and A will send their messages non-cooperatively at T6. A behavioral strategy for S (A), following the acceptance of the grand contract, is a function that maps her (his) type to a message.

Definition 1 *Let* $GC = \{M_S, M_A, x(\cdot), t(\cdot), w(\cdot)\}$ *be a grand contract; and* $\sigma : \mathbb{L} \to M_S$, $\alpha : \mathbb{N} \to M_A$ *be two functions defined on players' type spaces.* $\{\sigma(\cdot), \alpha(\cdot)\}$ *is a* ***non-cooperative equilibrium***[15] *of* GC *if*

$$\sigma(l) \in \arg\max_{m_s \in M_S} \left\{ \sum_{n \in d_l} f_n w(m_s, \alpha(n)) \right\} \text{ for all } l. \tag{1}$$

$$\alpha(n) \in \arg\max_{m_a \in M_A} \{ t(\sigma(l(n)), m_a) - c_n x(\sigma(l(n)), m_a) \} \text{ for all } n. \tag{2}$$

An agent of type n knows S's type is $l(n)$, and she will send the message $\sigma(l(n))$ in the equilibrium. Therefore, he chooses the message that will maximize his utility under this condition. However, S does not observe A's type directly. So, her equilibrium message solves an expected utility maximization problem.[16]

- **Collusion:**

Now, we will add the collusion stage to our analysis. At T4, S has already observed l, and knows that A's type is an element of d_l. Given the grand contract, the supervisor's choice of the side contract

well as the grand contract. For how a dynamic interaction may lead to commitment in covert contracts, see Martimort (1999).

[14]By not allowing lotteries in their respective definitions, we assume both the grand contract and the side contract to be deterministic. At the grand contract level this is without loss of generality: Since $W''(\cdot) < 0$ and the players are risk neutral in monetary transfers, stochastic grand contracts are dominated by the deterministic ones (from the principal's perspective). However, the possibility of stochastic side contracts increases the collusion opportunities and therefore shrinks the implementable set of outcomes further. See footnote 31 for more on the stochastic side contracts.

[15]We restrict attention to pure strategy Bayesian Nash equilibria.

[16]After observing l, the conditional probability of facing a type n agent is $\frac{f_n}{\sum_{i \in d_l} f_i}$ for S, provided that $n \in d_l$. In condition (1), we rescale S's problem by multiplying her expected utility by $\sum_{i \in d_l} f_i$.

is a mechanism design problem. Therefore, the revelation principle is a valid tool in this setting: We can restrict attention to direct contracts, where the message space for A is identical to the information set S observes ($M'_A = d_l$) and to truthful behavior for A, where A finds it optimal to reveal his type to S through the message he sends.

The remaining choice variables for S are the collective message function, $m(\cdot)$, and the bribe function, $b(\cdot)$, both of which are defined on M'_A. In order to induce truthful behavior, S should set these functions such that A does not prefer to imitate another type in the same information set. Moreover, S should also make sure that A accepts the side contract. This requires leaving him a rent that is at least the same as what he could get in the non-cooperative equilibrium[17] without the side contract.[18] Since the non-cooperative rent level for A depends on his type, S's optimization problem will be a mechanism design problem with type-specific reservation values.

Definition 2 *Let $GC = \{M_S, M_A, x(\cdot), t(\cdot), w(\cdot)\}$ be a grand contract. $\{\mu(\cdot), \beta(\cdot)\}$ is called a **collusive equilibrium** of GC if there exists a non-cooperative equilibrium, $\{\sigma(\cdot), \alpha(\cdot)\}$, of GC such that*

$$\{\mu(n), \beta(n)\}_{n \in d_l} \in \arg \max_{\{m(n), b(n)\}_{n \in d_l}} \left\{ \sum_{n \in d_l} f_n \left[w(m(n)) + b(n) \right] \right\} \text{ s.t.}$$

$$t(m(n)) - b(n) - c_n x(m(n)) \geq t(m(n')) - b(n') - c_n x(m(n')) \text{ for all } n, n' \in d_l \quad (3)$$

$$t(m(n)) - b(n) - c_n x(m(n)) \geq t(\sigma(l), \alpha(n)) - c_n x(\sigma(l), \alpha(n)) \text{ for all } n \in d_l \quad (4)$$

for all l.

We will also refer to $\{\sigma(\cdot), \alpha(\cdot)\}$ as the non-cooperative equilibrium that **supports** $\{\mu(\cdot), \beta(\cdot)\}$.

Constraint (4) guarantees A's acceptance of the side contract offer, whereas (3) implies A's truthful revelation of his type to S.[19]

The solution concept above employs "passive beliefs" on out-of-equilibrium paths. Whenever the side contract offer is rejected, both S and A are assumed to make no update on their interim beliefs of how their rival would play in the non-cooperative subgame that follows.[20] This is potentially problematic, since certain side contract offers can be rejected by certain types of A and accepted by some others.

[17] A grand conract may have multiple non-cooperative equilibria, and each non-cooperative equilibrium may lead to multiple collusive equilibria.

[18] Note that any outcome that results from A's rejection of the side contract can also be achieved by A's acceptance of an "expanded" side contract that induces A's non-cooperative behavior as an additional choice for A.

[19] We could equivalently model the collusion stage as a two player game, rather than as a mechanism design problem. In that case, the behavioral strategy for S, following the grand contract, would be her side contract offer as a function of the information set she observes. Similarly, the behavioral strategy for A would be his side contract acceptance and message choice decisions, both of which are functions of A's type and the side contract offer.

[20] For solution concepts that depend on belief updating of the mechanism designer, see Cramton and Palfrey (1995).

Therefore, A's rejection of a side contract can reveal some relevant information about his type. We will discuss the validity of the passive beliefs assumption in the next subsection, after characterizing the outcomes that can be supported by the collusive equilibria.

3.1 Collusion Feasibility

The ex-post utility levels for the players are determined by the level of production and allocation of the rent created through this production. $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ is sufficient to identify a production and distribution rule; where x_n and r_n are the output and the agent's utility levels when the realization of the agent's type is n, and u_n is the total information rent, or the summation of utility levels for S and A, associated with the same state of nature. We will refer to $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ as an **outcome**. The following definition identifies those outcomes that can be induced by the principal, given the possibility of collusion between S and A, but ignoring participation constraints.

Definition 3 *$\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ is a **collusion feasible outcome** if there exists a grand contract $GC = \{M_S, M_A, x(\cdot), t(\cdot), w(\cdot)\}$, and a collusive equilibrium $\{\mu(n), \beta(n)\}_{n\in\mathbb{N}}$ of GC such that:*

$$\begin{aligned} x_n &= x[\mu(n)] \\ u_n &= t[\mu(n)] + w[\mu(n)] - c_n x[\mu(n)] \\ r_n &= t[\mu(n)] - \beta(n) - c_n x[\mu(n)] \end{aligned}$$

The terminology here follows that of Holmstrom and Myerson (1983), where they define the property of "incentive feasibility," to refer to the set of outcomes such that each type of the agent is voluntarily truthful about his type. Similarly, for collusion feasibility, we require S's voluntary choice of the side contract that would make the S - A coalition reveal A's type. Note that the definition for collusion feasibility depends on three other concepts: grand contract, non-cooperative equilibrium and collusive equilibrium. Our next task is characterization of collusion feasible outcomes without referring to these primitive concepts.

One common theme in the earlier literature on collusion is "the collusion-proofness principle."[21] According to this principle, any collusion feasible outcome can be induced through a collusive equilibrium which replicates the underlying non-cooperative equilibrium.[22] We can think of the collusion proofness principle as an extended version of the revelation principle. Any outcome that can be supported with collusion can also be supported as a collusion-proof non-cooperative outcome.

[21] See Tirole (1986 and 1992) among others.

[22] Formally, $\mu(n) = \{\sigma(l(n)), \alpha(n)\}$ and $\beta(n) = 0$ for all n, where $\{\sigma(\cdot), \alpha(\cdot)\}$ is the non-cooperative equilibrium that supports $\{\mu(\cdot), \beta(\cdot)\}$.

The collusion-proofness principle is also valid in our environment. We can show that the task of designing a grand contract reduces to designing a direct revelation game where the set of available messages for each party is isomorphic to her (or his) respective type space. That is, there is no loss of generality in considering only the grand contracts that induce $M_A = N$, $M_S = L$. For an outcome to be collusion feasible, it must be the outcome to truthful reporting in this direct revelation game, and truthful reporting must constitute a non-cooperative equilibrium as well as a collusive equilibrium (together with zero as the uniform bribe level) supported by itself as the outside option. The following characterization result builds on this discussion.

Proposition 1 *(**Characterization of the collusion feasible outcomes**)* $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ *is collusion feasible if and only if*

$$\begin{aligned}
\{n, r_n\}_{n\in d_l} \quad &\in \quad \arg \max_{\{\hat{n}(n),\hat{r}(n)\}_{n\in d_l}\in\{\mathbb{N}\times\Re\}^{\#d_l}} \sum_{n\in d_l} f_n \left[u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} - \hat{r}\,(n)\right] \;\; s.t. \qquad (5)\\
\mathbf{AIC}\,(n'|n) \quad &: \quad \hat{r}\,(n) \geq \hat{r}\,(n') + (c_{n'} - c_n)\, x_{\hat{n}(n')} \text{ for all } n, n' \in d_l\\
\mathbf{AIR}\,(n) \quad &: \quad \hat{r}\,(n) \geq r_n \text{ for all } n \in d_l
\end{aligned}$$

for all l.

The proof for the proposition is in the appendix. With this new formulation of collusion feasibility, S still maximizes her expected payoff, given the information set she observes. For an agent of type n in her information set, S chooses a type to report, $\hat{n}\,(n)$, and a utility level for the agent, $\hat{r}\,(n)$. Since the report will be made to P, and since P does not directly observe any information on A's type, $\hat{n}\,(n)$ can be any type in N. Similarly, since there is no exogenous constraint on the side payment, $\hat{r}\,(n)$ can be any real number. The constraint $\mathbf{AIC}\,(n'|n)$ implies that type n would not lie to S, by pretending to be type n', which is in the same information set with n. And, $\mathbf{AIR}\,(n)$ requires S's leaving A a utility level that is at least as large as r_n. The proposition claims that $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ is a collusion feasible outcome if and only if the truthful reporting of the type ($\hat{n}\,(n) = n,$ for all n), and leaving the intended rent to each type ($\hat{r}\,(n) = r_n,$ for all n) constitute optimal behavior for S.

As discussed earlier, one remaining concern is the reasonableness of the passive beliefs assumption. This assumption rules out any update of S's and A's interim beliefs on each other's next play following a rejection of a side contract offer. To prove the sufficiency part of our characterization result in the appendix, we provide a class of direct revelation grand contracts that would induce truthful reporting as the non-cooperative equilibrium for any subgame that starts after a rejection of a side contract. Therefore, it suffices to consider how reasonable truthful reporting is as an equilibrium for each of those subgames. In the appendix, we show that truthful reporting is a weakly dominant strategy for all types of A. Thus it is an optimal strategy for A regardless of A's belief on S's play. Provided that A reports

truthfully, we also show that it is optimal for S to be truthful under any belief she may have on A's type, consistent with the information set she observes. Therefore, even if she was allowed to perform a Bayesian update of her beliefs after the rejection of the side contract, she would not be willing to change her best response to A's weakly dominant strategy. In other words, the truthful equilibrium we employ to characterize the set of collusion feasible outcomes is an "ex-post equilibrium," which is robust to any consistent update in the belief structure.[23]

We will end our discussion of collusion feasibility with the following corollary which is readily available from Proposition 1.

Corollary 1 *Suppose* $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ *is collusion feasible. Then, the following "within partition incentive compatibility (***WPIC***)" and "within partition monotonicity (***WPM***)" conditions are satisfied by* $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$.

$$\begin{aligned} \mathbf{WPIC}\,(n'|n) \quad &: \quad r_n \geq r_{n'} + (c_{n'} - c_n)\, x_{n'} \text{ for all } n, \text{ and } n' \in d_{l(n)} \\ \mathbf{WPM} \quad &: \quad n' \in d_{l(n)} \text{ and } n > n' \text{ imply } x_n \geq x_{n'} \end{aligned}$$

Proof. $\mathbf{WPIC}\,(n'|n)$ is identical to $\mathbf{AIC}\,(n'|n)$ at $\{\hat{n}\,(n)\,, \hat{r}\,(n)\}_{n\in\mathbb{N}} = \{n, r_n\}_{n\in\mathbb{N}}$. And **WPM** follows from $\mathbf{WPIC}\,(n'|n)$, $\mathbf{WPIC}\,(n|n')$ and the fact that $c_n < c_{n'}$. ■

Every type of A should be given the incentive not to imitate another type from the same information set. The output levels they will produce are the only tools to separate any such two types. The **WPIC** and **WPM** constraints formalize this argument. Note that these constraints are implied by the global incentive compatibility and monotonicity constraints that are standard in the absence of supervision. The difference with supervision is that we need not have either incentive compatibility or monotonicity across the boundaries of the partition.

3.2 Delegation Feasibility

The grand contract we define in T2 corresponds to an institutional design, where P contracts with both S and A at the same time. While constructing the grand contract, P also accounts for the fact that S and A would write a side contract to collude on the messages they would send to P. The interaction between these three parties can be considered as a triangular contract, where each player has a chance to communicate with and make a monetary transfer to the other two. Delegation is a special form of this

[23]The truthful equilibrium might not be the unique non-cooperative equilibrium of the direct grand contract. In case of multiple equilibria, the truthful equilibrium is the focal equilibrium since every player's strategy is "telling the truth." A stronger notion of feasibility would require a unique non-cooperative equilibrium and therefore would lead to a smaller feasible set. Nevertheless, the closure of that smaller set is also identical to the set of collusion feasible outcomes we identify.

triangular interaction, where there is no direct communication or transfer of money between P and A. Under delegation, P contracts with S only, and delegates her the authority to contract with A through a side contract.

The defining feature of delegation is the shut down of production in case that a side contract is not signed between S and A. The outside option for A provides him with a reservation utility of zero regardless of his realized type**.** In order to identify the outcomes that are feasible under delegation, we need to replace the **AIR** constraints of program (5) with the $\mathbf{d-AIR}$ constraints.

Definition 4 $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ *is a* ***delegation feasible outcome*** *if*

$$\begin{aligned} \{n, r_n\}_{n\in d_l} &\in \arg \max_{\{\hat{n}(n),\hat{r}(n)\}_{n\in d_l}\in\{\mathbb{N}\times\Re\}^{\#d_l}} \sum_{n\in d_l} f_n\left[u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} - \hat{r}\left(n\right)\right] \ \ s.t. \qquad (6)\\ \mathbf{AIC}\left(n'|n\right) &: \hat{r}\left(n\right) \geq \hat{r}\left(n'\right) + \left(c_{n'} - c_n\right) x_{\hat{n}(n')} \text{ for all } n, n' \in d_l \\ \mathbf{d-AIR}\left(n\right) &: \hat{r}\left(n\right) \geq 0 \text{ for all } n \in d_l \end{aligned}$$

for all l*.*

Note that condition (6) implies condition (5). This is due to the fact that delegation can be considered as a special case of collusive supervision.[24]

3.3 Implementation

Collusion feasibility determines what P could implement, given that both S and A have already accepted to participate in the grand contract he would propose. However, securing their participation requires a further contraction of the implementable set. Recall that the opportunity cost of accepting the grand contract is 0 for both S and A. Therefore all types of agents should be given a non-negative utility level:

$$r_n \geq 0 \text{ for all } n \in \mathbb{N} \qquad (7)$$

And S needs to have a non-negative expected utility for each possible realization of his information l:

$$\sum_{n\in d_l} f_n\left(u_n - r_n\right) \geq 0 \text{ for all } l \in \mathbb{L} \qquad (8)$$

Now, we can state our first implementation concept.

Definition 5 $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ *is an* ***implementable outcome*** *if it is collusion feasible and it satisfies the participation constraints (7) and (8). Moreover, an implementable outcome is* ***delegation implementable*** *if it is delegation feasible.*

[24] In this paper, we define delegation feasibility with Condition (6). Alternatively, we could introduce a class of **decentralized grand contracts** and define delegation feasibility as the property of being induced by those contracts. Delegation feasible outcomes are characterized by (6), under this alternative definition as well. See Celik (2002) for this approach.

Under the interim participation constraints, 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 the ones with positive surplus. However, if the supervisor can walk out of the contract after A's type is revealed or she is protected by limited liability, the relevant constraints for her would be the ex-post participation constraints. A stronger implementation concept can be defined by replacing the interim participation constraints of S with the ex-post ones.

Definition 6 *An implementable outcome, $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$, is **implementable with ex-post participation** if*

$$(u_n - r_n) \geq 0 \text{ for all } n \in \mathbb{N}. \tag{9}$$

The set of implementable outcomes can be regarded as a budget set for P. From this set, he wishes to choose the outcome that supports the highest expected payoff for him. We can formalize this problem as follows:

$$\max_{\{x_n,u_n,r_n\}_{n\in\mathbb{N}}} \sum_{n\in\mathbb{N}} f_n \left[W(x_n) - c_n x_n - u_n\right] \text{ s.t. } \{x_n, u_n, r_n\}_{n\in\mathbb{N}} \text{ is implementable.}$$

Since P does not have any preference on the distribution of the information rent he is giving up, his objective function is in terms of $\{x_n, u_n\}_{n\in\mathbb{N}}$ only. However, distribution of the rent, and therefore $\{r_n\}_{n\in\mathbb{N}}$ are still relevant for the optimization because of the implementability constraint.

Next, we will restate a standard implementation concept and a well-known result from the literature on mechanism design without supervision. Since the only players present in this case are P and A, an output - information rent profile, where the entire information rent is consumed by A this time, is sufficient to define an outcome in this setup**.**

Definition 7 *$\{x_n, u_n\}_{n\in\mathbb{N}}$ is **no-supervision implementable** if*

$$\begin{aligned} \mathbf{IC}(n'|n) &: u_n \geq u_{n'} + (c_{n'} - c_n) x_{n'}, \text{ for all } n, n' \in \mathbb{N} \\ \mathbf{IR}(n) &: u_n \geq 0, \text{ for all } n \in \mathbb{N} \end{aligned}$$

The above definition determines the extent of P's rent extraction power over A in the absence of a third player.

Proposition 2 *i) For any weakly monotonic sequence $\{x_n\}_{n\in\mathbb{N}}$ there exists a $\{u_n\}_{n\in\mathbb{N}}$ such that $\{x_n, u_n\}_{n\in\mathbb{N}}$ is no-supervision implementable. The smallest information rent levels that make such implementation possible are given by $\left\{u_n^o\left(\{x_n\}_{n\in\mathbb{N}}\right)\right\}_{n\in\mathbb{N}}$, where $u_n^o(\cdot)$ is defined recursively as:*[25]

$$\begin{aligned} u_1^o(\{x_n\}) &= 0 \\ u_n^o(\{x_n\}) &= u_{n-1}^o(\{x_n\}) + (c_{n-1} - c_n) x_{n-1} \text{ for } n > 1 \end{aligned}$$

[25]To ease the notational representation, hereafter we will write the argument of function $u_n^o(\cdot)$ as $\{x_n\}$.

ii) Let $\{x_n^{ns}\}_{n\in\mathbb{N}}$ *be the optimal implementable set of output levels for P. Then,*

$$\begin{aligned} \{x_n^{ns}\}_{n\in\mathbb{N}} &\in \arg\max_{\{x_n\}_{n\in\mathbb{N}}} \sum f_n\left[W\left(x_n\right) - c_n x_n - u_n^o\left(\{x_n\}\right)\right] \\ s.t.\ x_n &\geq x_{n-1} \textit{ for all } n > 1 \textit{ (monotonicity)} \end{aligned}$$

The proof is a standard one in the mechanism design literature and therefore omitted. The construction of function $u_n^o(\cdot)$ reveals that the downward adjacent **IC** constraints ($\mathbf{IC}(n-1|n)$) and the **IR** constraint of the least productive type ($\mathbf{IR}(1)$) are always binding for the no supervision problem. (Note that this last constraint can be regarded also as a downward adjacent constraint between the least productive type and a hypothetical type that does not participate in the mechanism.) Given these constraints are binding, part (ii) of the proposition states that monotonicity of the output profile is necessary and sufficient for all the other constraints.

Since the optimal information rent levels, identified by the functions $u_n^o(\cdot)$, are increasing in x_n for $1 \leq n < N$, the optimal output levels for these types are distorted downward from their respective "first best" levels: The optimal solution to the no-supervision problem induces "underproduction" with respect to the total welfare maximizing output levels. This underproduction phenomenon is a common property for the mechanism design problems where the productive agent has zero utility as his outside option.

The monotonicity constraint for the most productive types in the support of the distribution is never binding (i.e., $x_N^{ns} > x_{N-1}^{ns}$). Whether the monotonicity constraints for the other types are binding depends on the specification of the type distribution. For instance, if $\frac{1-F_n}{f_n}(c_n - c_{n+1})$ is declining in n, all the monotonicity constraints are slack (i.e., $x_n^{ns} > x_{n-1}^{ns}$ for all n). The continuous type space version of this last property is known as the monotone hazard rate property.

With or without supervision, the objective function for P remains the same. And the following argument reveals that the no-supervision problem has stronger constraints. Let $\{x_n, u_n\}_{n\in\mathbb{N}}$ be no-supervision implementable. It is easy to see that, under supervision, the same output - information rent pairs can be induced as a part of an implementable outcome where $r_n = u_n$. Such an implementation can be regarded as ignoring the existence of S and contracting with A only. But the more appropriate question here is whether the principal can benefit from the potentially weaker constraints associated with supervision. In other words, whether supervision makes it possible to induce a set $\{x_n, u_n\}_{n\in\mathbb{N}}$ that is not no-supervision implementable and that provides a payoff for P higher than the no-supervision optimal. The answer to this question will determine the "relevance" of supervision in collusive environments.

For supervision to be relevant, one necessary condition is the existence of an implementable outcome that violates some of the constraints of the no-supervision problem. This indicates that, for the relevance of supervision, the S - A coalition must fail to behave like a single individual and maximize their total gains by means of a side contract. As we discussed earlier, the literature on collusion sustains such

a coalitional inefficiency by adopting some form of exogenous transactional imperfection between the colluding parties. In what follows we do not make any such assumption on the transaction technology. The only potential source for a collusion failure is the asymmetric information between the colluding parties.

The coalitional inefficiency discussed above should not be mistaken as a *sufficient* condition for relevance as well. The test for relevance does not reduce to checking the potential for *any* inefficiency at the collusion stage. For supervision to be relevant, the collusion failure must result in the violation of some *binding* constraint of the no-supervision problem. As an example to this point, consider a situation where the monotonicity constraints are slack for the no-supervision problem: The only binding constraints are the downward adjacent constraints that we discussed earlier. In this case, in order to improve over the no-supervision optimal outcome, at least one of these binding constraints should be violated: For at least one type realization of A, the total information rent for the S - A coalition should have increased if they had collectively behaved as though A's type is lower. Since the no-supervision output levels are weakly increasing in type, this amounts to a specific coalitional inefficiency, where the S - A coalition fails to reduce the output level even though it is coalitionally efficient to do so.

4 Failure of Delegation

In the previous section, we introduced the grand contract as the most general form of organizational design to contract with S and A. We mentioned delegation as a special case of this design where P contracts with S only, and delegates her the authority to contract with A. By delegating to S, P sustains a loss of control over A. Nevertheless, performance of this organizational form is of special interest to the economists, since P may be forced to follow this path due to a variety of reasons including communication and information processing costs. The analysis in this section will also help to compare the message of this paper with the other recent developments in the literature on collusion.

We will start our analysis by stating three necessary conditions for delegation feasibility.

Lemma 1 *If* $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ *is delegation feasible, then*

$$r_n = \begin{cases} 0 & \text{if } n = \underline{n}\left(l\left(n\right)\right) \\ r_{n-1} + \left(c_{n-1} - c_n\right) x_{n-1} & \text{otherwise} \end{cases} \tag{10}$$

$$u_n \geq u_{n-1} + \left(c_{n-1} - c_n\right) x_{n-1} \text{ for } n, n-1 \in d_l \tag{11}$$

$$\sum_{n\in d_l} f_n u_n \geq \sum_{n\in d_l} f_n \left[u_{\overline{n}(l-1)} + \left(c_{\overline{n}(l-1)} - c_{\underline{n}(l)}\right) x_{\overline{n}(l-1)} + r_n\right] \text{ for } l > 1 \tag{12}$$

The proof of the lemma is in the appendix. Equation (10) pins down the feasible utility profile for A under delegation. It is not surprising that the construction of this profile mimics the construction of

the function $u_n^o(\cdot)$ of the no-supervision implementation problem: As we discussed earlier, the defining feature of delegation is the shut down of production whenever S and A cannot agree on a side contract. This implies that the outside option of side contract for A is zero utility, as is the outside option of a no-supervision contract. Condition (11) is necessary for S not to find a profitable to offer a side contract that misreports type n as type $n-1$. And similarly, condition (12) is necessary for S not to find it profitable to offer a side contract that misreports all the types in d_l as the type $\overline{n}(l-1)$.[26]

In light of the discussion at the end of the previous section, condition (11) equips us with the first piece of bad news regarding the relevance of delegation: The downward **IC** constraints between two adjacent types within the same information set cannot be violated if delegation is adopted as an organizational form. Nevertheless, delegation feasibility (or delegation implementability) is not sufficient to reproduce the downward adjacent **IC** constraints across different information sets. To provide a conclusive result for the performance of delegation, we will need to identify a lower bound for the expected information rent conditional on each information set. Once the participation constraint of S for $l=1$ is taken into account, Lemma 1 recursively identifies such a lower bound in terms of the functions $u_n^o(\cdot)$.

Lemma 2 *Suppose $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ is a delegation implementable outcome. Then,*

$$\sum_{n\in d_l} f_n u_n \geq \sum_{n\in d_l} f_n u_n^o(\{x_n\}) \tag{13}$$

holds for every l.

The proof of the lemma is in the appendix. Note that the output profile $\{x_n\}$ does not need to be monotonic for the functions $u_n^o(\cdot)$ to be well defined. The central result of this section immediately follows from this lemma.

Proposition 3 ***(Failure of Delegation)*** *Suppose monotonicity constraints are slack for the no-supervision problem. Then P's expected payoff from a delegation implementable outcome is not higher than his no-supervision optimal expected payoff.*

Proof. In the absence of the monotonicity constraints, the no-supervision problem can be written as

$$\{x_n^{ns}\}_{n\in\mathbb{N}} \in \arg\max_{\{x_n\}_{n\in\mathbb{N}}} \sum_{n\in\mathbb{N}} f_n\left[W(x_n) - c_n x_n - u_n^o(\{x_n\})\right].$$

Let $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ be delegation implementable. Since (13) holds for all l,

$$\begin{aligned}\sum_{n\in\mathbb{N}} f_n\left[W(x_n) - c_n x_n - u_n\right] &\leq \sum_{n\in\mathbb{N}} f_n\left[W(x_n) - c_n x_n - u_n^o(\{x_n\})\right] \\ &\leq \sum_{n\in\mathbb{N}} f_n\left[W(x_n^{ns}) - c_n x_n^{ns} - u_n^o(\{x_n^{ns}\})\right]\end{aligned}$$

[26]The lemma above, and the results following, would still be valid under an alternative timing of events, where S sends a report to P prior to side-contracting with A. This latter timing is due to Melumad, Mookherjee, and Reichelstein (1995).

A delegation implementable outcome cannot improve on the no-supervision optimal outcome. ■

Delegation reduces the side contract selection problem of S to a standard mechanism design problem similar to the no-supervision implementation problem. Therefore we expect the optimal side contract to fail to be efficient for the involved parties, as does the optimal no-supervision contract. However, the inefficiency in the side contract selection process reflects to the optimal outcome as an underproduction with respect to the coalitionally efficient output level. In other words, the optimal side contract under delegation has a potential not to increase the output level even though it is coalitionally efficient to do so. As a result, the inefficiency caused by delegating to S compounds with the inefficiency inherent in the no supervision problem rather than alleviating it. Therefore delegation fails to be a useful organizational choice for P regardless of the partitional structure of S's information.

The failure of delegation to improve over no-supervision can also be explained in terms of "double marginalization" of the information rents.[27] When S is interacting with A at the side contract selection stage, she behaves as the designer of a mechanism for a productive agent with zero reservation utility. Since S is generally not as informed as A, the optimal solution to this design problem requires leaving an information rent to A. It is due to this information rent that the "virtual" cost of production for S is higher than the real cost that A incurs. And P's interaction with S through the grand contract would be identical to his contracting with a productive agent where the cost of production is given as this virtual cost. Therefore, delegating to S, instead of following a no-supervision implementation, would have a similar effect as a rise in the production costs. Accordingly, it will be dominated by no-supervision.[28]

Before ending our discussion for this section, we should note the significance of the slackness of the monotonicity constraints in the derivation of our failure of delegation result. We will later observe in the context of our three type model that delegation might indeed improve over no-supervision if some of these constraints are binding.

5 Relevance of Supervision

So far we have proved that delegating to S is not a viable strategy to increase P's payoff over its no-supervision optimal level under the standard assumption that only the downward local constraints matter. As we have mentioned before, delegation is only a special case of implementation with super-

[27] See Melumad, Mookherjee, and Reichelstein (1995), Mookherjee and Tsumagari (2002) on a discussion of double marginalization in the context of two productive agents.

[28] If S is fully informed or completely uninformed on the type of A, it is possible to construct a delegation implementable outcome (with interim participation) that yields the same payoff for P as does the optimal no-supervision outcome. However, if supervisory information is between these extremes, delegation generally performs *strictly* worse than no-supervision.

vision. In order to say more on the relevance of supervision, we will need to consider the more general grand contracting approach. In this section we will pursue this task.

The first result of this section has the same negative nature as the failure of delegation result. Recall that a necessary condition to improve the principal's rent extraction over the no-supervision environment is enlarging the set of attainable output - information rent pairs. With the following proposition, we prove that such an enlargement is not possible if S is fully informed on A's type.

Proposition 4 *(**Full Information Irrelevance**) If S's information structure is as fine as A's (if all d_l are singleton), then any $\{x_n, u_n\}_{n\in\mathbb{N}}$ that is induced by an implementable outcome is also no-supervision implementable.*

Proof. With full information, the type space for S is isomorphic to the type space for A. We can rewrite the participation constraints (7) and (8) as follows:

$$\begin{aligned} r_n &\geq 0 \text{ for all } n \\ u_n - r_n &\geq 0 \text{ for all } n \end{aligned}$$

The next step is writing (5), the collusion feasibility constraint. Since all information sets are singleton, there is no **AIC** constraint and therefore all **AIR** constraints are binding at the optimal solution. Therefore (5) is identical to:

$$n \in \arg\max_{\hat{n}(n)\in\mathbb{N}} \left\{ u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} - r_n \right\} \text{ for all } n$$

The participation constraints imply $u_n \geq 0$, and collusion feasibility implies $u_n \geq u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)}$ for all n. Note that these are identical to the **IR** and **IC** constraints for the no supervision implementation. Since **IR** and **IC** are the only constraints for the no supervision implementation, supervision cannot enlarge the set of no supervision implementable output - information rent profiles. ■

The reason for the supervisor's irrelevance here is the coalitional efficiency of the supervisor - agent interaction. Any efficient trade between S and A is realized in the equilibrium. From the principal's point of view, the S - A coalition is behaving as if it is a single player. Therefore, inferring the information through the coalition is as costly as inferring it from a single agent. It is trivial that we have a similar irrelevance result at the other end of the spectrum, where S does not have more information than what P already knows about A (i.e., where $d_1 = N$). If the supervisory information is ever relevant for P's mechanism, it must be that the information is neither too good nor too bad. Our next task will be illustrating how P can benefit from such an informational structure. To prove that supervision is relevant for the principal's payoff, it is sufficient to show that supervision can sustain information rent levels smaller than $\{u_n^o(\{x_n\})\}_{n\in\mathbb{N}}$, when $\{x_n\}$ is a weakly monotonic output profile.

Proposition 5 *(**Relevance**) Let $\{x_n\}_{n\in\mathbb{N}}$ be a weakly monotonic profile of output levels. Suppose there exists $\tilde{l} \in \mathbb{L} - \{1\}$ such that $k, k' \in d_{\tilde{l}}$ and $x_k \neq x_{k'}$. Then, there also exists an outcome, $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$, which is implementable (with ex-post participation) and which satisfies $u_n \leq u_n^o(\{x_n\})$ for all n, with strict inequality for some.*

From the analysis of the previous sections, we already know that any no-supervision implementable output profile is also implementable under supervision, with the same information rent levels. This is because of the fact that P can replicate no-supervision implementation by ignoring the existence of S. The result above identifies a sufficient condition for implementation of those output levels with an expected information rent smaller than the smallest possible no-supervision expected rent level.[29] If the optimal no-supervision output profile satisfies this condition, relevance of supervision follows.

Corollary 2 *There exists an implementable outcome (with ex-post participation) that leaves a strictly higher expected payoff to P than does the optimal no-supervision outcome if either one of the two conditions below holds.*

i) The highest element of the partition, $d_{l(N)}$, is neither a singleton, nor the entire type space $\mathbb{N}$.

ii) The monotonicity constraints are slack for the no-supervision implementation problem and there exists $\tilde{l} \in \mathbb{L} - \{1\}$ such that $d_{\tilde{l}}$ contains at least two types.

The complete proof for Proposition 5 is in the appendix. Here, we provide a sketch. The hypothesis implies the existence of $\tilde{l} \neq 1$ such that $k, k+1 \in d_{\tilde{l}}$ and $x_{k+1} > x_k$. Consider the following $\{u_n, r_n\}_{n\in\mathbb{N}}$:

$$u_n = \begin{cases} u_n^o(\{x_n\}) & n \leq k \\ u_n^o(\{x_n\}) - \Delta & n > k \end{cases} \tag{14}$$

$$r_n = \begin{cases} u_n^o(\{x_n\}) - u_{\underline{n}(l(n))}^o(\{x_n\}) & n \leq k \\ u_n^o(\{x_n\}) - u_{\underline{n}(l(n))}^o(\{x_n\}) + \Psi & n > k \end{cases} \tag{15}$$

where Δ and Ψ are strictly positive real numbers that satisfy the following inequalities:

$$\Delta + \Psi \leq u_{\underline{n}(\tilde{l})}^o(\{x_n\}) \tag{16}$$

$$\Psi \leq (c_k - c_{k+1})(x_{k+1} - x_k) \tag{17}$$

$$\Delta \leq (c_{k+1} - c_{k+2})(x_{k+1} - x_k) \text{ if } k+1 < N \tag{18}$$

$$f_{k+1}(\Delta + \Psi) \leq (f_k + f_{k+1})\Psi \tag{19}$$

Since $\underline{n}\left(\tilde{l}\right)$ is larger than 1, and since x_{k+1} is larger than x_k, the right hand sides of inequalities (16) to (18) are strictly positive. This guarantees the existence of strictly positive Δ and Ψ. For types

[29] The hypothesis of Proposition 5 is also a *necessary* condition for implementability of such an outcome with ex-post participation.

smaller than or equal to k, the profile $\{u_n, r_n\}_{n\in\mathbb{N}}$ provides a coalitional rent that is equal to u_n^o. For these types, S's share of the rent is $u^o_{\underline{n}(l(n))}(\{x_n\})$, which is the no-supervision information rent level of the least productive type in her information set; and the rest of the rent is left for A. For types higher than k, the coalitional rent levels are reduced by the amount Δ, while A's share of the rent is increased by Ψ. Since Δ is strictly positive, the profile $\{u_n\}_{n\in\mathbb{N}}$ induces a lower information rent for types higher than k than does the profile $\{u_n^o(\{x_n\})\}_{n\in\mathbb{N}}$.[30]

Recall that $\{u_n, r_n\}_{n\in\mathbb{N}}$ above constitutes an outcome together with the preset output profile $\{x_n\}_{n\in\mathbb{N}}$. This outcome is not coalitionally efficient for the S - A pair. To see this, observe that $u_{k+1} < u_k + (c_k - c_{k+1})\,x_k$, since the right hand side is equal to $u_k^o + (c_k - c_{k+1})\,x_k$ and the left hand side is the same amount less Δ. Whenever the realized type for A is $k+1$, the output level (x_{k+1}) is higher than what is optimal (x_k) for the S - A coalition. Therefore, unlike what we have encountered for the case of delegation implementable outcomes, the collusion failure here distorts the output levels in the direction that would improve P's rent extraction.

Now the remaining question is why this outcome is implementable. More specifically, why the side contract fails to takes advantage of this potential for the improvement of the coalitional gains. If S could perfectly identify the type of A, she would be willing to pay a bribe by the amount Ψ to type $k+1$ and persuade him to misreport his type as k. Such a side contract would not change the agent's rent but would increase the supervisor's expected surplus (net of the bribe she pays) by $f_{k+1}\Delta$.

However, the information technology for S does not allow her to distinguish types k and $k+1$. Bribing type $k+1$ changes the incentives for type k at the side contracting stage. When type $k+1$ misreports himself as type k, both types produce the same output level and get the same transfer level through the grand contract. This requires S to pay the same amount of bribe to type k as she pays to type $k+1$. Otherwise, when colluding with S, type k would imitate type $k+1$, rather than revealing his type truthfully to S. When we take this indirect effect of bribing into account, the expected cost of the bribe for S turns out to be $(f_k + f_{k+1})\,\Psi$. Inequality (19) states that S's expected gain from the misreport of type $k+1$ is not high enough to cover the expected cost of the bribe that would induce the misreport.[31]

[30] The construction of $\{u_n, r_n\}_{n\in\mathbb{N}}$ indicates that S's payoff does not only depend on her observation of the information set, but also on the exact realization of A's type. In the absence of side transfers in the equilibrium, this suggests a wage function for S that is responsive to A's type. In fact, the same outcome is implementable with an inflexible wage function and equilibrium collusion. See Celik (2002) for this.

[31] Allowing for stochastic side contracts would require a stronger hypothesis for Proposition 5, but would not change the essence of our analysis. With stochastic side contracts, it would be possible for S to offer a side contract that misreports type $k+1$ as type k with a *nondegenerate* probability. As indicated in footnote 14, such a potential would increase the collusion possibilities. Under this alternative formulation of collusion, the outcome identified in (14) and (15) is still implementable given conditions (16), (18), (19) are satisfied and (17) holds as an equality. Therefore, if

The discussion above reveals an interesting feature of S's problem at the side contracting stage. $\mathbf{AIC}(k+1|k)$ is a binding constraint for S's maximization. In other words, S should exert an extra effort to make sure that a less productive type is not willing to imitate a more productive type at the collusion stage. Note that this is not the usual direction for the incentive compatibility constraints to bind in mechanism design problems. Under the basic paradigm of adverse selection, the relevant incentive compatibility constraints arise from the agent's preference to understate his productivity in order to increase the compensation for the variable production cost. However, for the supervisor's problem, the type of the agent signals not only his productivity, but also the outside option he could sustain by refusing the side contract offer. For the collusion-proof implementation of the above outcome, r_n is the final utility level for type n as well as his reservation utility at the collusion stage. Recall that the profile $\{r_n\}_{n\in\mathbb{N}}$ is constructed to support a "jump" in this reservation utility in-between the types k and $k+1$. This provides a "countervailing incentive" for type k: He may prefer to overstate his productivity in order to increase his compensation from S for the outside option he is foregoing. This new source of incentives culminates in a different form of inefficiencies for the side contract selection process: In order to prevent type k from imitating type $k+1$, S tends to keep the output level for the latter type higher. This results in a production level that is higher than what is coalitionally efficient.

6 Discussion

Our results so far indicate supervision can be relevant to improve P's payoff over its no-supervision optimal level provided that S is not perfectly informed (or completely uninformed) on the realized type of A. The mechanism that sustains this improvement employs a certain manipulation of A's outside option at the collusion stage. If, however, P insists on delegating to S, he would deprive himself of the instrument for such a manipulation. As a result, delegation is weakly dominated by no-supervision implementation and strictly dominated by a more general mechanism that does not restrict P's direct contracting with A.

Our result pertaining to delegation is quite in contrast with a result by Faure-Grimaud, Laffont, and Martimort (2002) (hereafter FLM), who examine the relevance of supervision in a different environment and conclude that delegation is an optimal organizational structure for the principal. The FLM environment is characterized by a unit production cost, which can assume two possible values (low cost or high cost), and a signal, which can also assume two possible 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 supervisor and the agent. This envi-

we allowed for stochastic side contracts, the hypothesis of Proposition 5 should be amended to include the condition $(c_k - c_{k+1})(x_{k+1} - x_k) < u^o_{\underline{n}(\bar{l})}(\{x_n\})$ so that a positive Δ exists when (17) holds as an equality.

ronment induces four different states of nature, each involving a different cost - signal pair. Note that a no-supervision implementation in this environment requires the **IC** constraints to be satisfied in-between these different states. It follows from our earlier arguments on coalitional inefficiency that improving over this no-supervision implementation demands violating some of these no-supervision constraints.

The optimal implementable outcome, which FLM identify, induces a lower output level for the low cost - high signal state than for the low cost - low signal state. It turns out that the only violated **IC** constraint by this outcome is the one between these two states: Whenever the realized state is the former of these two, it would be a coalitional improvement for the S - A pair to behave as though the state is the latter. For this improvement not to be realized, collusion must suffer from a failure that leads to a production level that is *lower* than what is coalitionally efficient. This points to a coalitional inefficiency that is in the opposite direction to the inefficiency we identified as the sufficient condition of relevance in our environment. Moreover, this underproduction feature of the required collusion failure is compatible with the performance of the optimal side contract under delegation. As a result, the principal does not need to manipulate the agent's reservation values further to sustain the relevant collusion failure. Therefore delegation is an adequate organizational choice in the FLM environment.

In relation to the FLM paper, this current paper attempts to further our understanding of mechanism design under collusion through (i) introduction of a different environment, where relevance of supervision requires violation of the (more standard) downward adjacent **IC** constraints of no-supervision implementation, (ii) explanation of why delegation would fail in this environment, and (iii) description of a more elaborate mechanism that would lead to relevant supervision even with ex-post participation.

7 The Three Type Model

With our relevance result, we established that supervision is beneficial for the principal, even if collusion between the supervisor and supervised agent is a possibility. That is because there exists an implementable outcome under supervision that improves over the optimal no-supervision outcome. In this section we will employ a more structured model to identify the *optimal* outcome that is implementable under supervision. We will also state our results on social welfare and reexamine the performance of delegation when the standard assumptions do not hold.

There are three possible types for A: $\mathbb{N} = \{1, 2, 3\}$; and S can only distinguish the least productive type from the others: $\mathbb{L} = \{1, 2\}$ with $d_1 = \{1\}$, $d_2 = \{2, 3\}$. We will start with the no-supervision implementation in this setup. Suppose $\{x_1, x_2, x_3\}$ is a profile of weakly increasing output levels. From Proposition 2, we know that $\{x_1, x_2, x_3\}$ is no-supervision implementable with the following information

rent levels:

$$u_1^o(\{x_n\}) = 0 \tag{20}$$

$$u_2^o(\{x_n\}) = (c_1 - c_2)\, x_1 \tag{21}$$

$$u_3^o(\{x_n\}) = (c_1 - c_2)\, x_1 + (c_2 - c_3)\, x_2 \tag{22}$$

And the optimal no-supervision implementable output profile is a solution to

$$\begin{aligned} \max_{x_1,x_2,x_3} & \sum_{\{x_n\}} f_n \left[W(x_n) - c_n x_n - u_n^o(\{x_n\}) \right] \\ \text{s.t. } x_3 \quad \geq \quad & x_2 \geq x_1 \end{aligned} \tag{23}$$

Following the discussion after the statement of Proposition 5, we can state that any monotonic $\{x_1, x_2, x_3\}$ is implementable (with ex-post participation) together with the following information rent profile:

$$\begin{array}{ll} u_1 = u_1^o(\{x_n\}) & r_1 = 0 \\ u_2 = u_2^o(\{x_n\}) & r_2 = 0 \\ u_3 = u_3^o(\{x_n\}) - \frac{f_2}{f_3}\Pi & r_3 = (c_2 - c_3)\, x_2 + \Pi \end{array} \tag{24}$$

where $\Pi = \min\left\{ \frac{f_3}{f_2+f_3}(c_1 - c_2)\, x_1, (c_2 - c_3)(x_3 - x_2) \right\}$. To see this, observe that the rent profile above is the same as the profile we constructed to prove Proposition 5, provided that $k = 2$, $\Psi = \Pi$, and $\Delta = \frac{f_2}{f_3}\Pi$. Since the value assigned for Π satisfies inequalities (16), (17), and (19) , we conclude that $\{x_1, x_2, x_3\}$ is implementable with the above rent profile. (Since $k + 1 = 3 = N$, inequality (18) is not relevant here.)[32] Provided that $\{x_1, x_2, x_3\}$ is set optimally, we will also prove that this particular outcome is the optimal implementable outcome as long as the monotonicity constraints are not binding for the no-supervision implementation problem.

Proposition 6 *Suppose the monotonicity constraints are slack for (23), the no-supervision implementation problem. A solution to the following problem is an optimal implementable outcome with ex-post participation:*

$$\begin{aligned} \max_{\{x_n, u_n, r_n\}} & f_3 \left[W(x_3) - c_3 x_3 - u_3 \right] + f_2 \left[W(x_2) - c_2 x_2 - u_2 \right] + f_1 \left[W(x_1) - c_1 x_1 - u_1 \right] \\ \textit{s.t. } x_3 \quad \geq \quad & x_2 \geq x_1 \textit{ and (24)} \end{aligned}$$

The proof of the proposition is in the appendix.[33] Since the optimal outcome above respects the monotonicity constraints of the no-supervision optimization, and u_3 is strictly smaller than $u_3^o(\{x_n\})$

[32] If $(c_2 - c_3)(x_3 - x_2) \leq \frac{f_3}{f_2+f_3}(c_1 - c_2)\, x_1$, this outcome is implementable even if stochastic side contracts are allowed.

[33] If the relevant participation constraints are interim, the maximization problem above still yields the optimal implementable outcome provided that the definition for Π is amended as $\Pi = \min\{(c_1 - c_2)\, x_1, (c_2 - c_3)(x_3 - x_2)\}$ and monotonicity constraints are slack for no-supervision implementation.

(recall that $x_3 > x_2$, and $x_1 > 0$), P improves upon his optimal no-supervision payoff. The improvement is due to the coalitional inefficiency when the realized type for A is 3. On this state of nature, the coalition would be better off if it behaved as if the type were 2. However, S sacrifices this coalitional gain to improve her rent extraction when the type for A is 2.

The introduction of S not only increases P's payoff, but also alters the output levels he is willing to implement. This change in the optimal output levels is also relevant in order to identify the effect of supervision on social welfare. We will address these issues in the following subsection.

7.1 Distortions and Social Welfare

With our relevance results we established that the principal's payoff increases with the introduction of supervision despite the potential for collusion. Another question of interest is the overall effect of supervision on the social welfare. Since utility functions of all players are quasilinear in money, the extent of output distortions is a good measure for social welfare. Substituting in the constraints for $\{u_n, r_n\}_{n=1,2,3}$, we can rewrite the optimization problem that is given in Proposition 6 as:

$$\max_{x_1,x_2,x_3,\Pi} \left\{ \begin{array}{l} f_3 \left[W(x_3) - c_3 x_3 - (c_1 - c_2)\, x_1 - (c_2 - c_3)\, x_2\right] \\ +f_2 \left[W\left(x_2\right) - c_2 x_2 - (c_1 - c_2)\, x_1\right] + f_1 \left[W\left(x_1\right) - c_1 x_1\right] + f_2 \Pi \end{array} \right\} \text{ s.t.}$$

$$f_2\Pi \leq f_2 \frac{f_3}{f_2 + f_3} (c_1 - c_2)\, x_1 \tag{25}$$

$$f_2\Pi \leq f_2 (c_2 - c_3)(x_3 - x_2) \tag{26}$$

$$x_3 \geq x_2 \geq x_1 \tag{27}$$

Since the derivative of the objective function with respect to $f_2\Pi$ is 1, and (25) and (26) are the only constraints on $f_2\Pi$, the Lagrange multipliers for these constraints should add up to 1. Let λ and $1-\lambda$ be the Lagrange multipliers for (26) and (25) respectively. We will ignore the monotonicity constraint (27), and perform a first-order analysis:[34]

$$W'(x_3^*) = c_3 - \lambda \frac{f_2}{f_3} (c_2 - c_3) \tag{28}$$

$$W'(x_2^*) = c_2 + \frac{f_3}{f_2} (c_2 - c_3) + \lambda (c_2 - c_3) \tag{29}$$

$$W'(x_1^*) = c_1 + \frac{f_3 + f_2}{f_1} (c_1 - c_2) - (1 - \lambda) \frac{f_2}{f_1} \frac{f_3}{f_2 + f_3} (c_1 - c_2) \tag{30}$$

$\lambda = 1$ implies (26) is binding and $\lambda = 0$ implies (25) is binding. If λ assumes an interior value then both constraints are binding.[35]

[34] Since $W''(\cdot) < 0$, the second order conditions for the maximization are satisfied.

[35] The value of λ is determined by the parametrization of the problem. However, given any value for the other exogenous variables, if c_3 is small enough, then $(c_2 - c_3)\left(x_3^* - x_2^*\right) > \frac{f_3}{f_2+f_3}(c_1 - c_2)\, x_1^*$ and $\lambda = 0$, since $\lim_{c_3 \to 0} x_3^* = \infty$. On

For simplicity of the analysis, we will assume $x_2^* \geq x_1^*$, so that constraint (27) is not binding and $\{x_n^*\}_{n=1,2,3}$ is the optimal implementable output profile. Except for the last terms in their right-hand side, equations (28) to (30) are identical to the first-order equations that would determine $\{x_n^{ns}\}$, the no-supervision optimal output levels. Therefore, the effect of the supervision on the optimal output levels can be identified by examining these last terms. We will define $\left\{x_n^{fb}\right\}$ as the profile of "first best" output levels, where $W'(x_n^{fb}) = c_n$.

If $\lambda < 1$ and therefore constraint (25) is binding, then equation (30) implies $x_1^{ns} < x_1^* < x_1^{fb}$. That is, under supervision the optimal output level for type 1 is distorted less relative to its no-supervision optimal level. Recall that constraint (25) was derived from the participation constraint of S for the state where $n = 3$. Since the surplus for S is $(c_1 - c_2)\, x_1 - \frac{f_2+f_3}{f_3}\Pi$ in that state of nature, an increase in x_1 relaxes constraint (25). Therefore, increasing x_1 is not as costly as it had been under no-supervision.

If $\lambda > 0$ and therefore constraint (26) is binding, then equations (29) and (28) imply that the optimal output levels for types 2 and 3 are further distorted relative to their no-supervision levels: $x_2^* < x_2^{ns} < x_2^{fb}$ and $x_3^* > x_3^{ns} = x_3^{fb}$. Constraint (26) determines the bound on the difference between the reservation utilities for types 2 and 3 at the collusion stage, given the output levels they are supposed to produce. By increasing the difference between the output levels, P can relax this bound. Therefore, x_2^* is lower and x_3^* is higher than their respective no-supervision optimal levels.[36]

The closer the realized output level to its first-best value, the more efficient the overall economy would be. To determine the effect of supervision on the social welfare, we should compare the output distortions induced under supervision with the distortions associated with the no-supervision benchmark. If $\lambda = 0$, x_1^* is the only output level that will be different than its no-supervision level. Since its value is closer to its first-best level, the optimal supervision outcome creates a higher social welfare than the optimal no-supervision outcome, whenever $\lambda = 0$. However, if $\lambda = 1$, output levels x_2^* and x_3^* are moving in socially undesirable directions. Therefore, there is a social welfare loss with the introduction of supervision in this latter case. For the intermediate values of λ, the overall effect on social welfare is ambiguous.

The possibility that supervision can decrease social welfare is another differentiating aspect of this paper. In models of exogenous transaction costs[37] or in models using asymmetric information to justify the transaction cost approach[38] supervision would always increase the overall efficiency. To see this, recall that the cost of acquiring the relevant information is the only reason for output distortions in

---

the other hand, if c_3 is higher than a threshold level, then $(c_2 - c_3)\left(x_3^* - x_2^*\right) < \frac{f_3}{f_2+f_3}\,(c_1 - c_2)\, x_1^*$ and $\lambda = 1$, since $\lim_{c_3 \to c_2}\left(x_3^* - x_2^*\right) = 0$.

[36]The upward distortion for the most productive type is a common feature of models that employ countervailing incentives.

[37]Such as Laffont and Tirole (1991), Laffont and Martimort (1999), Laffont and Meleu (2001).

[38]Such as Faure-Grimaud, Laffont, and Martimort (2000 and 2001).

adverse selection models. Under the transaction cost assumption, supervision reduces to a technology to provide the information to P at a cost that is lower than the cost of getting the information directly from the agent. In other words, the information rent P has to leave is scaled down with respect to the no-supervision benchmark. Therefore, the output distortions P will impose will be smaller than their no-supervision levels.

On the other hand, in our setup, P sustains the collusion failure by manipulating the type specific information rent levels for A. This manipulation has to respect the incentive compatibility constraints of the agent types within the same information set. By changing the type specific output levels, P can relax these constraints. However, this change might be in the direction of distorting the output levels further from their first best levels. For example, by increasing the output level of type 3 (over its first best and no-supervision level), P can increase that type's information rent without worrying about type 2 trying to imitate type 3. Although such a distortion on the output levels proves to be beneficial for P, it would induce a loss in overall welfare.

7.2 Delegation when Monotonicity Constraints are Binding

In section 4, we established a result regarding the failure of delegation as a rent extraction tool for P whenever the monotonicity constraints of the no-supervision problem are slack. In this subsection, we will use our three type model to analyze how the performance of delegation reacts to the existence of some binding monotonicity constraints.

Proposition 7 *Assume that* $f_2(c_1 - c_2) < f_3(c_2 - c_3)$. *A non-monotonic output profile* $\{x_n\}_{n=1,2,3}$, *such that* $x_3 \geq x_1 > x_2$ *is delegation implementable together with the following* $\{u_n, r_n\}_{n=1,2,3}$:

$$\begin{aligned} u_1 &= 0 & r_1 &= 0 \\ u_2 &= (c_1 - c_2) x_1 - \tfrac{f_3}{f_2}(c_2 - c_3)(x_1 - x_2) & r_2 &= 0 \\ u_3 &= (c_1 - c_2) x_1 + (c_2 - c_3) x_1 & r_3 &= (c_2 - c_3) x_2 \end{aligned}$$

The proof for the proposition is in the appendix. This proposition demonstrates the implementability of a non-monotonic output profile through delegation. It also reveals some information about the desirability of such an implementation for P. With the outcome above, the expected information rent P has to pay is $f_3 u_3 + f_2 u_2 + f_1 u_1$, which is identical to $f_3 u_3^0(\{x_n\}) + f_2 u_2^0(\{x_n\}) + f_1 u_1^0(\{x_n\})$. Therefore, this leaves P with the same objective function as in the no-supervision problem (23). When the monotonicity constraint $x_2 \geq x_1$ is not binding for (23), implementing a non-monotonic profile through the above procedure will not be desirable for P. This is also consistent with the statement of Proposition 3. However, if $x_2 \geq x_1$ is a binding constraint, the above outcome may improve P's payoff since P does not have to respect that constraint any more. Once S is present, output levels need only be monotonic

within each information set of S, but not across them. This introduces a channel for delegation to improve upon no-supervision.

8 Conclusion

In this paper, we have provided 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 amounts to a special class of grand contracts which are not responsive to the agent's report. Under delegation, the outside option for collusion becomes the shut down of production and therefore the principal loses his power to manipulate the type dependent opportunity cost of collusion. As a result, delegation performs as badly as no-supervision for the principal as long as the monotonicity constraints of the no-supervision implementation problem are slack.

The increase in the principal's payoff does not necessarily indicate an overall efficiency gain for the society. We show that social welfare may decline with the introduction of the supervisor.

By modeling collusion as a take it or leave it offer from the supervisor, we assumed that the supervisor is the party with all of the bargaining power at the collusion stage. Alternatively, we could let the agent or a benevolent outsider make the collusion offer. In either case, informational asymmetry is still a source of coalitional inefficiency[39], and our failure and relevance results are still valid under these alternative formulations of collusion.

[39]In case that the agent is the party who is making the collusion offer, his problem can be regarded as the problem of an informed mechanism designer who could signal his information with the offer he makes. See Maskin and Tirole (1990, 1992) on this.

9 Appendix

9.1 Proof for Proposition 1 (Characterization of Collusion Feasible Outcomes)

- **Necessity**

Let $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$ be a collusion feasible outcome. From the definition of collusion feasibility, we know that there exists a grand contract $GC = \{M_S, M_A, x(\cdot), t(\cdot), w(\cdot)\}$, and a collusive equilibrium of GC, $\{\mu(\cdot), \beta(\cdot)\}$, that attains $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$.

$\{\mu(n), \beta(n)\}_{n\in d_l}$ must be a solution to the maximization problem stated in the definition of collusive equilibrium. Consider the following modification of the problem, where the objective function is the same but the choice set is further constrained:

$$\max_{\{m(n),b(n)\}_{n\in d_l}} \left\{ \sum_{n\in d_l} f_n \left[w(m(n)) + b(n) \right] \right\} \text{ s.t.}$$

$$t(m(n)) - b(n) - c_n x(m(n)) \geq t(m(n')) - b(n') - c_n x(m(n')) \text{ for all } n, n' \in d_l \quad (3)$$

$$t(m(n)) - b(n) - c_n x(m(n)) \geq t(\mu(n)) - \beta(n) - c_n x(\mu(n)) \text{ for all } n \in d_l \quad (31)$$

$$m(n) \in \{\mu(n') : n' \in \mathbb{N}\} \text{ for all } n \in d_l \quad (32)$$

Constraint (32) dictates that any message pair suggested by the supervisor should be the equilibrium messages for some type. And constraint (31) is a further strengthening of (4). Since $\{\mu(n), \beta(n)\}_{n\in d_l}$ satisfies all the additional constraints, it is an optimal solution of this modified version of the problem as well. Given $\{\mu(n), \beta(n)\}_{n\in d_l}$ and any $\{m(n), b(n)\}_{n\in d_l}$ which satisfies (32), we can define functions $\{\hat{n}(n), \hat{r}(n)\}_{n\in d_l}$ such that:[40]

$$\hat{n}(n) : \mu(\hat{n}(n)) = m(n)$$

$$\hat{r}(n) : \hat{r}(n) = t(m(n)) - c_n x(m(n)) - b(n)$$

Note that if $\{m(n), b(n)\}_{n\in d_l} = \{\mu(n), \beta(n)\}_{n\in d_l}$, then $\hat{n}(n) = n$ and $\hat{r}(n) = t(\mu(n)) - \beta(n) - c_n x(\mu(n)) = r_n$. With a change of the choice variables, we can rewrite the modified optimization problem above:

$$\max_{\{\hat{n}(n),\hat{r}(n)\}_{n\in d_l}\in\{\mathbb{N}\times\Re\}^{\#d_l}} \sum_{n\in d_l} f_n \left[u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n \right) x_{\hat{n}(n)} - \hat{r}(n) \right] \text{ s.t.}$$

$$\mathbf{AIC}(n'|n) : \hat{r}(n) \geq \hat{r}(n') + (c_{n'} - c_n) x_{\hat{n}(n')} \text{ for all } n, n' \in d_l$$

$$\mathbf{AIR}(n) : \hat{r}(n) \geq r_n \text{ for all } n \in d_l$$

[40] If $\{\mu(\cdot), \beta(\cdot)\}$ instructs to send the same messages for different states of the nature, there may be more than one n' that satisfies the equation $\mu(n') = m(n)$. In that case the function $\hat{n}(n)$ can assume the value of any such n'.

where $\{n, r_n\}_{n \in d_l}$ is an optimal solution. Since this is true for any arbitrary l, any collusion feasible outcome satisfies condition (5).

- **Sufficiency**

Suppose $\{x_n, u_n, r_n\}_{n \in \mathbb{N}}$ satisfies condition (5). We have to provide a grand contract, a non-cooperative equilibrium, and a collusive equilibrium that would induce $\{x_n, u_n, r_n\}_{n \in \mathbb{N}}$. As the grand contract, we will use $GC^* = \{M_S^*, M_A^*, x^*(\cdot), t^*(\cdot), w^*(\cdot)\}$ such that $M_A^* = \mathbb{N}$, $M_S^* = \mathbb{L}$, and

$$x^*\left(\hat{l}, \hat{n}\right) = \begin{cases} x_{\hat{n}}, \text{ if } \hat{l} = l(\hat{n}) \\ x_{\underline{n}(\hat{l})}, \text{ if } \hat{l} > l(\hat{n}) \\ x_{\overline{n}(\hat{l})}, \text{ if } \hat{l} < l(\hat{n}) \end{cases}$$

$$t^*\left(\hat{l}, \hat{n}\right) = \begin{cases} r_{\hat{n}} + c_{\hat{n}} x_{\hat{n}}, \text{ if } \hat{l} = l(\hat{n}) \\ r_{\underline{n}(\hat{l})} + c_{\underline{n}(\hat{l})} x_{\underline{n}(\hat{l})}, \text{ if } \hat{l} > l(\hat{n}) \\ r_{\overline{n}(\hat{l})} + c_{\overline{n}(\hat{l})} x_{\overline{n}(\hat{l})}, \text{ if } \hat{l} < l(\hat{n}) \end{cases}$$

$$w^*\left(\hat{l}, \hat{n}\right) = \begin{cases} u_{\hat{n}} - r_{\hat{n}}, \text{ if } \hat{l} = l(\hat{n}) \\ \underline{w}, \text{ if } \hat{l} \neq l(\hat{n}) \end{cases}$$

where $\underline{w}$ is a real number smaller than $u_n - r_n$ for all n. Also recall $\underline{n}(l) = \min\{n \in d_l\}$ and $\overline{n}(l) = \max\{n \in d_l\}$ are the least and the most productive types in information set d_l respectively.

GC^* is a direct contract, where the players are asked to reveal their information to P. If both S and A respond truthfully, GC^* leads to the outcome $\{x_n, u_n, r_n\}_{n \in \mathbb{N}}$. Suppose S and A send messages that contradict, i.e., $l(\hat{n}) \neq \hat{l}$. In that case, GC^* treats A as if he is a type in information set $d_{\hat{l}}$. Among the output - transfer pairs consistent with $d_{\hat{l}}$, A is assigned to the one that maximizes type $\hat{n}$'s utility. Whenever the messages contradict, S is punished by being paid $\underline{w}$ as well.

We will show that the truthful behavior, $\{\sigma^*(l) = l, \alpha^*(n) = n\}$, is a non-cooperative equilibrium of GC^*. Moreover, the truthful behavior also supports a truthful collusive equilibrium, where the messages reveal the information that the coalition has and there is no side transfer within the coalition. Formally, $\{\mu^*(\cdot), \beta^*(\cdot)\}$ is a collusive equilibrium supported by $\{\sigma^*(\cdot), \alpha^*(\cdot)\}$, where

$$\mu^*(n) = (l(n), n)$$
$$\beta^*(n) = 0$$

for all n.

i) $\{\sigma^*(\cdot), \alpha^*(\cdot)\}$ is a non-cooperative equilibrium of GC^*.

First, we will show that A's truthful revelation of his type is weakly dominant. Here is the payoff for

type n agent as a function of the messages, $\hat{n}$ and $\hat{l}$:

$$\begin{aligned} r_{\hat{n}} + (c_{\hat{n}} - c_n)\, x_{\hat{n}}, \text{ if } \hat{l} &= l(\hat{n}) \\ r_{\overline{n}(\hat{l})} + \left(c_{\overline{n}(\hat{l})} - c_n\right) x_{\overline{n}(\hat{l})}, \text{ if } \hat{l} &< l(\hat{n}) \\ r_{\underline{n}(\hat{l})} + \left(c_{\underline{n}(\hat{l})} - c_n\right) x_{\underline{n}(\hat{l})}, \text{ if } \hat{l} &> l(\hat{n}) \end{aligned}$$

If S behaves truthful and sends $\hat{l} = l(n)$ to P, optimality of A's truthful revelation ($\alpha(n) = n$) follows from $\mathbf{AIC}(n'|n)$ for all $n' \in d_l(n)$ at $\{\hat{n}(n), \hat{r}(n)\}_{n\in\mathbb{N}} = \{n, r_n\}_{n\in\mathbb{N}}$. If S sends $\hat{l} < l(n)$, agent type n can secure any utility level in the form of $r_{n'} + (c_{n'} - c_n)\, x_{n'}$, where $n' \in d_{\hat{l}}$. From the constraints $\mathbf{AIC}\left(n'|\overline{n}\left(\hat{l}\right)\right)$ and $\mathbf{AIC}\left(\overline{n}\left(\hat{l}\right)|n'\right)$, we know that $x_{\overline{n}(\hat{l})} \geq x_{n'}$ for all such n'. This fact, together with $\mathbf{AIC}\left(n'|\overline{n}\left(\hat{l}\right)\right)$ and $c_n < c_{n'}$ imply $\overline{n}\left(\hat{l}\right) \in \arg\max_{n'\in d_{\hat{l}}} \{r_{n'} + (c_{n'} - c_n)\, x_{n'}\}$. Note that by sending his true type to P, type n agent can achieve his maximum possible payoff under the assumption that S sends $\hat{l}$. So $\alpha(n) = n$ is still a best response to S's message. With a symmetric argument, same is true for $\hat{l} > l(n)$. Therefore, the strategy $\alpha(n) = n$ is weakly dominant.

Given $\alpha(n) = n$, S gets a payoff of $u_n - r_n$ if she sends l, or $\underline{w}$ if she sends another message. By construction, $\underline{w}$ is smaller than $u_n - r_n$ for all n. Therefore, the truthful strategy, $\sigma(l) = l$, is optimal in the ex post sense.

ii) $\{\mu^*(\cdot), \beta^*(\cdot)\}$ is a collusive equilibrium of GC^* (supported by $\{\sigma^*(\cdot), \alpha^*(\cdot)\}$).

First, note that $\{\mu^*(n), \beta^*(n)\}_{n\in d_l}$ satisfies the constraints (3) and (4) for all l. The first constraint is identical to **AIC** at $\{\hat{n}(n), \hat{r}(n)\}_{n\in\mathbb{N}} = \{n, r_n\}_{n\in\mathbb{N}}$, and the second one is a tautology. For $\{\mu^*(\cdot), \beta^*(\cdot)\}$ not to be a collusive equilibrium, there must exist $\tilde{l}$ and $\{m(n), b(n)\}_{n\in d_{\tilde{l}}}$, that would satisfy constraints (3) and (4), and also give a higher value for the objective function, i.e.,

$$\sum_{n\in d_{\tilde{l}}} f_n\left[w(m(n)) + b(n)\right] > \sum_{n\in d_{\tilde{l}}} f_n\left[w(\mu^*(n)) + \beta^*(n)\right] \tag{33}$$

Let $\{\tilde{x}(n), \tilde{u}(n), \tilde{r}(n)\}_{n\in d_{\tilde{l}}}$ be the production and rent division rule following $\{m(n), b(n)\}_{n\in d_{\tilde{l}}}$:

$$\begin{aligned} \tilde{x}(n) &= x^*(m(n)) \\ \tilde{u}(n) &= t^*(m(n)) + w^*(m(n)) - c_n x^*(m(n)) \\ \tilde{r}(n) &= t^*(m(n)) - c_n x^*(m(n)) - b(n) \end{aligned}$$

when we substitute these and $\{x_n, u_n, r_n\}_{n\in d_{\tilde{l}}}$ in inequality (33),

$$\sum_{n\in d_{\tilde{l}}} f_n\left[\tilde{u}(n) - \tilde{r}(n)\right] > \sum_{n\in d_{\tilde{l}}} f_n\left[u_n - r_n\right] \tag{34}$$

Notice that GC^* allows for only the output levels that are induced by $\{x_n, u_n, r_n\}_{n\in\mathbb{N}}$. For the S - A coalition to get the output level $\tilde{x}(n)$, there should exist $n' \in \mathbb{N}$ such that $\tilde{x}(n) = x_{n'}$. Moreover,

the highest gross transfer (that is transfer plus wage) that the coalition can get by production of $x_{n'}$ is $u_{n'} + c_{n'} x_{n'}$. Therefore, for every $n \in d_{\tilde{l}}$, there exists $\tilde{n}(n) \in \mathbb{N}$ such that:

$$\begin{aligned} \tilde{x}(n) &= x_{\tilde{n}(n)} \\ \tilde{u}(n) &\leq u_{\tilde{n}(n)} + \left(c_{\tilde{n}(n)} - c_n\right) x_{\tilde{n}(n)} \end{aligned}$$

From this fact and inequality (34),

$$\sum_{n \in d_{\tilde{l}}} f_n \left[u_{\tilde{n}(n)} + \left(c_{\tilde{n}(n)} - c_n\right) x_{\tilde{n}(n)} - \tilde{r}(n)\right] > \sum_{n \in d_{\tilde{l}}} f_n \left[u_n - r_n\right] \tag{35}$$

By construction of $\{m(n), b(n)\}_{n \in d_{\tilde{l}}}$, constraints (3) and (4) are satisfied. Also note that the same constraints are identical to **AIC** and **AIR** at $\{\hat{n}(n), \hat{r}(n)\}_{n \in d_{\tilde{l}}} = \{\tilde{n}(n), \tilde{r}(n)\}_{n \in d_{\tilde{l}}}$. Then, (35) is a contradiction to (5). Therefore, $\{\mu^*(\cdot), \beta^*(\cdot)\}$ is a collusive equilibrium of GC^*.

iii) $\{x_n, u_n, r_n\}_{n \in \mathbb{N}}$ is a collusion feasible outcome.

This follows from the definition of collusion feasibility and the equations below:

$$\begin{aligned} x_n &= x^*[\mu^*(n)] \\ u_n &= t^*[\mu^*(n)] + w^*[\mu^*(n)] - c_n x^*[\mu^*(n)] \\ r_n &= t^*[\mu^*(n)] - \beta^*(n) - c_n x^*[\mu^*(n)] \end{aligned}$$

9.2 Proof for Lemma 1

Equation (10) states that the least productive type in an information set receives his reservation value. The utility for the other types are determined by downward adjacent **AIC** constraints. If r_n was smaller than that is specified in (10), either a $\mathbf{d-AIR}$ or a downward adjacent **AIC** constraint would be violated. If r_n was larger instead, S could decrease it without violating any **AIC** or $\mathbf{d-AIR}$ constraints.

To prove (11), consider $n, n-1 \in d_l$. Suppose S offers a side contract that misreports type n as type $n-1$. Formally, consider $\{\hat{n}(i), \hat{r}(i)\}_{i \in d_l}$ such that $\hat{n}(n) = n-1$, $\hat{n}(i) = i$ for all $i \neq n$, and $\hat{r}(i) = r_i$ for all i. **AIC** and $\mathbf{d-AIR}$ constraints at (6) defined for l are satisfied.[41] The change in the objective function is $f_n[u_{n-1} + (c_{n-1} - c_n) x_{n-1} - u_n]$. For $\{\hat{n}(i), \hat{r}(i)\}_{i \in d_l}$ not to raise the objective function, (11) must hold.

To prove (12), consider (6) for d_l again. Suppose S offers a side contract that misreports all types in d_l as $\overline{n}(l-1)$. Formally, consider $\{\hat{n}(n), \hat{r}(n)\}_{n \in d_l}$ such that $\hat{n}(n) = \overline{n}(l-1)$ and $\hat{r}(n) =$

[41] $\mathbf{d-AIR}$ are intact since there is no change in r_i. Since a delegation feasible outcome is also a collusion feasible one, **WPM** holds for $i \in d_l$. Under **WPM**, **AIC**$(n|i)$ is intact, if $i < n$ and relaxed otherwise.

$\left(c_{\underline{n}(l)} - c_n\right) x_{\overline{n}(l-1)}$, for all $n \in d_{\tilde{l}}$. **AIC** and $\mathbf{d-AIR}$ constraints at (6) defined for l are satisfied.[42] This deviation gives $\sum_{n\in d_l} f_n \left[u_{\overline{n}(l-1)} + \left(c_{\overline{n}(l-1)} - c_{\underline{n}(l)}\right) x_{\overline{n}(l-1)}\right]$ as the value of the objective function.[43] For this value to be smaller than the original one,

$$\sum_{n\in d_l} f_n \left(u_n - r_n\right) \geq \sum_{n\in d_l} f_n \left[u_{\overline{n}(l-1)} + \left(c_{\overline{n}(l-1)} - c_{\underline{n}(l)}\right) x_{\overline{n}(l-1)}\right]$$

which is identical to (12).

9.3 Proof for Lemma 2

- **Step 1: (13) is satisfied for $l = 1$.**

From (10) and the construction of $u_n^o(\cdot)$, $u_n^o(\{x_n\}) = r_n$ for $n \in d_1$. And from the participation constraint of S for $l = 1$, $\sum_{n\in d_1} f_n (u_n - r_n) \geq 0$.

- **Step 2: If (13) is satisfied for l, then $u_{\overline{n}(l)} \geq u^o_{\overline{n}(l)}(\{x_n\})$.**

First observe that we can rewrite (10) as

$$r_n - r_{n'} = u_n^o(\{x_n\}) - u_{n'}^o(\{x_n\}) \text{ for } n, n' \in d_l \tag{36}$$

The hypothesis of the claim requires the existence of $k \in d_l$ such that $u_k \geq u_k^o(\{x_n\})$. From (11) and (36),

$$\begin{aligned} u_{\overline{n}(l)} - u_k &\geq r_{\overline{n}(l)} - r_k \\ u_{\overline{n}(l)} - u_k &\geq u^o_{\overline{n}(l)}(\{x_n\}) - u_k^o(\{x_n\}) \\ u_{\overline{n}(l)} - u^o_{\overline{n}(l)}(\{x_n\}) &\geq u_k - u_k^o(\{x_n\}) \geq 0 \end{aligned} \tag{37}$$

We need one more step to start an iteration:

- **Step 3: If $u_{\overline{n}(l-1)} \geq u^o_{\overline{n}(l-1)}(\{x_n\})$, then (13) holds for l.**

The hypothesis of the claim and (12) imply

$$\sum_{n\in d_l} f_n u_n \geq \sum_{n\in d_l} f_n \left[u^o_{\overline{n}(l-1)}(\{x_n\}) + \left(c_{\overline{n}(l-1)} - c_{\underline{n}(l)}\right) x_{\overline{n}(l-1)} + r_n\right] \tag{38}$$

From (10) and construction of the function $u_n^o(\cdot)$:

$$u^o_{\overline{n}(l-1)}(\{x_n\}) + \left(c_{\overline{n}(l-1)} - c_{\underline{n}(l)}\right) x_{\overline{n}(l-1)} + r_n = u_n^o(\{x_n\}). \tag{39}$$

When we substitute this into the right-hand side of (38), we get (13).

Step 1 shows that (13) holds for $l = 1$. Steps 2 and 3 show that if (13) holds for $l-1$, it holds for l as well. By iteration, we conclude that (13) is satisfied by all l.

[42] $\{\hat{n}(n), \hat{r}(n)\}_{n\in d_l}$ is constructed as a result of a side contract that offers a single output - transfer pair, rather than a menu of such pairs. **AIC** constraints follow from this fact. $\mathbf{d-AIR}$ are satisfied, since $\hat{r}(n)$ is non-negative for all n.

[43] Note that the surplus for S is uniform over d_l.

9.4 Proof for Proposition 5 (Relevance)

We need to show that a set of weakly monotonic output levels, together with the information rent and utility levels specified in (14) and (15) are implementable given conditions (16) to (19).

The ex-post participation constraints, (7) and (9), are satisfied since u_n^o is increasing in n and (16) holds.[44] To establish implementability we should also show that the outlined outcome is collusion feasible, and therefore satisfies condition (5). At $\{\hat{n}(n), \hat{r}(n)\}_{n\in\mathbb{N}} = \{n, r_n\}_{n\in\mathbb{N}}$, **AIR** constraints are tautologically satisfied. For $n \notin d_{\tilde{l}}$, $\mathbf{AIC}(n'|n)$ follows from $\mathbf{IC}(n'|n)$.[45] For $n \in d_{\tilde{l}}$, and $n > k$, constraint $\mathbf{IC}(n'|n)$ and $\Psi > 0$ imply $\mathbf{AIC}(n'|n)$. For $n \in d_{\tilde{l}}$, and $n, n' \leq k$, $\mathbf{IC}(n'|n)$ is again sufficient for $\mathbf{AIC}(n'|n)$. However, for $n \in d_{\tilde{l}}$ and $n \leq k < n'$, we need a more detailed analysis:

We can write the $\mathbf{AIC}(n'|n)$ constraint as

$$(c_n - c_{n'})\, x_{n'} \geq r_{n'} - r_n. \tag{40}$$

Recall that $n \leq k$. From the construction of $\{r_n\}_{n\in d_l}$, $\mathbf{AIC}(n'|n)$ can be rewritten as

$$\begin{aligned} (c_n - c_{n'})\, x_{n'} &\geq (u_{n'}^o - u_n^o) + \Psi \\ (c_n - c_{n'})\, x_{n'} &\geq \left(u_{n'}^o - u_{k+1}^o\right) + \left(u_{k+1}^o - u_k^o\right) + (u_k^o - u_n^o) + \Psi \end{aligned} \tag{41}$$

From $\mathbf{IC}(n'|k+1)$, Proposition 2[46], and $\mathbf{IC}(k|n)$, we can derive a sufficient condition for (41):

$$\begin{aligned} (c_n - c_{n'})\, x_{n'} &\geq (c_{k+1} - c_{n'})\, x_{n'} + (c_k - c_{k+1})\, x_k + (c_n - c_k)\, x_k + \Psi \\ (c_n - c_{k+1})\,(x_{n'} - x_k) &\geq \Psi \end{aligned} \tag{42}$$

which is implied by (17), since $c_n \geq c_k$, and $x_{n'} \geq x_{k+1}$.

Therefore, the constraints of the programs in (5) are satisfied at $\{\hat{n}(n), \hat{r}(n)\}_{n\in\mathbb{N}} = \{n, r_n\}_{n\in\mathbb{N}}$. Thus, the only reason for (5) to fail could be the existence of some l and $\{\hat{n}(n), \hat{r}(n)\}_{n\in d_l}$ that would satisfy the constraints **AIC**, **AIR** and give a higher value for the objective function, i.e.,

$$\sum_{n\in d_l} f_n \left[u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} - u_n\right] - \sum_{n\in d_l} f_n \left[\hat{r}(n) - r_n\right] > 0. \tag{43}$$

Given **AIR** constraints, $\sum_{n\in d_l} f_n \left[\hat{r}(n) - r_n\right]$ is non-negative. Therefore, a necessary condition to induce a higher value for the objective function is the existence of a room for the coalitional efficiency. Formally, we need n such that

$$u_n < u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)}, \text{ where } \hat{n}(n) \in \mathbb{N}. \tag{44}$$

[44] In this section we will drop the argument of function $u_n^o(\cdot)$, to ease the notation.

[45] Recall that **IC** constraints of the no supervision implementation are satisfied by $\{u_n^o\}$.

[46] Proposition 2 implies $u_{k+1}^o = u_k^o + (c_k - c_{k+1})\, x_k$.

The **IC** constraints on $\{u_n^o\}$ and $\Delta \geq 0$ imply that there exists no such n, that is weakly smaller than k. For $n > k$, we can rewrite (44) as

$$\left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} > u_n - u_{\hat{n}(n)}. \tag{45}$$

From the construction of $\{u_n\}_{n \in d_l}$, a necessary condition for (45) is

$$\begin{aligned} \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} &> u_n^o - u_{\hat{n}(n)}^o - \Delta \\ \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} &> \left(u_n^o - u_{k+1}^o\right) + \left(u_{k+1}^o - u_{\hat{n}(n)}^o\right) - \Delta. \end{aligned} \tag{46}$$

From $\mathbf{IC}\left(k+1|n\right)$ and $\mathbf{IC}\left(\hat{n}\left(n\right)|k+1\right)$, we can also derive a necessary condition for (46):

$$\begin{aligned} \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} &> \left(c_{k+1} - c_n\right) x_{k+1} + \left(c_{\hat{n}(n)} - c_{k+1}\right) x_{\hat{n}(n)} - \Delta \\ \Delta &> \left(c_{k+1} - c_n\right)\left(x_{k+1} - x_{\hat{n}(n)}\right) \end{aligned} \tag{47}$$

Given condition (18), two necessary conditions for the inequality (47) are $n = k+1$ and $\hat{n}\left(k+1\right) \leq k$. Since $k+1 \in d_{\tilde{l}}$, the only information set that has the potential to support a coalitionally efficient deviation is $d_{\tilde{l}}$. And such a deviation $\{\hat{n}\left(n\right), \hat{r}\left(n\right)\}_{n \in d_{\tilde{l}}}$ would require $\hat{n}\left(k+1\right) \leq k$.

Rewrite (43) for $\tilde{l}$:

$$\sum_{n \in d_{\tilde{l}}} f_n \left[u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} - u_n\right] > \sum_{n \in d_{\tilde{l}}} f_n \left[\hat{r}\left(n\right) - r_n\right] \tag{48}$$

From the argument above, an upper bound for the left-hand side is $f_{k+1}\Delta$. We can also construct a lower bound for the right-hand side too. Since $\{\hat{n}\left(n\right), \hat{r}\left(n\right)\}_{n \in d_{\tilde{l}}}$ satisfies the constraints of (5), $\mathbf{AIC}\left(k+1|k\right)$ implies

$$\hat{r}\left(k\right) \geq \hat{r}\left(k+1\right) + \left(c_{k+1} - c_k\right) x_{\hat{n}(k+1)}. \tag{49}$$

We will substitute in $\mathbf{AIR}\left(k+1\right)$, invoke the fact that $\hat{n}\left(k+1\right) \leq k$, and subtract r_k from each side:

$$\hat{r}\left(k\right) - r_k \geq \left(r_{k+1} - r_k\right) + \left(c_{k+1} - c_k\right) x_k \tag{50}$$

It follows from the construction of $\{r_n\}_{n \in \mathbb{N}}$ that

$$\hat{r}\left(k\right) - r_k \geq u_{k+1}^o - u_k^o + \Psi + \left(c_{k+1} - c_k\right) x_k. \tag{51}$$

And finally, since $u_{k+1}^o = u_k^o + \left(c_k - c_{k+1}\right) x_k$,

$$\hat{r}\left(k\right) - r_k \geq \Psi. \tag{52}$$

Therefore, a lower bound for $\sum_{n \in d_{\tilde{l}}} f_n \left[\hat{r}\left(n\right) - r_n\right]$ is $f_k \Psi$. Condition (19) implies that (48) cannot be satisfied with $\{\hat{n}\left(n\right), \hat{r}\left(n\right)\}_{n \in d_{\tilde{l}}}$ that does not violate **AIC** and **AIR**.

9.5 Proof for Proposition 6

We already know that the outcome suggested by the proposition is implementable and it improves over the no-supervision optimal solution. Under the monotone hazard rate assumption, a necessary condition for improving over the no-supervision optimal solution is satisfying the following inequality:

$$\begin{aligned} f_3 u_3 + f_2 u_2 &< f_3 u_3^o(\{x_n\}) + f_2 u_2^o(\{x_n\}) \\ &= (f_3 + f_2)(c_1 - c_2) x_1 + f_3 (c_2 - c_3) x_2 \end{aligned} \tag{53}$$

Therefore, an upper bound for P's payoff with implementable outcomes that satisfy (53) will also be a global upper bound on P's payoff with any implementable outcome.

Let $\{x_n, u_n, r_n\}_{n=1,2,3}$ be an implementable outcome that satisfies (53). Assume $k \in \{1,2\}$. And consider the following deviation $\{\hat{n}(n), \hat{r}(n)\}_{n=2,3}$ to $\{n, r_n\}_{n=2,3}$, where both types 3 and 2 are pooled with type k, such that $\hat{n}(3) = \hat{n}(2) = k$. For the **AIC** constraints to hold, it must be $\hat{r}(3) = \hat{r}(2) + (c_2 - c_3) x_k$. By setting $\hat{r}(2) = \max\{r_2, r_3 - (c_2 - c_3) x_k\}$, we also satisfy the **AIR** constraints. For this deviation not to increase the value for the objective function:

$$\begin{aligned} f_2(u_2 - r_2) + f_3(u_3 - r_3) &\geq (f_2 + f_3)[u_k + (c_k - c_2) x_k - \hat{r}(2)] \\ f_2(u_2 - r_2) + f_3(u_3 - r_3) &\geq (f_2 + f_3)[u_k + (c_k - c_2) x_k - \max\{r_2, r_3 - (c_2 - c_3) x_k\}] \end{aligned} \tag{54}$$

Now we will argue that $r_3 - (c_2 - c_3) x_k \geq r_2$. The statement is immediate from **WPIC** $(2|3)$ for $k = 2$. As for $k = 1$, suppose $r_3 - (c_2 - c_3) x_1 < r_2$. Then inequality (54) can be written for $k = 1$ as:

$$\begin{aligned} f_2 u_2 + f_3 u_3 &\geq (f_2 + f_3)(u_1 + (c_1 - c_2) x_1) + f_3 (r_3 - r_2) \\ f_2 u_2 + f_3 u_3 &\geq (f_2 + f_3)(u_1 + (c_1 - c_2) x_1) + f_3 (c_2 - c_3) x_2 \end{aligned} \tag{55}$$

where the last inequality is attained by substituting in **WPIC** $(2|3)$. Since $u_1 \geq 0$ (this follows from the participation constraints), this is a contradiction to (53). Therefore,

$$r_3 - (c_2 - c_3) x_k \geq r_2. \tag{56}$$

Now, consider another deviation for S, where $\hat{n}(2) = 1$, $\hat{r}(2) = r_2$, $\hat{n}(3) = 3$ and $\hat{r}(3) = r_3$. That is, she misreports type 2 as type 1 to P and adjusts the bribe such that type 2 is indifferent to this deviation. Since $\hat{r}(n) = r_n$, **AIR** constraints are intact. Given (56) for $k = 1$, **AIC** constraints hold as well. With this deviation, the change in the value for the objective function is $f_2(u_1 + (c_1 - c_2) x_1 - u_2)$. Since $u_1 \geq 0$, for the change in the objective function not to be positive, it must be that

$$u_2 \geq u_1 + (c_1 - c_2) x_1. \tag{57}$$

Given (56), we can rewrite (54) as:

$$f_2 u_2 + f_3 u_3 \geq (f_2 + f_3)(u_k + (c_k - c_2) x_k + (c_2 - c_3) x_k) - f_2 (r_3 - r_2) \tag{58}$$

By substituting in (57) and $u_1 \geq 0$, we conclude that

$$f_2 u_2 + f_3 u_3 \geq (f_2 + f_3)\left((c_1 - c_2) x_1 + (c_2 - c_3) x_k\right) - f_2 (r_3 - r_2). \tag{59}$$

Inequality (59) holds for both values of k. We can write both of these inequalities as follows:

$$f_2 u_2 + f_3 u_3 \geq (f_2 + f_3)\left((c_1 - c_2) x_1 + (c_2 - c_3) \bar{x}\right) - f_2 (r_3 - r_2) \tag{60}$$

where $\bar{x} = \max\{x_1, x_2\}$. The bound that is established on $f_2 u_2 + f_3 u_3$ by inequality (60) is decreasing in $(r_3 - r_2)$. We will derive two upper bounds on the value of $(r_3 - r_2)$, too. The first one is

$$\textbf{WPIC}\,(3|2) : r_3 - r_2 \leq (c_2 - c_3) x_3.$$

The other results from participation constraints:

$$\begin{aligned} r_2 &\geq 0 \\ u_3 - r_3 &\geq 0 \end{aligned}$$

When we substitute these into (60), we get

$$f_3 u_3 + f_2 u_2 \geq \left\{ \begin{array}{c} f_3\left[(c_1 - c_2) x_1 + (c_2 - c_3) \bar{x}\right] + f_2 (c_1 - c_2) x_1 \\ -f_2 \min\left\{ \frac{f_3}{f_3 + f_2} (c_1 - c_2) x_1, (c_2 - c_3)(x_3 - \bar{x}) \right\} \end{array} \right\} \tag{61}$$

Recall that (61) is satisfied for any implementable outcome for which (53) holds. Therefore, solution to the following maximization problem is an upper bound on P's expected welfare

$$\max_{\{x_n, u_n, r_n\}_{n=1,2,3}} f_3\left[W(x_3) - c_3 x_3\right] + f_2\left[W(x_2) - c_2 x_2\right] + f_1\left[W(x_1) - c_1 x_1\right] - f_3 u_3 - f_2 u_2 \tag{62}$$

subject to (61).

The optimal solution to this maximization induces $\bar{x} = x_2$. To see this, suppose $x_2 < \bar{x} = x_1$. Then the first order conditions require $W'(x_2) = c_2 < W'(x_1) = c_1 + \frac{f_3 + f_2}{f_1}(c_1 - c_2) + \frac{f_2}{f_1}(c_2 - c_3)$, which is a contradiction to $x_2 < x_1$. Note also that $x_3 > \bar{x}$, with a similar argument. Therefore, another way to write (62) is

$$\max_{\{x_n, u_n, r_n\}_{n=1,2,3}} \left\{ \begin{array}{c} f_3\left[W(x_3) - c_3 x_3\right] + f_2\left[W(x_2) - c_2 x_2\right] + f_1\left[W(x_1) - c_1 x_1\right] \\ -f_3\left[(c_1 - c_2) x_1 + (c_2 - c_3) x_2\right] + f_2 (c_1 - c_2) x_1 \\ +f_2 \min\left\{(c_2 - c_3)(x_3 - x_2), (c_1 - c_2) x_1\right\} \end{array} \right\} \tag{63}$$

subject to $x_3 \geq x_2 \geq x_1$.

Since the outcome defined in Proposition 6 is an optimal solution to problem (63) as well, it is an optimal implementable outcome.

9.6 Proof for Proposition 7

The participation constraints for A (7) and S (8) are satisfied. (Note that $f_3(u_3 - r_3) + f_2(u_2 - r_2) = (f_3 + f_2)(c_1 - c_2)x_1 \geq 0$.) Given $f_2(c_1 - c_2) < f_3(c_2 - c_3)$, delegation feasibility (6) is satisfied for $l = 1$, since there is no room for coalitional improvement for $l = 1$.

The remaining condition is (6) for $l = 2$. We start with observing **AIC** and $\mathbf{d-AIR}$ constraints are satisfied at $\{\hat{n}(n), \hat{r}(n)\}_{n=2,3} = \{n, r_n\}_{n=2,3}$. For (6) to fail, there must exist $\{\hat{n}(n), \hat{r}(n)\}_{n=2,3}$ which satisfies **AIC** and $\mathbf{d-AIR}$, and gives a higher value for the objective function. $\mathbf{AIC}(2|3)$, together with $x_2 = \min\{x_n\}_{n=1,2,3}$ imply $\hat{r}(3) \geq r_3$. For $\{\hat{n}(n), \hat{r}(n)\}_{n=2,3}$ to improve on $\{n, r_n\}_{n=2,3}$, there must exist n and $\hat{n}(n)$ such that $u_{\hat{n}(n)} + \left(c_{\hat{n}(n)} - c_n\right) x_{\hat{n}(n)} > u_n$.

The only possible coalitional improvement is type 2's imitating another type. The maximum value for $u_{\hat{n}(2)} + \left(c_{\hat{n}(2)} - c_3\right) x_{\hat{n}(2)} - u_2$ under deviation $\{\hat{n}(n), \hat{r}(n)\}_{n=2,3}$ is $\frac{f_3}{f_2}(c_2 - c_3)(x_1 - x_2)$.

Due to $\mathbf{AIC}(2|3)$ and $\mathbf{d-AIR}(2)$, with such a deviation, the rent for type 3 is $\hat{r}(3) \geq r_2 + (c_2 - c_3)x_{\hat{n}(2)} \geq (c_2 - c_3)x_1$. Compared with the original side contract, this implies a minimum loss by the amount $\hat{r}(3) - r_3 = (c_2 - c_3)(x_1 - x_2)$ for S, whenever A is type 3. When weighted with the probabilities, S's gain whenever A is type 2 is totally consumed by her loss whenever A is type 3. Therefore, there is no profitable $\{\hat{n}(n), \hat{r}(n)\}_{n=2,3}$ for S.

References

[1] Baliga, S. and T. Sjostrom, 1998. "Decentralization and collusion" Journal of Economic Theory, 83, 196-232.

[2] Baron, D. P. and D. Besanko, 1994. "Informational hierarchies, self-remedying hidden gaming, and organizational neutrality," Working Paper, Stanford University.

[3] Caillaud, B., B. Jullien and P. Picard, 1996a. "National versus European incentive policies: bargaining, information and coordination" European Economic Review, 40, 91-111.

[4] Caillaud, B., B. Jullien and P. Picard, 1996b. "Hierarchical organization and incentives" European Economic Review, 40, 687-695.

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

[6] Cramton, P. C. and T. R. Palfrey, 1995. "Ratifiable mechanisms: learning from disagreement" Games and Economic Behavior, 10, 255-283.

[7] Faure-Grimaud A., J. J. Laffont and D. Martimort, 2000. "A theory of supervision with endogenous transaction costs" Annals of Economics and Finance, 1, 231-263.

[8] Faure-Grimaud A., J. J. Laffont and D. Martimort, 2002. "Collusion, delegation and supervision with soft information" forthcoming in The Review of Economic Studies.

[9] Felli, L., 1996. "Preventing collusion through discretion" LSE working paper.

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

[11] Jullien, B., 2000. "Participation constraints in adverse selecton models" Journal of Economic Theory, 93, 1-47.

[12] Kofman, F. and J. Lawarree, 1993. "Collusion in hierarchical agency" Econometrica, 61, 629-656.

[13] Kofman, F. and J. Lawarree, 1996. "A prisoner's dilemma model of collusion deterrence" Journal of Public Economics, 59, 117-136.

[14] Laffont, J. J. and D. Martimort, 1997. "Collusion under asymmetric information" Econometrica 65, 875-911.

[15] Laffont, J. J. and D. Martimort, 1998. "Collusion and delegation" Rand Journal of Economics, 29, 280-305.

[16] Laffont, J. J. and D. Martimort, 1999. "Separation of regulators against collusive behavior" Rand Journal of Economics 30, 232-262.

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

[18] Laffont, J. J. and M. Meleu, 2001. "Separation of powers and development" Journal of Development Economics 64, 129-145.

[19] Laffont, J. J. and J. Tirole, 1991. "The politics of government decision making: a theory of regulatory capture" Quarterly Journal of Economics, 106, 1089-1127.

[20] Laffont, J. J. and J. Tirole, 1993. The theory of incentives in procurement and regulation, MIT Press.

[21] Lewis, T. R. and D. E. M. Sappington, 1989. "Countervailing incentives in agency problems" Journal of Economic Theory, 49, 294-313.

[22] McAfee, P. and J. McMillan, 1995. "Organizational Diseconomies of Scope" Journal of Economics and Management Strategy 4, 399-426.

[23] Maggi, G. and A. Rodriguez-Clare, 1995. "On Countervailing Incentives" Journal of Economic Theory 66, 238-263.

[24] Martimort, D., 1999. "The life cycle of regulatory agencies: Dynamic capture and transaction costs" The Review of Economic Studies, 66, 929-947.

[25] Maskin, E. and J. Tirole, 1990. "The principal-agent relationship with an informed principal: the case of private values" Econometrica, 58, 379-409.

[26] Maskin, E. and J. Tirole, 1992. "The principal-agent relationship with an informed principal II: common values" Econometrica, 60, 1-43.

[27] Melumad, N., D. Mookherjee, and S. Reichelstein, 1995. "Hierarchical decentralization of incentive contracts" Rand Journal of Economics, 26, 654-692.

[28] Mookherjee, D., 2003. "Delegation and contracting hierarchies: an overview" invited session, Econometric Society Meetings, Northwestern University.

[29] Mookherjee, D. and M. Tsumagari, 2002. "The organization of supplier networks: effects of delegation and intermediation" Boston University working paper.

[30] Stole L., 1990 "Mechanism design and common agency" mimeo, University of Chicago.

[31] Tirole, J., 1986."Hierarchies and breaucracies: on the role of collusion in organizations" Journal of Law, Economics and Organization 2, 181-214.

[32] Tirole, J., 1992. "Collusion and the theory of organizations," in J. J. Laffont, ed., Advances in Economic Theory: Sixth World Congress, Vol. II, Cambridge University Press, Cambridge.