MECHANISM DESIGN UNDER COLLUSION AND RISK AVERSION

by

Gorkem Celik
Department of Economics
University of British Columbia

December 2003

Discussion Paper No.: 03-14
Mechanism Design under Collusion and Risk Aversion*

Gorkem Celik†

December 22, 2003

Abstract

In this paper, I study a multi-player mechanism design problem under the assumption that the players are able to collude: The principal commits to making a transfer to the productive agent increasing in the output level. The principal also hires a supervisor, whose wage (potentially) depends on the output level as well. To insure himself against the uncertainty of the transfer, the principal wants the supervisor’s wage to be declining in the output level. Such an interdependent compensation structure brings in the question of collusion between the supervisor and the agent. I characterize the set of wage profiles that are consistent with the potential of collusion. I identify the optimal wage profile corresponding to the intended output profile. The optimal wage is decreasing in the output, so that supervision provides some insurance for the principal. However, the rate of change of the wage does not completely offset the rate of change of the transfer. Therefore full insurance is not attainable if collusion is possible.

Key Words: Collusion, mechanism design.

JEL Classification: D82, C72.

1 Introduction

In this paper, I study how the potential for collusion creates additional incentives for the players within a hierarchy and how the organization of the hierarchy should respond to these incentives. To address these questions I employ the standard adverse selection framework. The principal (P) is the residual claimant of a commodity produced by the agent (A). The unit production cost (the type of the agent) is known by A, but unknown to P. In order to encourage production, P commits to a grand contract. This grand contract consists of a transfer schedule that maps output levels to monetary transfers P will

---

*I thank Mike Peters, Okan Yilankaya, seminar participants at UBC and SFU.

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

1I will use masculine pronouns for the principal as well as the agent, and use feminine pronouns for the supervisor.
make to $A$. Through these transfers, $P$ not only compensates $A$ for the production cost but also leaves him an *information rent*. It follows from the earlier analyses\(^2\) of this adverse selection problem that, for $A$ not to overstate the production cost, his information rent should be decreasing in the realization of the production cost. As a result, the transfer level $P$ makes to $A$ and the output level he receives from him are both decreasing in $A$’s type.

In this bilateral interaction, one rationale for the involvement of an unproductive supervisor ($S$) arises from $P$’s risk aversion in the monetary transfer he makes: In the environment I study, $P$’s disutility from the transfer increases with an increasing rate. This specific form of risk aversion may result from the increasing marginal cost of raising funds. If $P$’s grand contract includes a *wage schedule* for $S$ that is contingent on the level of output, this may provide insurance for $P$ against the uncertain transfer level: When the realization of $A$’s type is high, and therefore $P$ is required to make lower transfers to $A$, $P$ is willing to make a positive wage payment to $S$. This is in return for a negative wage level for $S$ whenever the lower types are realized and $P$ is required to make a higher transfer to $A$. Through such a scheme, $P$ cannot reduce the expected value of the total payment he makes to the other players,\(^3\) but he can reduce the variance of it and therefore increase his expected payoff.

Reducing the variance of the total payment requires $S$’s wage to be responsive to $A$’s output choice. This interdependent payoff structure introduces the potential for collusion between $S$ and $A$. If $S$’s wage is decreasing in the output level, she will have an incentive to bribe $A$ in order to encourage him to reduce the output. Therefore, $P$ should account for the possibility of such a collusion when designing his grand contract.

If collusion between $S$ and $A$ is perfectly efficient, in the sense that it sustains an outcome that is Pareto efficient for the $S$ - $A$ coalition, then these players behave as if they were a single entity. In this case, $S$’s existence will be totally irrelevant from $P$’s perspective: Any expected payoff that is available for $P$ under supervision would still be available in the absence of supervision. However, since $A$’s type is unknown to $S$, collusion between these players takes the form of a side contract between asymmetrically informed parties. Thus, generally, collusion falls short of attaining full efficiency. This presents the prospect of beneficial supervision despite the threat of collusion.\(^4\)

In recent years collusion between asymmetrically informed parties has been explored by many researchers. Laffont & Martimort (1997, 1998) study collusion between two agents producing complementary products. Faure-Grimaud, Laffont & Martimort (2003) extend this analysis to address collusion between a productive agent and a supervisor who is imperfectly informed on the agent’s type. In these

\(^2\)See Baron & Myerson (1982) among others.

\(^3\)Unless the supervisor is *risk lover* and therefore willing to pay a premium for the risk she assumes.

\(^4\)Inefficiency of collusion can also be sustained by assuming exogenous transaction costs for colluding parties. See Tirole (1992) for an extensive survey of the literature following this alternative approach.
papers the productive agents’ marginal cost levels are assumed to take two possible values. Under this assumption, the principal’s optimal grand contract takes the form of contracting with only one of the players and delegating to her the authority to contract with the other player.

When the marginal cost can take more than two possible values, this delegation scheme is prone to the problem of double marginalization of information rents. In the absence of a direct interaction between the principal and the productive agent, the principal himself cannot provide the agent with the information rent that is decreasing in his type. Therefore there is no direct incentive for the agent not to overstate the production cost. Instead, the principal has to motivate some other player to provide this incentive for the agent. For this other player to be willing to leave an information rent to lower types of the agent, her payoff should also be decreasing in the type of the agent. This requires the total information rent (information rent of the agent plus the payoff of the third player) to be more responsive to the type of the agent than it was in the presence of a direct interaction between the principal and the agent. Accordingly, when the marginal cost can take more than two possible values, Mookherjee & Tsumagari (2003) and Celik (2003) construct grand contracts that outperform delegation, respectively in the contexts of collusion between two agents and collusion between the partly informed supervisor and the agent.

The current paper differs from this earlier literature since it studies a hierarchy, where the third player neither takes part in production nor possesses any relevant information regarding the parameters of production. Nevertheless this third party is potentially useful in alleviation of the uncertainty in the monetary transfer that $P$ makes. In this environment, I characterize the set of outcomes, which are implementable through some grand contract. Since delegation constitutes a special class of grand contracts, the outcomes that are implementable under delegation are a subset of this implementable set. These delegation implementable outcomes are undesirable from $P$’s perspective, due to the double marginalization phenomenon outlined by Mookherjee & Tsumagari (2003) and Celik (2003). Moreover,

5 This player can be either one of the agents when collusion is between two productive agents. When collusion is between the partly informed supervisor and the productive agent, the principal should contract with the supervisor.

6 The term double marginalization characterizes the externality between upstream and downstream monopolies. (See Tirole (1998) page 174.) To my knowledge Melumad, Mookherjee & Reichelstein (1995) are the first to use the term in a delegation setup.

7 Mookherjee & Tsumagari (2003) employ a general production technology, where the products of the agents need not be perfect complements. In this environment, they also study the effect of delegation to a completely informed but unproductive “middleman.” They show that double marginalization is not relevant if the delegate has full information on the delegated party.

8 In Celik (2003), the type space of the agent is an arbitrary finite set. But the supervisor’s information structure is restricted to be a connected partition of this type space. Therefore the model there is not a direct extension of the earlier models on informed supervisor. In Celik (2003), delegation is outperformed by the absence of supervision, and the absence of supervision is outperformed by a more general grand contract.
unlike in these earlier papers, the characterization result here paves the way for the identification of the optimal implementable outcome rather than the construction of some implementable outcome that would dominate the delegation implementable ones.  

Under the optimal grand contract, $A$ has the option of refusing to collude with $S$ and contracting directly with $P$. This outside option will provide a reservation utility for $A$ that is decreasing in his type. In this setting, as shown by Lewis & Sappington (1989), Maggi & Rodriguez-Clare (1995), and Jullien (2000), the dominant incentive for $A$ would be understating the production cost in order to increase the compensation he would receive for the sacrificed outside option. As a consequence of this incentive reversal for $A$, the main concern for $S$ at the collusion stage becomes the deterrence of the understatement of the production cost rather than its overstatement. This is consistent with leaving $S$ a payoff that is increasing in $A$’s type. This phenomenon can be named as counter marginalization of information rents as opposed to the double marginalization inherent in delegation.

By employing a grand contract that would result in counter marginalization, $P$ can induce a total payment that is less responsive to $A$’s type than it was in the absence of supervision. Under the optimal grand contract, for lower realizations of $A$’s type, the total payment does not vary with the type at all. However, for higher realizations of the type, the total payment will be decreasing (but with a slower rate than it would have been in the absence of supervision). Therefore the optimal grand contract improves over the absence of supervision but fails to sustain full insurance for $P$.

The current paper is also related to a paper by Baliga & Sjostrom (1998), where they consider collusion between two productive agents, who sequentially decide on the intensity of the effort they will exert in the production process. Unlike in the current paper, the underlying informational problem in their model is moral hazard. Even though collusion takes place under symmetric information, the efficiency of collusion is hindered by the limited liability of the colluding parties. The negative results on delegation derived for the adverse selection framework does not carry on to this moral hazard setup. For a wide range of parameters, Baliga & Sjostrom (1998) show the optimality of delegation to one of the agents.

In the current paper, the productive agent receives two contracts from two different parties regarding the production level. First $P$ offers a grand contract and then, after observing this grand contract, $S$ offers a side contract. This structure resembles the structure of the sequential common agency game studied by Calzolari & Pavan (2001). However, unlike in their common agency framework, here $P$ can

---

9In Celik (2003), I provide a characterization of implementable outcomes when the supervisor is informed. However the characterization result there does not reveal a general formula to identify the optimal implementable outcome. Therefore, in that paper, I could derive the optimal outcome only for a specific example.

10The identification of the agent to delegate to (whether to delegate to the upstream agent or the downstream one) depends on the exact parameters of the model.
directly affect S’s preferences over A’s production level. This is due to the fact that S is also a party for the grand contract offered by P.

Organization of the rest of the paper is as follows: In section 2, I present the model. Even though the focus of this paper is collusive supervision, I start with stating well known results from the two player adverse selection setups. This will establish the basis for the analysis of the three player hierarchy and also serve as the no-supervision benchmark. Potential benefits of supervision and its vulnerability to collusion are introduced in this section as well. In section 3, I outline the outcomes that are implementable under delegation and discuss why they are dominated by no-supervision outcomes. In section 4, I outline the outcomes that are implementable through the general class of grand contracts. In section 5, I identify the optimal implementable transfer - wage profile and discuss its properties. Until section 6, I model collusion as a side contract that is offered by S to A. In section 6, I suggest two alternative formulations, where the collusive side contract is offered (i) by a fourth party, and (ii) by A. I show that in either of these alternative formulations, the set of implementable outcomes are not restricted much further. I conclude with section 7. Section 8 is the appendix, which contains the omitted proofs in the text.

2 The Model

2.1 No Supervision

The principal (P) is the residual claimant of a commodity produced by the agent (A). The constant unit cost of production (θ) is observed by A, but unknown by P. I will also refer to the variable θ as the type of A. θ is continuously distributed on the support [θ, θ], where 0 < θ < θ. This distribution is governed by the cumulative distribution function F(·), with the probability density function f(·). Since θ is continuously distributed, the latter function is strictly positive over [θ, θ]. I will impose a generalized monotone hazard rate condition for the distribution, which is standard in the type dependent participation constraints literature: $\frac{d}{d\theta} \left( \frac{F(\theta)}{f(\theta)} \right)$ and $\frac{d}{d\theta} \left( \frac{F(\theta)-1}{f(\theta)} \right)$ are well defined and nonnegative for all θ.

To induce A’s production, P commits to a transfer schedule $T(\cdot)$ that maps output levels to monetary transfers P will make. A’s utility is quasilinear in this monetary transfer, i.e., it can be written as $T(x) - \theta x$, where x is the output level and θ is A’s type. Given $T(\cdot)$, A chooses the output level that would maximize his utility conditional on his type. This decision induces an output profile $x(\cdot)$ and a transfer profile $t(\cdot)$, both as functions of the type of A.

**Definition 1** The output - transfer profile $\{x(\cdot), t(\cdot)\}$ is incentive compatible (through $T(\cdot)$) if there
exists a transfer schedule $T(\cdot)$ such that

$$x(\theta) \in \arg \max_x \{T(x) - \theta x\},$$

$$t(\theta) = T(x(\theta))$$

for all $\theta$.

A’s selection of the optimal output level corresponding to his type can also be considered as his making a statement of his realized type. It follows from the revelation principle that an output - transfer profile is incentive compatible if and only if there exists no type that strictly prefers the output - transfer pair of another type. This reveals the following “more familiar” representation for incentive compatibility:

$$IC(\hat{\theta}|\theta) : t(\hat{\theta}) - \theta x(\hat{\theta}) \geq t(\theta) - \theta x(\theta)$$

for all $\hat{\theta}, \theta$. Continuity of $t(\theta) - \theta x(\theta)$, together with the following first order and monotonicity conditions are necessary and sufficient for this:

$$FO : \frac{d[t(\theta) - \theta x(\theta)]}{d\theta} = -x(\theta) \text{ a.e.}$$

$$Mon : x(\theta) \text{ is non-increasing}$$

Condition $FO$ reveals that the rate of change in the incentive compatible transfer profile $t(\theta)$ is different than the rate of change in the total production cost $\theta x(\theta)$. This difference between the two rates is attributed to the information rent of $A$. In the absence of an incentive compatible scheme, $A$ would have an incentive to overstate his type (the unit cost of production) in order to increase his compensation from $P$ for the production cost he incurs. To preclude such a misrepresentation of A’s type, the incentive compatible contract should make imitation of higher types less desirable. This requires leaving an information rent to $A$ which is decreasing in his type.

It follows from the representation above that any non-increasing output profile is incentive compatible together with a transfer profile. This transfer profile is pinned down by $FO$ up to a constant. Accordingly, $t(\cdot)$ is constant whenever $x(\cdot)$ is constant, strictly decreasing whenever $x(\cdot)$ is strictly decreasing, discontinuous whenever $x(\cdot)$ is discontinuous. If $x(\cdot)$ and $t(\cdot)$ are both differentiable at $\theta$, the derivative of $t(\cdot)$ at $\theta$ is $\theta x'(\theta)$.

Suppose $x(\cdot)$ is a non-increasing output profile. Define function $x^{-1}(\cdot)$ as

$$x^{-1}(x) = \begin{cases} 
\sup \{\theta : x(\theta) \geq x\} & \text{if } x \leq x(\theta) \\
0 & \text{otherwise}
\end{cases}. \quad (1)$$

Note that if $x(\cdot)$ has an inverse, its inverse will be equal to $x^{-1}(\cdot)$ on the relevant support. Provided that \{x(\cdot), t(\cdot)\} is incentive compatible, by employing the function $x^{-1}(\cdot)$, we can construct a continuous
transfer schedule $T(\cdot)$ such that

$$T(x(\theta)) = t(\theta) \quad \text{and} \quad T'(x) = x^{-1}(x) \quad \text{for all} \quad x. \tag{2}$$

Since $x^{-1}(\cdot)$ is a non-increasing function of $x$, the constructed transfer schedule is concave, and $x(\theta)$ maximizes the utility of agent type $\theta$ under this transfer schedule. It follows from this argument that $\{x(\cdot), t(\cdot)\}$ is incentive compatible through the continuous transfer schedule constructed above. Since $\{x(\cdot), t(\cdot)\}$ is an arbitrary incentive compatible output - transfer profile, it is without loss of generality to allow for only the continuous functions as transfer schedules.

Incentive compatibility outlines the set of available output - transfer profiles for $P$, conditional on $A$’s consent to participate in production under the proposed transfer schedule. Securing $A$’s participation requires leaving him a non-negative utility regardless of his type. This is stipulated by the following individual rationality constraint:

$$IR(\theta) : t(\theta) - \theta x(\theta) \geq 0$$

**Definition 2** The output - transfer profile $\{x(\cdot), t(\cdot)\}$ is **no-supervision implementable** if

i) $\{x(\cdot), t(\cdot)\}$ is incentive compatible.

ii) $\{x(\cdot), t(\cdot)\}$ satisfies $IR(\theta)$ for all $\theta$.

The set of no-supervision implementable output - transfer profiles can be considered as the budget set for $P$. His design problem reduces to choosing the optimal profile within this set. $P$’s utility consists of two separable parts: He receives a direct utility from production, as well as an indirect disutility due to the transfer he makes to induce production. Conditional on the implemented output - transfer profile, his expected utility can be written as

$$\int_{\theta} [U(x(\theta)) - V(t(\theta))] f(\theta) d\theta. \tag{3}$$

$U$ is continuous, increasing and strictly concave. $V$ is continuous, increasing and convex. The convexity of $V$ reflects risk aversion of $P$ in money. Note that quasilinear utility, which is a rather standard specification in the adverse selection literature, is a special case of this representation, where $V$ is linear, and $V'$ is normalized at 1.

$P$ chooses $\{x(\cdot), t(\cdot)\}$ to maximize his expected utility subject to constraints $FO$, $Mon$, and $IR(\theta)$ for all $\theta$.\footnote{This representation of the mechanism design problem does not allow $P$ to offer a stochastic mechanism that maps $A$’s type to probability distributions over output and transfer pairs. Since $P$’s payoff is concave in output and transfer pairs and $A$’s payoff is linear in these variables, limiting attention to deterministic mechanisms is without loss of generality.} This problem can be separated into two parts: First, $P$ treats $x(\cdot)$ as exogenous and he chooses $t(\cdot)$ among the transfer profiles that are implementable with $x(\cdot)$ to minimize $\int_{\theta} V(t(\theta)) f(\theta) d\theta$. I will
refer to the solution of this minimization problem as the **optimal no-supervision transfer profile** that induces $x(\cdot)$. Then he chooses a non-increasing output profile $x(\cdot)$ that would maximize his expected payoff together with the optimal no-supervision transfer profile that induces $x(\cdot)$. The results of this paper will follow from the analysis of the first part of this optimization problem.

We have already seen that selection of $x(\cdot)$ determines the incentive compatible $t(\cdot)$ up to a constant. Therefore, in this no-supervision setup, the first part of $P$’s optimization problem reduces to choosing this constant without violating the individual rationality constraints. Since $t(\theta) - \theta x(\theta)$ is declining in $\theta$ for an incentive compatible $\{x(\cdot), t(\cdot)\}$, the individual rationality constraint of the highest type ($IR(\bar{\theta})$) is sufficient for all the other $IR$ constraints. And since $P$’s utility is declining in the monetary transfer, $IR(\bar{\theta})$ is binding, i.e., $t(\bar{\theta}) = \bar{\theta} x(\bar{\theta})$. This reveals the optimal no-supervision transfer profile that induces $x(\cdot)$. (See Figure 1.)

Any incentive compatible transfer profile, and therefore the optimal one, is decreasing in $\theta$. Since $V$ is increasing and convex, $P$’s marginal disutility of transfer is declining in $\theta$ as well. This indicates a potential gains from trade if $P$ has access to a third party that he can write a contract contingent on the realized output level. Particularly, $P$ can promise a positive monetary transfer to the third party whenever $A$ reveals a high $\theta$, in return for a positive transfer from the third party whenever $A$ reveals a low $\theta$. Such a contract would increase the expected utility of $P$.

Before formalizing the idea of introducing this third party, let me briefly comment on the characterization of the optimal output profile. For the time being, I will ignore the monotonicity constraint. By treating $x(\theta)$ as the control variable and $t(\theta) - \theta x(\theta)$ as the state variable, $P$’s output profile selection problem can be transformed into an optimal control program. Accordingly, $x(\cdot)$ is optimal if

$$x(\theta) \in \arg \max_x \left\{ U(x) - V(\theta x + r(\theta)) - \int_\theta^\theta V'(s x(s) + r(s)) f(s) ds \right\}$$

(4)

for all$^{12}$ $\theta$, where $r(\theta) = \int_\theta^\theta x(\theta) d\theta$. Under the special case of quasilinear utility for $P$, i.e. $V$ is linear with $V' = 1$, this last condition boils down to the following familiar optimality condition:

$$x(\theta) \in \arg \max_x \left\{ U(x) - \theta x - \frac{F(\theta)}{f(\theta)} x \right\}$$

(5)

for all $\theta$. The output profile that is derived from this last condition is decreasing since $U(\cdot)$ is strictly concave and $\frac{F(\theta)}{f(\theta)}$ is increasing. Therefore the ignored monotonicity constraint is slack under the quasi-

---

$^{12}$Technically, an output profile is optimal if and only if it satisfies this condition almost everywhere (for all $\theta$ except for a subset with probability zero). Any such output profile is essentially equivalent (identical for expected payoff relevant purposes) to $x(\cdot)$ that satisfies the condition for all $\theta$. Therefore it is without loss of generality to suppress the term almost everywhere and to consider the output profiles that satisfy this condition for all $\theta$. I will follow the same approach in the statement of Condition (14).
linearity assumption. The optimal value for $x(\theta)$ is distorted downwards from its Pareto efficient value that maximizes $U(x) - \theta x$ (for $\theta > 2$). Note that the magnitude of this distortion is increasing in $\theta$.

2.2 Supervision under Collusion

In this subsection, I introduce the supervisor ($S$) as a third party to the principal-agent interaction. $S$ does not incur any production cost or enjoy any direct benefit from production. Her ex-post utility is equal to the monetary payment she receives.\(^\text{13}\) $P$ is still considered as the mechanism designer with commitment power. A grand contract for $P$ has two components: As before, $T(\cdot)$ is the transfer schedule for $A$. The new component, $W(\cdot)$, is the wage schedule for $S$. Both schedules are continuous functions of the level of output produced by $A$.

There is no change in the utility function for $A$. He chooses an output level that maximizes his type dependent utility. As a result of $A$’s optimization, the grand contract $\{T(\cdot), W(\cdot)\}$ yields an output profile $x(\cdot)$, a transfer profile $t(\cdot)$, and a wage profile $w(\cdot)$, all as functions of $\theta$. I will refer to $\{x(\cdot), t(\cdot), w(\cdot)\}$ as an outcome. The total transfer $P$ makes to the players is $t(\theta) + w(\theta)$ whenever $A$’s type is $\theta$. Therefore, we can write down the objective function of $P$ as

$$
\int_{\theta} U(x(\theta)) - V(t(\theta) + w(\theta)) f(\theta) d\theta.
$$

As before, $A$’s participation in the grand contract is guaranteed by the $IR(\theta)$ constraints. $S$ does not know the type of $A$, but is informed on the type distribution. Therefore, $S$’s participation is assured by the following ex-ante individual rationality constraint:

$$
IR - S : \int_{\theta} w(\theta) f(\theta) d\theta \geq 0.
$$

$S$’s individual rationality constraint reveals that the expected value of $w(\theta)$ is at least 0. Therefore, supervision does not help in reducing the expected value of the total payment $P$ has to make to the other players $(t(\theta) + w(\theta))$. However, by using $S$ as an insurer, the risk averse principal can reduce the variance of the total payment and increase his expected payoff without reducing the expected total payment. If the supervisor-agent collusion were not a concern, the optimal grand contract would induce a flat total transfer from $P$, regardless of the output level.\(^\text{14}\)

Provision of insurance through $S$ requires her ex-post payoff to depend on the production level of $A$. The interdependence of this payoff structure brings in the question of collusion. After $P$’s announcement

\(^{13}\)This indicates that, unlike the principal, the supervisor is risk neutral. This assumption is to keep the exposition simple. The results to follow are valid as long as $S$ is not infinitely risk averse.

\(^{14}\)Note that insuring the principal requires smoothing the total payment he makes, not his ex-post utility. Therefore making the risk neutral supervisor residual claimant of production is not an optimal mechanism in this environment.
of the grand contract \( \{T(\cdot), W(\cdot)\} \), \( S \) approaches to \( A \) and commits to a side contract, \( B(\cdot) \), which specifies the bribe \( S \) will pay to \( A \) as a function of the output level. The timing for the resulting game is as follows:

T0: \( \theta \) is selected by nature and observed by \( A \).

T1: \( P \) announces a grand contract \( \{T(\cdot), W(\cdot)\} \) to \( S \) and \( A \). Each of them accepts or rejects the grand contract. If both accept, the game proceeds to the next stage. Otherwise, the game ends without any production and any monetary payment.

T2: \( S \) announces a side contract \( B(\cdot) \) to \( A \). \( A \) accepts or rejects the side contract.\(^{15} \)

T3: \( A \) decides on \( x \), the level of production. \( P \) pays \( T(x) \) to \( A \), and \( W(x) \) to \( S \). If \( A \) accepted the side contract, \( S \) pays \( B(x) \) to \( A \). If \( A \) rejected the side contract, \( S \) does not make any payment to him.

If \( A \) accepts \( S \)'s side contract, his output choice is affected by both the direct transfer he receives from \( P \), and the bribe he receives from \( S \). Accordingly, the output - transfer profile will be determined as

\[
\begin{align*}
x(\theta) & \in \arg \max_x \{ T(x) + B(x) - \theta x \}, \\
t(\theta) & = T(x(\theta)) + B(x(\theta)).
\end{align*}
\]

Note that in this setup \( t(\theta) \) is defined as the net transfer for type \( \theta \) agent including the bribe he receives from \( S \). The profile \( \{x(\cdot), t(\cdot)\} \) would satisfy the above conditions only if it is incentive compatible through \( T(\cdot) + B(\cdot) \). Conversely, if \( \{x(\cdot), t(\cdot)\} \) is incentive compatible, for any transfer schedule \( T(\cdot) \), there exists some bribe schedule \( B(\cdot) \) such that the above conditions are satisfied. Therefore the above conditions reduce to \( \{x(\cdot), t(\cdot)\} \) being incentive compatible.

The side contract \( S \) offers should also provide the incentive for \( A \) to collude with \( S \) rather than rejecting \( S \)'s offer.\(^{16} \) In case of such a rejection, \( A \) responds directly to the grand contract and receives a type dependent utility of \( \max_x \{ T(x) - \theta x \} \). To guarantee \( A \)'s participation, his ex-post utility from colluding with \( S \) should be greater than this reservation utility for each \( \theta \).

These incentive compatibility and participation constraints outline the available output - transfer profiles for \( S \) at the collusion stage. By choosing a side contract, she picks one of those available profiles to maximize her expected surplus. Provided that \( \{T(\cdot), W(\cdot)\} \) is the grand contract, and \( S \) induces \( \{x(\cdot), t(\cdot)\} \) through her side contract, her ex-post surplus is \( T(x(\theta)) + W(x(\theta)) - t(\theta) \) as a function

\(^{15}\)Equivalently, \( S \) can be restricted to make only non-negative bribe commitments to \( A \). \( A \) would not have an incentive to reject any such side contract.

\(^{16}\)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 \( A \). Therefore there is no loss of generality in considering only the outcomes that result from \( A \)'s acceptance of the side contract.
of \( \theta \). Accordingly, \( S \)'s side contract selection problem is identified as follows:

\[
\max_{\ddot{x}(\cdot), \ddot{t}(\cdot)} \int_{\theta} \left[ T(\dot{x}(\theta)) + W(\dot{x}(\theta)) - \dot{t}(\theta) \right] f(\theta) d\theta \quad \text{s.t.} \quad \text{IC} \quad \left\{ \dot{x}(\theta), \dot{t}(\theta) \right\} \quad \text{is incentive compatible} \quad \text{Part}(\theta) \quad \dot{t}(\theta) - \dot{\theta} \cdot \dot{\theta} \geq \max_{x} \left\{ T(x) - \theta x \right\} \quad \text{for all} \quad \theta
\]

For \( \{x(\cdot), t(\cdot), w(\cdot)\} \) to be a feasible outcome under the threat of supervisor-agent collusion, \( x(\cdot) \) and \( t(\cdot) \) must constitute a solution to the above problem and \( w(\cdot) \) must identify the ex-post utility of \( S \) net of the bribe she pays.

**Definition 3** The outcome \( \{x(\cdot), t(\cdot), w(\cdot)\} \) is **collusion feasible** (through \( \{T(\cdot), W(\cdot)\} \)) if there exist a transfer schedule \( T(\cdot) \) and a wage schedule \( W(\cdot) \) such that

\[
i) \left\{ x(\cdot), t(\cdot) \right\} \quad \text{is a solution to (7)}, \quad \text{(8)}
\]

\[
ii) w(\theta) + t(\theta) = W(x(\theta)) + T(x(\theta)) \quad \text{for all} \quad \theta. \quad \text{(9)}
\]

Implementability is defined by combining collusion feasibility with the individual rationality constraints.

**Definition 4** The outcome \( \{x(\cdot), t(\cdot), w(\cdot)\} \) is **implementable** if

\[
i) \left\{ x(\cdot), t(\cdot), w(\cdot) \right\} \quad \text{is collusion feasible.}
\]

\[
ii) \left\{ x(\cdot), t(\cdot) \right\} \quad \text{satisfies} \quad \text{IR}(\theta) \quad \text{for all} \quad \theta.
\]

\[
iii) w(\cdot) \quad \text{satisfies} \quad \text{IR} - S.
\]

Once again, \( P \)'s mechanism design problem reduces to choosing the implementable outcome that maximizes his expected payoff (6). This problem can also be dismantled into two parts: The first part is \( P \)'s selection of the transfer-wage profile \( \{t(\cdot), w(\cdot)\} \) that would be implementable together with the exogenous \( x(\cdot) \). As before, this part requires minimizing the expected disutility of payments by \( P \), which is represented by the function \( V(\cdot) \). I will refer to the solution to the first part of \( P \)'s optimization as the **optimal implementable transfer-wage profile that induces** \( x(\cdot) \). The main concern of this paper is identifying this optimal profile. The second part of \( P \)'s problem, which will be suppressed most of the time, is his selection of the implementable output profile \( x(\cdot) \).

If the output-transfer profile \( \{x(\cdot), t(\cdot)\} \) is induced by an implementable outcome, it follows from the definition of this concept that \( \{x(\cdot), t(\cdot)\} \) is incentive compatible and satisfies the \( \text{IR}(\theta) \) constraints. In other words, any implementable outcome induces a no-supervision implementable output-transfer profile. Moreover, \( \text{IR} - S \) constraint imposes a lower bound for the expected wage profile. As a result, any implementable transfer-wage profile that induces \( x(\cdot) \) yields an expected total payment at least as
high as the expected value of the optimal no-supervision transfer profile that induces the same output profile. Nevertheless, the risk averse principal may still benefit from the existence of the supervisor through the implementation of a smoother total payment \((t(\cdot) + w(\cdot))\) profile than the no-supervision transfer profile.

We have already seen that incentive compatibility requires the transfer profile \(t(\cdot)\) to be a non-increasing (decreasing if \(x(\cdot)\) is not constant) function. Therefore a wage profile \(w(\cdot)\) is beneficial for \(P\) only if it is increasing. The optimal implementable wage profile that will be identified in section 5 will bear this property. The identification of this optimal wage profile requires characterization of the set of all implementable outcomes. But before moving to the characterization result, I will analyze a certain subset of the implementable outcomes, which consists of outcomes that are attainable under delegation.

### 3 Delegation

In this section, I will identify the outcomes that are collusion feasible through a special class of grand contracts, where \(T(\cdot) = 0\). Under such a grand contract, there is no direct incentive for \(A\) to produce the commodity. However, through the appropriate selection of the wage schedule, \(P\) can induce \(S\) to provide an indirect incentive for \(A\) at the side contracting stage. This special class of grand contracts can be regarded as \(P\)’s delegating to \(S\) the authority to contract with \(A\). Since delegation imposes a restriction on the grand contract, it is clear that it represents a loss of control for \(P\) relative to the general case. Nevertheless, delegation provides a useful benchmark for the analysis of mechanism design under collusion.

**Definition 5**

i) The outcome \(\{x(\cdot), t(\cdot), w(\cdot)\}\) is **delegation feasible** (through \(W(\cdot)\)) if there exists a wage schedule \(W(\cdot)\) such that \(\{x(\cdot), t(\cdot), w(\cdot)\}\) is collusion feasible through \(\{T_0(\cdot), W(\cdot)\}\), where \(T_0(x) = 0\) for all \(x\).

ii) The outcome \(\{x(\cdot), t(\cdot), w(\cdot)\}\) is **delegation implementable** if it is delegation feasible, it satisfies \(IR(\theta)\) for all \(\theta\), and it satisfies \(IR - S\).

Similarly, the **optimal delegation implementable transfer - wage profile that induces** \(x(\cdot)\) is defined as the transfer - wage profile which yields the highest expected payoff for \(P\) within the class of delegation implementable \(\{t(\cdot), w(\cdot)\}\) together with \(x(\cdot)\).

Under delegation, zero is the only output level that maximizes the direct utility \(A\) can acquire by rejecting the side contract. Therefore the outside option at the side contracting stage is shut down of production, regardless of the type of the agent. This implies that \(\max_x \{T_0(x) - \theta x\} = 0\) for all \(\theta\). When this is substituted in constraint \(Part(\theta)\), \(S\)’s side contract selection problem (7) turns out to be identical to the optimization problem of a risk neutral principal (principal with the quasilinear utility).
in the no-supervision game, who has \( W(\cdot) \) as his direct utility of output. Accordingly, the solution to problem (7) inherits the properties of the optimal output - transfer profile that solves the risk neutral principal’s problem: For an incentive compatible \( \{ x(\cdot), t(\cdot) \} \) to be a solution to problem (7), it must be that

\[
\begin{align*}
t(\overline{\theta}) &= \overline{\theta} x(\overline{\theta}), \quad (10) \\
x(\theta) &\in \arg \max_x \left\{ W(x) - \theta x - \frac{F(\theta)}{f(\theta)} x \right\} \quad (11)
\end{align*}
\]

for all \( \theta \).

Condition (11) indicates that \( S \)'s output selection problem is the same as the output selection problem of a productive agent, who has the unit production cost \( \theta + \frac{F(\theta)}{f(\theta)} \). The term \( \frac{F(\theta)}{f(\theta)} \) measures the extent of the double marginalization of information rents resulting from delegating to \( S \). Condition (11) can be combined with equation (9) in the definition of collusion feasibility to yield a first order necessary condition for \( w(\cdot) \) to be delegation feasible together with \( \{ x(\cdot), t(\cdot) \} \).

**Lemma 1** If \( \{ x(\cdot), t(\cdot), w(\cdot) \} \) is delegation feasible, then \( w(\theta) - \frac{F(\theta)}{f(\theta)} x(\theta) \) is a continuous function of \( \theta \) and

\[
\frac{d}{d\theta} \left[ w(\theta) - \frac{F(\theta)}{f(\theta)} x(\theta) \right] = - \left[ \frac{d}{d\theta} \frac{F(\theta)}{f(\theta)} \right] x(\theta) \quad \text{a.e.} \quad (12)
\]

**Proof.** See the appendix. ■

The first order condition (12) identifies the rate of change in the delegation feasible \( w(\cdot) \) as a function of \( x(\cdot) \), in the same manner that condition \( FO \) identifies the rate of change in the incentive compatible \( t(\cdot) \). In other words, the output profile reveals the delegation feasible wage profile up to a constant.

The discussion above exposes incentive compatibility of \( \{ x(\cdot), t(\cdot) \} \), (10), and (12) as necessary conditions for an outcome to be delegation feasible. With the following proposition, I show that these conditions are also sufficient for delegation feasibility and identify the optimal delegation implementable transfer - wage profile that induces a non-increasing output profile.

**Proposition 1** i) \( \{ x(\cdot), t(\cdot), w(\cdot) \} \) is delegation feasible if and only if \( \{ x(\cdot), t(\cdot) \} \) is incentive compatible, \( t(\overline{\theta}) \) is identified by (10), and \( w(\cdot) \) satisfies (12).

ii) The optimal delegation implementable transfer - wage profile that induces the non-increasing \( x(\cdot) \) is fully identified by delegation feasibility and the binding \( IR - S \) constraint.

**Proof.** i) Necessity follows from Lemma 1 and the preceding discussion. Proving sufficiency requires construction of a continuous wage schedule \( W(\cdot) \) such that \( x(\theta) \) satisfies (11) and \( W(x(\theta)) = w(\theta) + t(\theta) \) for all \( \theta \). Recall that \( x^{-1}(\cdot) \) is defined by (1). Consider the following \( W(\cdot) \):

\[
W(x(\overline{\theta})) = w(\overline{\theta}) + \overline{\theta} x(\overline{\theta}) \quad \text{and} \quad W'(x) = x^{-1}(x) + \frac{F(x^{-1}(x))}{f(x^{-1}(x))} \quad \text{for all } x.
\]

\[
W (x (\overline{\theta})) = w (\overline{\theta}) + \overline{\theta} x (\overline{\theta}) \quad \text{and} \quad W' (x) = x^{-1} (x) + \frac{F (x^{-1} (x))}{f (x^{-1} (x))} \quad \text{for all } x.
\]

\[
13
\]
Since \( x^{-1}(\cdot) \) is non-increasing and the monotone hazard rate condition is satisfied, \( W(\cdot) \) is concave. From the first order conditions, \( W(x) − \theta x − \frac{F(\theta)}{f(\theta)} \) is maximized by \( x(\theta) \). Condition \( W(x(\theta)) = w(\theta) + t(\theta) \) follows from \( FO \) and (12).

ii) Given \( x(\cdot) \), conditions (10) and \( FO \) fully identify the delegation feasible transfer profile. Condition (12) identifies the rate of change in the delegation feasible wage profile. Since \( P \)'s payoff is declining in \( w(\theta) \), the \( IR − S \) constraint is binding for his optimization problem. ■

As mentioned above, condition (12) reveals the rate of change in the delegation feasible \( w(\cdot) \) as a function of \( x(\cdot) \). (See Figure 2 for the optimal delegation implementable transfer - wage profile.) For instance, if \( x(\cdot) \) and \( w(\cdot) \) are differentiable at \( \theta \), the latter function has the derivative \( \frac{F(\theta)}{f(\theta)}x'(\theta) \) at \( \theta \). Since \( x(\cdot) \) is a non-increasing function, so is \( w(\cdot) \). Moreover \( w(\cdot) \) is strictly decreasing at \( \theta \) if \( x(\cdot) \) is strictly decreasing. This implies the failure of delegation to improve the risk averse principal’s expected payoff with respect to no-supervision implementation.

**Corollary 1** Suppose \( x(\cdot) \) is a non-increasing output profile. The optimal delegation implementable transfer - wage profile that induces \( x(\cdot) \) yields a weakly lower expected payoff for \( P \) than does the optimal no-supervision transfer profile. The ordering is strict if \( V(\cdot) \) is strictly convex and \( x(\cdot) \) is not constant.

The discussion above and the subsequent delegation failure result relate to the first part of \( P \)'s maximization problem. That is, they do not require the output profile to be chosen optimally. Fix a non-increasing output profile \( x(\cdot) \). Consider \( P \)'s expected payoff under the optimal delegation implementable transfer - wage profile that would induce \( x(\cdot) \). There exists a transfer profile that is no-supervision implementable together with \( x(\cdot) \) and that yields a higher expected payoff for \( P \).\(^{17}\)

To understand why delegation fails to provide insurance to \( P \), recall that \( S \)'s side contract selection problem is identical to the optimization problem of a risk neutral principal in the absence of supervision. To acquire \( A \)'s private information, \( S \) has to leave him an information rent decreasing in his type. On the other hand, for \( S \) to reveal this acquired information to \( P \), she should be left with an additional information rent also decreasing in \( A \)'s type. This double marginalization of information rents leads to a steeper total payment (as a function of \( A \)'s type) for \( P \) and consequently a lower expected payoff under delegation.

This decline in the principal’s payoff due to the cascading information rents is also in line with an earlier result by McAfee and McMillan (1995), who study the effects of delegating to a middle principal.

\(^{17}\)The optimal delegation implementable output profile can be unraveled by an optimal control program, where (6) is the objective function, \( x(\theta) \) is the control variable, \( t(\theta) − \theta x(\theta) \) and \( w(\theta) − \frac{F(\theta)}{f(\theta)} x(\theta) \) are the state variables. The dynamics of the system are given by conditions \( FO \) and (12). \( IR \) constitutes the terminal condition for the first state variable. There is no terminal condition for the second state variable. Instead, the second state variable has to satisfy \( IR − S \). In addition to these, \( x(\theta) \) must be non-increasing.
Unlike the supervisor in the current paper, their middle principal is limitedly liable. The top principal has to leave a strictly positive expected payoff to the middle principal if he chooses to delegate to her. Therefore delegating to the middle principal is outperformed by directly contracting with the productive agent. In the present model, since there is no limited liability constraint, any positive rent can be taxed away from the supervisor. Thus the supervisor’s expected payoff is 0 under the optimal delegation implementable outcome. Nevertheless, delegation still hurts the risk averse principal since it requires the supervisor’s payoff to be decreasing in the agent’s type.

4 Collusion Feasible Outcomes

In this section, I will not make any restrictions on the transfer or the wage schedules. Instead, I will consider $P$’s selection from the set of all the available grand contracts. I will start with characterization of the collusion feasible outcomes without referring to the grand contracts, through which they are feasible. Such a characterization requires a further analysis of the side contract selection process in (7). This process is an example to mechanism design problems with type dependent reservation utility. For this problem, identification of the binding participation constraints is not immediate. Depending on how the reservation utility responds to $\theta$, the constraint $Part(\theta)$ can be slack for $\theta$ and/or binding for types lower than $\theta$.

Type dependence of reservation utility also affects the incentives that govern the behavior of $A$. When the reservation utility is uniformly zero, we have already seen that $A$ has an incentive to overstate his type in order to increase the compensation he receives for the production costs. However, type dependent reservation utility may be a source for an additional incentive that countervails the original one: If the reservation utility is declining in type, $A$ may prefer to understate his type in order to increase the compensation for the forgone reservation utility. The optimal mechanism depends on which one of these incentives is dominant for each type. This type of mechanism design problems had been introduced by Lewis & Sappington (1989), and developed further by Maggi & Rodriguez-Clare (1995), and Jullien (2000).

In problem (7), even though the reservation utility depends on $A$’s type, this dependence is not completely arbitrary. The right hand side of constraint $Part(\theta)$ is the maximum value function of $A$’s payoff maximization if he directly responds to the transfer schedule $T(x)$ without colluding with $S$. Under this additional qualification, (7) can be transformed into an optimal control program, where $x(\theta)$ is the control variable and $t(\theta) - \theta x(\theta)$ is the state variable. Following the analyses of Maggi & Rodriguez-Clare (1995) and Jullien (2000), the solution to (7) is identified as follows:

---

18 As in Melumad, Mookherjee & Reichelstein (1995).

19 See Seierstadt & Sydsaeter (1987), Chapter 5 for how to account for the type dependent reservation utility.
Lemma 2 Given \( W(\cdot) \) and \( T(\cdot) \), an incentive compatible output-transfer profile \( \{x(\cdot), t(\cdot)\} \) that satisfies Part (\( \theta \)) for all \( \theta \) is a solution to S’s maximization problem (7) if and only if there exists a function \( \gamma(\cdot) \) defined on \([\theta, \bar{\theta}]\) such that

a) \( \gamma(\cdot) \) is non-decreasing,

b) \( \gamma(\cdot) \) is constant on any interval where Part (\( \theta \)) constraints are slack,

c) \( \gamma(\bar{\theta}) \leq 1 \) (with equality if Part (\( \bar{\theta} \)) is slack), \( \gamma(\theta) \geq 0 \) (with equality if Part (\( \theta \)) is slack),

and

\[
x(\theta) \in \arg \max_x \left\{ W(x) + T(x) - \theta x - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x \right\}
\]  \hspace{1cm} (14)

for all \( \theta \).


Jullien (2000) interprets \( \gamma(\theta) \) as the shadow value associated with the uniform marginal reduction of the reservation utility for all types between \( \theta \) and \( \bar{\theta} \). With this interpretation, it should be by no surprise that under delegation, where the only relevant participation constraint is Part (\( \bar{\theta} \)), \( \gamma(\theta) \) is equal to 0 for all \( \theta \). For an arbitrary side contract selection problem, if \( \gamma(\theta) \) assumes a value other than 0, this indicates that the participation constraint is binding for some type between \( \theta \) and \( \bar{\theta} \).

Following the same steps as in the analysis of delegation feasible outcomes, I will combine (14) above with equation (9) to provide a first order condition for collusion feasibility of \( w(\cdot) \).

Lemma 3 If \( \{x(\cdot), t(\cdot), w(\cdot)\} \) is collusion feasible, then there exists a function \( \gamma(\cdot) \) defined on \([\theta, \bar{\theta}]\) such that

a) \( \gamma(\cdot) \) is non-decreasing,

b) \( \gamma(\cdot) \) assumes values in \([0, 1]\),

c) \( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \) is non-decreasing,

\[
w(\theta) - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x(\theta) \text{ is continuous and}
\]

\[
\frac{d}{d\theta} \left[ w(\theta) - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x(\theta) \right] = - \left[ \frac{d}{d\theta} \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \right] x(\theta) \text{ a.e.}
\]  \hspace{1cm} (15)

Proof. See the appendix. ■

Unlike condition (12), condition (15) does not reveal the rate of change in \( w(\cdot) \) as a function of \( x(\cdot) \) only. The function \( \gamma(\cdot) \) is also relevant for the identification of the wage profile. For instance, whenever \( x(\cdot) \) is differentiable at \( \theta \), \( w(\cdot) \) is differentiable as well and its derivative is equal to \( \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x'(\theta) \). Note that the change in the wage profile at \( \theta \) has the same sign as \( \gamma(\theta) - F(\theta) \).

This is also consistent with the shadow value interpretation of \( \gamma(\theta) \). If \( \gamma(\theta) \) is smaller than \( F(\theta) \), then the dominant incentive for the type \( \theta \) agent is still exaggerating the production cost to increase his compensation. In this case, S’s payoff must be declining in the revealed type of A to encourage her
to prevent $A$'s overstatement of his type. The double marginalization effect is still present, i.e., $w(\cdot)$ is decreasing at $\theta$. But this effect is less severe than it was under delegation, unless $\gamma(\theta)$ is equal to 0.

On the other hand, if $\gamma(\theta)$ is larger than $F(\theta)$, it means that the participation constraints for types lower than $\theta$ have more weight in $S$’s optimization problem than do the participation constraints for types higher than $\theta$. This indicates the existence of some type, which is lower than $\theta$ and which has a large enough reservation utility. In this case, the countervailing incentive for the type $\theta$ agent dominates the original incentive: The agent of type $\theta$ should be motivated not to understate his type. This in turn requires that $S$ must be compelled to provide this motivation for the type $\theta$ agent. $S$ would be willing to prevent $A$ from understating his type only if her own payoff is increasing in the revealed type of $A$. Accordingly, $w(\cdot)$ must be increasing at $\theta$ if $\gamma(\theta) > F(\theta)$. For this latter case, counter marginalization replaces double marginalization.

It must be evident from the above discussion that condition (15) would be critical in establishing collusion feasibility of an increasing wage profile and proving that supervision is beneficial. However, Lemma 3 identifies (15) only as a necessary condition. With the following proposition I will prove that (15) is also sufficient for collusion feasibility, together with incentive compatibility of the output - transfer profile.

**Proposition 2** \{ $x(\cdot), t(\cdot), w(\cdot)$ \} is collusion feasible if and only if \{ $x(\cdot), t(\cdot)$ \} is incentive compatible and $w(\cdot)$ satisfies (15) where $\gamma(\cdot)$ is a function defined on $[\underline{\theta}, \overline{\theta}]$ such that

a) $\gamma(\cdot)$ is non-decreasing,

b) $\gamma(\cdot)$ assumes values in $[0, 1]$,

c) $\theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)}$ is non-decreasing.

**Proof.** Necessity follows from Lemma 3 and the incentive compatibility requirement of collusion feasibility. I will prove sufficiency by constructing a grand contract \{ $T(\cdot), W(\cdot)$ \} such that $Part(\theta)$ is binding, $x(\theta)$ satisfies (14) and $W(x(\theta)) + T(x(\theta)) = w(\theta) + t(\theta)$ for all $\theta$. Consider the following $T(\cdot)$ and $W(\cdot)$:

\[ T(x(\overline{\theta})) = t(\overline{\theta}), \text{ and } T'(x) = x^{-1}(x) \text{ for all } x, \] (16)

\[ W(x(\overline{\theta})) = w(\overline{\theta}), \text{ and } W'(x) = \frac{F(x^{-1}(x)) - \gamma(x^{-1}(x))}{f(x^{-1}(x))} \text{ for all } x. \] (17)

(16) implies that $t(\theta) - \theta x(\theta) = \max_x \{ T(x) - \theta x \}$ for all $\theta$. Since $x^{-1}(\cdot)$ is non-increasing and $\theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)}$ is non-decreasing, $T(\cdot) + W(\cdot)$ is concave. Therefore $W(x) + T(x) - \theta x - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x$ is maximized by $x(\theta)$. Equation $W(x(\theta)) + T(x(\theta)) = w(\theta) + t(\theta)$ follows from FO and (15).

Under the grand contract constructed to prove the sufficiency part of the above proposition, $x(\theta)$ is the optimal output choice for the type $\theta$ agent even when there is no interaction with $S$, i.e., $x(\theta) \in \arg\max_x \{ T(x) - \theta x \}$ for all $\theta$. Similarly, when $A$’s type is $\theta$, $t(\theta)$ and $w(\theta)$ are equal to the transfer
and wage levels that A and S would receive respectively, if there were no opportunity for them to collude. Therefore, we can interpret the grand contract \( \{ T(\cdot), W(\cdot) \} \) above as a collusion proof contract, where S is unable to find a side contract that would increase her expected payoff relative to the non-collusive play of the grand contract. Thus, a corollary to the construction of \( \{ T(\cdot), W(\cdot) \} \) is that any collusion feasible outcome can be induced by a collusion proof contract. This result is known as the collusion proofness principle in the literature.\(^{20}\)

Recall that Lemma 2 had stated some restrictions on the values that function \( \gamma(\cdot) \) can assume at \( \theta \), when the participation constraint is slack at \( \theta \). We do not see any such restriction in the statement of Proposition 2. This is due to the fact that, under the transfer schedule constructed in (16), \( t(\theta) - \theta x(\theta) = \max_x \{ T(x) - \theta x \} \) for all \( \theta \). Since all the participation constraints in S’s side contract selection problem (7) are binding, there is no reason to state restrictions for the regions where these constraints are slack.

For any non-increasing output profile \( x(\cdot) \) and any \( \gamma(\cdot) \) that satisfies the conditions in Proposition 2, there exists a transfer profile \( t(\cdot) \) and a wage profile \( w(\cdot) \) that are collusion feasible together with \( x(\cdot) \). This indicates that \( \gamma(\cdot) \) can be dealt as a choice variable for \( P \) at the grand contract selection stage. Moreover, selection of \( x(\cdot) \) and \( \gamma(\cdot) \) describes profiles \( t(\cdot) \) and \( w(\cdot) \) up to a constant.

From Proposition 2, it is easy to see that delegation feasible outcomes are a subset of collusion feasible outcomes that are supported by \( \gamma(\theta) = 0 \) for all \( \theta \).\(^{21}\) As mentioned above, this specific \( \gamma(\cdot) \) reflects the fact that the only relevant participation constraint is \( \text{Part} (\overline{\theta}) \) in the delegation problem. The other extreme case, where the only relevant constraint is \( \text{Part} (\bar{\theta}) \), is represented by \( \gamma(\theta) = 1 \) for all \( \theta \). One other situation that also calls for our attention is depicted by \( \gamma(\theta) = F(\theta) \) for all \( \theta \). This intermediate case can be considered as a replication of the no-supervision implementation, since the resulting \( w(\cdot) \) is constant everywhere.

We already know that the risk averse \( P \) is willing to implement an increasing wage profile to reduce the absolute value of the rate of change in the total payment he is making to the \( S - A \) pair. Such a wage profile can be achieved by choosing \( \gamma(\theta) \) larger than \( F(\theta) \) for all \( \theta \). In the following section I will identify the optimal \( \gamma(\cdot) \), and consequently the optimal \( w(\cdot) \), that would minimize \( P \)'s exposure to risk subject to the implementability constraints.

## 5 Optimal Wage Profile

In this section, I will identify the optimal implementable transfer - wage profile that would maximize \( P \)'s objective function (6) given a non-increasing output profile. Definition 4 outlines implementable

---

\(^{20}\)For the collusion proofness principle, see Tirole (1986) among others.

\(^{21}\)Delegation feasibility also requires constraint \( IR (\overline{\theta}) \) to be binding, i.e., \( t(\overline{\theta}) = \overline{\theta} x(\overline{\theta}) \). But this requirement is not material for \( P \)'s expected payoff maximization problem.
outcomes as the set of collusion feasible \( \{x(\cdot), t(\cdot), w(\cdot)\} \) that satisfies the individual rationality constraints \( IR - S \) and \( IR(\theta) \) for all \( \theta \). Incentive compatibility, which is a requirement of collusion feasibility, implies that \( IR(\theta) \) is sufficient for the other \( IR(\theta) \) constraints. Recall that collusion feasibility imposes conditions on the rate of change for both \( t(\cdot) \) and \( w(\cdot) \), but is silent on the intercepts of these functions. Since \( P \)'s objective function (6) is decreasing in both \( t(\theta) \) and \( w(\theta) \), optimality requires constraints \( IR(\theta) \) and \( IR - S \) to be binding.

The optimal transfer profile is fully characterized by the binding \( IR(\theta) \) constraint and incentive compatibility. And the binding \( IR - S \) constraint identifies the expected value of \( w(\theta) \) as 0. What is left out is the rate of change in \( w(\cdot) \), which is revealed by (15) as a function of \( \gamma(\cdot) \). Therefore \( P \)'s task of choosing the optimal transfer - wage profile reduces to choosing \( \gamma(\cdot) \) that would minimize the absolute value of the rate of change in \( t(\cdot) + w(\cdot) \).

In order to identify the optimal \( \gamma(\cdot) \), first note that the rate of change in \( t(\cdot) + w(\cdot) \) is revealed by adding up conditions \( FO \) and (15):

\[
\frac{d}{d\theta} \left[ t(\theta) + w(\theta) - \left( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \right) x(\theta) \right] = - \left[ \frac{d}{d\theta} \left( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \right) \right] x(\theta) \text{ a.e.} \quad (18)
\]

Whenever \( x(\cdot), t(\cdot), \) and \( w(\cdot) \) are differentiable at \( \theta \), this implies that \( t'(\theta) + w'(\theta) = \left( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \right) x'(\theta) \). Regardless of \( x(\cdot) \), a smaller absolute value for \( \left( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \right) \) indicates a lower rate of change for \( t(\cdot) + w(\cdot) \) at \( \theta \), and therefore a higher expected payoff for \( P \).

For smaller values of \( \theta \), setting \( \gamma(\theta) \) equal to \( F(\theta) + f(\theta) \theta \) reduces the term \( \left( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \right) \) and therefore the rate of change in total payment to 0. However, this is not possible for larger values of \( \theta \), since \( \gamma(\theta) \) is bounded from above by 1. For these values of \( \theta \), it would be optimal to set \( \gamma(\theta) \) equal to 1. This argument is formalized with the following proposition.

**Proposition 3** The optimal implementable transfer - wage profile that induces the non-increasing output profile \( x(\cdot) \) is fully identified by the binding \( IR(\theta) \) and \( IR - S \) constraints, as well as conditions \( FO \) and (15) for

\[
\gamma(\theta) = \begin{cases} 
F(\theta) + f(\theta) \theta, & \text{if } \theta \leq \theta^* \\
1, & \text{otherwise}
\end{cases} \quad (19)
\]

where \( \theta^* \) is the solution to \( F(\theta) + f(\theta) \theta = 1 \) if \( f(\theta) \theta < 1 \), and is equal to \( \theta \) otherwise.

**Proof.** (19) identifies \( \gamma(\cdot) \) that pointwise minimizes the absolute value of the rate of change in \( t(\cdot) + w(\cdot) \) subject to the constraint that \( \gamma(\cdot) \) assumes values in \([0,1]\). The proof of the proposition requires showing that the function in (19) satisfies the other constraints in Proposition 2, i.e., monotonicity of \( \gamma(\theta) \) and \( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \).
For $\theta \leq \theta^*$, $\theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)}$ is constant (and equal to 0). To prove that $\gamma(\cdot)$ is non-decreasing, consider $\theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)}$ as a function $\theta$ and $\gamma$, and totally differentiate it. This reveals

$$d\theta + \frac{d}{d\theta} \frac{F(\theta) - \gamma(\theta)}{f(\theta)} d\theta - \frac{1}{f(\theta)} d\gamma = 0 \quad (20)$$

for $\theta \leq \theta^*$ and $\gamma = \gamma(\theta)$. For values of $\gamma$ between 0 and 1, the monotone hazard rate condition implies $\frac{d}{d\theta} \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \geq 0$. Thus, it follows from the above equality that $\frac{d\gamma}{d\theta}$ is strictly positive for $\theta \leq \theta^*$.

For $\theta > \theta^*$, $\gamma(\cdot)$ is constant and equal to 1. Therefore, $\theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)}$ is equal to $\theta + \frac{F(\theta) - 1}{f(\theta)}$. It directly follows from the monotone hazard rate condition that this is a non-decreasing function of $\theta$.

For $\theta$ smaller than $\theta^*$, the optimal outcome induces a flat total payment from $P$ as a function of $A$’s type. (See Figure 3 for the optimal implementable transfer - wage profile.) Such an outcome is clearly Pareto inefficient for the supervisor - agent coalition, since the coalition could have reduced the output level (and therefore the total production cost) without changing the total payment $P$ would make to the coalition members. If $S$ were informed on the type of $A$, she would have been happy to offer him a bribe to encourage him to reduce the production level. Lacking the information on $A$’s type, $S$ is not willing to make such an offer. Because reducing the output level of type $\theta$ would make it easier to imitate type $\theta$ for all the types higher than $\theta$. This would require $S$ to pay a higher information rent to all these types to preclude such imitations. From $S$’s point of view, the cost of this rise in the information rent is as high as the benefit of reducing the output for type $\theta$. In other words, $S$’s expected surplus maximization distorts the outcome away from coalitional Pareto efficiency.

For $\theta$ larger than $\theta^*$, the distortion from coalitional efficiency is not large enough to sustain a flat payment to the coalition. This is due to the fact that there are not that many types higher than $\theta$. Therefore $S$ is less concerned about types that would consider imitating $\theta$ following a decline in type $\theta$ agent’s output level. Even though the payment profile is not completely flat, it is still smoother than the no-supervision transfer profile (since $\gamma(\theta)$ is uniformly larger than $F(\theta)$). This leads to the following corollary.

**Corollary 2** Suppose $x(\cdot)$ is a non-increasing output profile. The optimal implementable transfer - wage profile that induces $x(\cdot)$ yields a weakly higher expected payoff for $P$ than does the optimal no-supervision transfer profile. The ordering is strict if $V(\cdot)$ is strictly convex and $x(\cdot)$ is not constant.

As is the case for Corollary 1, the corollary above does not require the output profile to be chosen optimally. Supervision is beneficial regardless of the output profile $P$ intends to implement. Also note

---

22Note that the magnitude of the distortion from coalitional Pareto efficiency declines in the type of the agent and completely vanishes for type $\theta$. This is in contrast with what we have observed for the risk neutral principal’s optimal no-supervision outcome. This distinction arises from $A$’s incentive reversal under type dependent reservation utility.
that the optimal $\gamma(\cdot)$ in (19) is fully revealed by the distribution of $\theta$ and does not depend on the output profile.\footnote{The optimal implementable output profile can be unraveled by an optimal control program, where (6) is the objective function, $x(\theta)$ is the control variable, $t(\theta) - \theta x(\theta)$ and $w(\theta) - F(\theta) - \frac{\gamma(\theta)}{f(\theta)} x(\theta)$ are the state variables. $\gamma(\theta)$ here is identified as in (19). The dynamics of the system are given by conditions $FO$ and (15). $IR(\overline{\theta})$ constitutes the terminal condition for the first state variable. There is no terminal condition for the second state variable. Instead, the second state variable has to satisfy $IR - S$. In addition to these, $x(\theta)$ must be non-increasing.}

As is clear from the construction of the optimal $\gamma(\cdot)$, provision of an outside option to $A$ plays an important role in the implementation of the optimal wage profile. This is similar to Mookherjee & Tsumagari’s (2003) construction of the grand contract that outperforms delegation to one of the two productive agents in their model. Under the grand contract they characterize, they show that the participation constraints for all types can be replaced by a single ex-ante participation constraint that requires leaving $A$ a predetermined expected utility level. With the notation of the present paper, this corresponds to considering outcomes that can be supported by $\gamma(\theta) = \lambda F(\theta)$ for all $\theta$, where $\lambda \in [0, 1]$. In their environment, Mookherjee & Tsumagari (2003) show that $P$ can reduce the effect of double marginalization by choosing $\lambda$ larger than 0 and therefore improve over delegation (which corresponds to $\lambda = 0$). This result is valid in the current environment as well. However, here, by using such a grand contract, $P$ cannot improve over his optimal no-supervision payoff. Beneficial supervision is possible only with counter marginalization of information rents, which in turn requires the manipulation of the type dependence of the reservation utility.

Mookherjee & Tsumagari (2003) also study the effect of delegation to a completely informed but unproductive middleman. Since this middleman knows the type of the agents he faces, the side contract he offers supports an efficient outcome for the middleman - agents coalition. Neither double marginalization nor counter marginalization is possible under such a scenario. When the products of the two agents are substitutes, Mookherjee & Tsumagari (2003) show that this alternative form of delegation is also dominated by a grand contract. Counter marginalization plays a role in the construction of this second grand contract.

The way that beneficial supervision is sustained in the current paper is also similar to how it could be sustained when $P$ is risk neutral, but $S$ is partly informed on $A$’s type. To illustrate this point, I will construct an example that originates from Celik (2003). Suppose $S$’s information structure takes the form of the partition $\{[\tilde{\theta}, \bar{\theta}], (\bar{\theta}, \overline{\theta}]\}$, where $\theta < \tilde{\theta} < \overline{\theta}$. That is, $S$ does not know the exact realization of $A$’s type but she is informed on whether it is smaller or larger than $\tilde{\theta}$. Suppose $P$ is willing to implement the non-increasing output profile $x(\cdot)$. In the absence of supervision, the optimal transfer profile that would induce $x(\cdot)$ is identified by equation $t(\overline{\theta}) = \overline{\theta} x(\overline{\theta})$ and condition $FO$, as it is identified in the current model. On the other hand, when the partly informed supervisor is present, the type $\tilde{\theta}$
agent cannot imitate higher types without $S$’s consent. Therefore there is no reason to leave type $\tilde{\theta}$ a positive information rent. As a result, $P$ can sustain a downward shift in $t(\cdot)$ for $\theta \in [\underline{\theta}, \tilde{\theta}]$ by setting $t(\tilde{\theta}) = \tilde{\theta}x(\tilde{\theta})$. (See Figure 4.) If collusion is not a possibility, $S$ will truthfully reveal her information to $P$ even in the absence of a wage payment from $P$.

If $S$ and $A$ can sign a side contract to coordinate their interaction with $P$ and make side transfers to each other, then $P$ has to make a wage payment to $S$ whenever she reveals that $A$’s type is in the interval $[\bar{\theta}, \tilde{\theta}]$. Otherwise $S$ will have an incentive to lie to $P$ in exchange of a small bribe from $A$. Therefore, the above discontinuity in the total payment disappears. However this does not indicate that supervision is useless. By employing the type dependence of $A$’s reservation utility, $P$ can sustain a wage profile that is decreasing in $A$’s type in $[\bar{\theta}, \tilde{\theta}]$. The resulting total payment is less responsive to $\theta$ on the interval $[\bar{\theta}, \tilde{\theta}]$ than is the agent’s transfer. Accordingly, the total payment that $P$ makes under supervision is uniformly (weakly) lower than the no-supervision transfer profile.

6 Robustness of Collusion Feasibility

With Definition 3, I outlined the collusion feasible outcomes as the output - transfer - wage profiles that are invulnerable to deviations that satisfy incentive compatibility and participation constraints of $A$, and make $S$ strictly better off. A stronger feasibility concept is defined by preclusion of deviations that would make $S$ weakly better off as well. It turns out that this additional requirement does not reduce the set of feasible outcomes much further.

**Proposition 4** Suppose $x(\cdot)$ is “strictly” decreasing and incentive compatible together with $t(\cdot)$. Suppose $w(\cdot)$ satisfies (15) where $\gamma(\cdot)$ is a non-decreasing function, which assumes values in $[0, 1]$ and which yields a “strictly” increasing $\theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)}$. Let $T(\cdot)$ and $W(\cdot)$ be given as in (16) and (17). Then $\{x(\cdot), t(\cdot)\}$ is the essentially unique solution to $S$'s side contract selection problem (7).\(^{25}\)

**Proof.** It follows from the construction of $\{T(\cdot), W(\cdot)\}$ and Lemma 2 that $\{x(\cdot), t(\cdot)\}$ is a solution to (7). To see that this solution is essentially unique, note that incentive compatibility of $\{x(\cdot), t(\cdot)\}$ requires $x(\cdot)$ to be continuous a.e. Let $\theta$ be such a point where the output profile is continuous. It follows from the construction of $x^{-1}(\cdot)$ and strict monotonicity of the output profile that $x^{-1}(\cdot)$ is strictly decreasing around $x(\theta)$. Since $\theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)}$ is strictly increasing, $T'(\cdot) + W'(\cdot)$ is also strictly decreasing around $x(\theta)$. Therefore there exists an interval where $T(\cdot) + W(\cdot)$ is strictly concave and

\(^{24}\)To be accurate, the discontinuity in the total information rent disappears. If $x(\cdot)$ is discontinuous at $\tilde{\theta}$, the total payment is still discontinuous.

\(^{25}\)That is, there exists no $\{\hat{x}(\cdot), \hat{t}(\cdot)\}$ that is a solution to $S$’s problem and that differs from $\{x(\cdot), t(\cdot)\}$ on a subset of $[\theta, \tilde{\theta}]$, which has positive probability.
\( x(\theta) \) is an interior point. Recall that this is true for almost every \( \theta \). Since (7) is a concave program (the objective function is concave and the constraints outline a convex set), strict concavity of the objective function around \( x(\theta) \) requires the optimal output profile to be essentially unique. Condition \( FO \) indicates that the optimal output profile reveals the optimal transfer profile up to \( t(\theta) \). And the constant \( t(\theta) \) is uniquely identified by the participation constraints (\( Part(\theta) \) should be binding for some \( \theta \)).

As long as \( \{x(\cdot), t(\cdot), w(\cdot)\} \) is collusion feasible and the additional requirements of Proposition 4 on \( x(\cdot) \) and \( w(\cdot) \) are satisfied, there exists no incentive compatible \( \{\hat{x}(\cdot), \hat{t}(\cdot)\} \) essentially different than \( \{x(\cdot), t(\cdot)\} \),

(i) that does not decrease \( S \)'s expected surplus level:

\[
\int_\theta T(\hat{x}(\theta)) + W(\hat{x}(\theta)) - \hat{t}(\theta) \ d\theta \geq \int_\theta [T(x(\theta)) + W(x(\theta)) - t(\theta)] f(\theta) d\theta. 
\] (21)

and

(ii) that does not decrease \( A \)'s ex-post utility levels:

\[
\hat{t}(\theta) - \theta \hat{x}(\theta) \geq t(\theta) - \theta x(\theta) \text{ for all } \theta. 
\] (22)

The strict monotonicity requirements of Proposition 4 reduce the set of feasible outcomes to an open set. However the closure of this set is identical to the original collusion feasible set characterized in Proposition 2. Even though the strict monotonicity requirements rule out the optimal wage profile identified in Proposition 3, \( P \) can obtain this wage profile as the limit of a sequence of feasible wage profiles.

In light of this discussion, an alternative formulation of collusion can be analyzed. Suppose a fourth party makes the side contract offer that consists of a suggested bribe schedule in period \( T_2 \). Then, \( S \) and \( A \) decide to accept or reject this offer. This outside facilitator approach to collusion has been used by Laffont & Martimort (1997 and 1998). The question of interest is whether the collusion feasible outcomes (satisfying Definition 3) are still attainable under this alternative assumption.

Suppose \( P \) wants to implement the collusion feasible outcome \( \{x(\cdot), t(\cdot), w(\cdot)\} \). To do so he offers the grand contract \( \{T(\cdot), W(\cdot)\} \) that is given by (16) and (17). If the fourth party does not offer a side contract (or offers the null side contract \( B_0(x) = 0 \) for all \( x \)), the grand contract induces \( \{x(\cdot), t(\cdot)\} \) as the resulting output - transfer profile. As before, this constitutes the outside option of collusion for \( S \) and \( A \).

The side contract \( B(\cdot) \) may induce another output - transfer profile \( \{\hat{x}(\cdot), \hat{t}(\cdot)\} \). For type \( \theta \) agent not to imitate another type, this profile has to be incentive compatible. Moreover, the fourth party must make sure that \( S \) is willing to collude. Therefore \( \{\hat{x}(\cdot), \hat{t}(\cdot)\} \) must provide an expected surplus for \( S \) at least as high as the surplus from \( \{x(\cdot), t(\cdot)\} \). Accordingly, (21) must hold. Similarly, to compel \( A \) to
participate in collusive contract regardless of his realized type, \( \{ \hat{x}(\cdot), \hat{t}(\cdot) \} \) must provide all types of \( A \) with an ex-post utility at least as high as his reservation utility. Accordingly, (22) must hold. It follows from Proposition 4 that any such \( \{ \hat{x}(\cdot), \hat{t}(\cdot) \} \) is essentially equivalent to \( \{ x(\cdot), t(\cdot) \} \), as long as \( x(\theta) \) and \( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \) are strictly monotonic. In other words, the fourth party cannot find a side contract that would implement an outcome that is essentially different than \( \{ x(\cdot), t(\cdot), w(\cdot) \} \). Note that this is regardless of the agenda (the objective function) of the fourth party, who designs the side contract.

Another alternative formulation of collusion would have \( A \) as the party that makes the collusive offer in period \( T2 \). Since \( A \) is already informed on the realization of his type, his side contract offer may reveal his type (or some information about his type) to \( S \). However, it follows from Myerson’s (1983) inscrutability principle that there is no loss of generality in assuming all types of \( A \) offer the same side contract. As above, suppose that \( P \) offers the grand contract \( \{ T(\cdot), W(\cdot) \} \) in (16) and (17) to implement the collusion feasible outcome \( \{ x(\cdot), t(\cdot), w(\cdot) \} \). The side contract \( B(\cdot) \) would induce \( \{ \hat{x}(\cdot), \hat{t}(\cdot) \} \). Again, for type \( \theta \) agent not to imitate another type, \( \{ \hat{x}(\cdot), \hat{t}(\cdot) \} \) must be incentive compatible. \( S \) must be given the incentive to participate in the side contract. Therefore (21) must hold. And all types of \( A \) must prefer making such a side contract offer to responding to the grand contract non-cooperatively. Therefore (22) must hold. In the end, \( \{ x(\cdot), t(\cdot) \} \) is still the essentially unique profile that would satisfy all these conditions, as long as \( x(\theta) \) and \( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \) are strictly monotonic.

7 Conclusion

In an organization, there are several reasons for the compensation of a player taking part in the organization to depend on the conduct of another player:

1) There may be an externality related to the actions taken by different players: The value added to the organization through these actions may not be a separable function. For instance, if these actions are of complementary nature, and one of the players has indicated that his action will be high, the organization is willing to compensate the other players more in order to increase their actions as well.

2) One of the players may bear some relevant information about the type (say, the production cost) of another player. In this case, the optimal compensation scheme that the latter player will encounter depends on the former player’s report to the principal of the organization regarding the type of the latter player.

3) One of the players (supervisor) may be used in order to insure the organization against the uncertainty arising from the conduct of another player (agent). For instance, if the agent takes an action that requires a high compensation for him, the supervisor’s compensation can be decreased to balance the total compensation made to these players.

If any of these interdependent compensation structures is enforced, the affected players can increase
their payoffs by coordinating their response to the organization and by making side payments to each other. For instance, in case 3 above, the supervisor has the incentive to bribe the agent in return for the agent’s not taking the action that calls for a low compensation for the supervisor. The standard incentive compatibility constraints are not sufficient to account for this additional incentive that arises from the collusion potential. The organization has to take this collusion potential into consideration in the design of the compensation structure.

In this paper, I studied an environment, where the interdependence within the compensation structure stems from the desire to insure the organization (as in case 3 above). The productive agent, who knows the production cost, chooses an output level to produce. The organization (the principal of the organization) makes a transfer payment to him increasing in the output level. Therefore the uncertainty of the production cost translates into the uncertainty of the transfer level. To alleviate this uncertainty, the principal tries implementing a wage payment for the supervisor that is decreasing in the output level.

In this environment, I characterized the set of outcomes that are implementable despite the threat of collusion. Unlike for the cases 1 and 2 that are studied elsewhere, the characterization result here made it possible to identify the optimal compensation structure for the organization. This structure induces a negative correlation between the transfer and wage levels. Therefore, the supervisor can indeed provide some insurance for the organization. However, the optimal compensation structure falls short of attaining full insurance. Even though there may be a region of output levels, where the total payment made to these two players is constant, there is always another region, where it is strictly increasing.

Implementation of the optimal compensation structure requires the organization’s contracting with both of the players. That is, the optimal structure cannot be replicated by delegating to the supervisor the authority to contract with the agent on the organization’s behalf. In fact, the principal of the organization prefers not having any access to the supervisor to following such a delegation scheme.

As a final note I want to point out an implication of the paper on the size of bureaucracies. Since employing a single supervisor does not totally resolve the uncertainty in the total payment, the organization may benefit from employing an additional supervisor. This second supervisor of course is allowed to collude with the agent in the same manner that the first supervisor does. However, by adding this second supervisor to its ranks, the organization reduces the variance of the total payment further. However, full insurance is still not attainable. Actually, regardless of the number of already employed supervisors, there is always a gain for the organization to employ one more supervisor. Full insurance is approximated as the number of supervisors approaches to infinity. This suggests that the potential for collusion can be a rationale for designing a bureaucratic structure within the organization, which consists of supervisors.

26However, the supervisors are not allowed to collude with each other to bribe the agent collectively. Otherwise, the two-supervisor setup is qualitatively identical to the single supervisor setup.
that do not take a direct part in the production process.

8 Appendix

8.1 Proof of Lemma 1

Suppose \( \{x(\cdot), t(\cdot), w(\cdot)\} \) is delegation feasible. It follows from (11) that

\[
W(x(\hat{\theta})) - \theta x(\theta) - \frac{F(\theta)}{f(\theta)} x(\theta) \geq W(x(\hat{\theta})) - \theta x(\hat{\theta}) - \frac{F(\theta)}{f(\theta)} x(\hat{\theta})
\]

(23)

for all \( \hat{\theta} \) and \( \theta \). When we substitute in (9) in the above inequality, considering \( T_0(x) = 0 \) for all \( x \),

\[
t(\theta) + w(\theta) - \theta x(\theta) - \frac{F(\theta)}{f(\theta)} x(\theta) \geq t(\hat{\theta}) + w(\hat{\theta}) - \theta x(\hat{\theta}) - \frac{F(\theta)}{f(\theta)} x(\hat{\theta})
\]

(24)

or

\[
r(\theta) \geq r(\hat{\theta}) - \left[ \theta + \frac{F(\theta)}{f(\theta)} - \hat{\theta} - \frac{F(\hat{\theta})}{f(\hat{\theta})} \right] x(\hat{\theta})
\]

(25)

for all \( \hat{\theta} \) and \( \theta \), where \( r(\theta) = t(\theta) + w(\theta) - \theta x(\theta) - \frac{F(\theta)}{f(\theta)} x(\theta) \). Changing the roles of \( \theta \) and \( \hat{\theta} \), and merging the inequalities,

\[
- \left[ \theta + \frac{F(\theta)}{f(\theta)} - \hat{\theta} - \frac{F(\hat{\theta})}{f(\hat{\theta})} \right] x(\theta) \geq r(\theta) - r(\hat{\theta}) \geq - \left[ \theta + \frac{F(\theta)}{f(\theta)} - \hat{\theta} - \frac{F(\hat{\theta})}{f(\hat{\theta})} \right] x(\hat{\theta}).
\]

(26)

Since \( \{x(\cdot), t(\cdot)\} \) is incentive compatible, it must be that \( x(\cdot) \) is monotonic and therefore continuous almost everywhere. For all points of continuity for \( x(\theta) \), as \( \hat{\theta} \) approaches to \( \theta \), the above inequalities imply

\[
r'(\theta) = - \frac{d}{d\theta} \left[ \theta + \frac{F(\theta)}{f(\theta)} \right] x(\theta).
\]

(27)

The function \( r(\cdot) \) is continuous, since it is a maximum value function. Also recall that incentive compatibility requires that \( t(\theta) - \theta x(\theta) \) is continuous and differentiable almost everywhere. Therefore, \( w(\theta) - \frac{F(\theta)}{f(\theta)} x(\theta) \) is continuous and differentiable almost everywhere. Another equation describing \( r'(\cdot) \) can be constructed from the definition of \( r(\cdot) \):

\[
r'(\theta) = \frac{d}{d\theta} [t(\theta) - \theta x(\theta)] + \frac{d}{d\theta} \left[ w(\theta) - \frac{F(\theta)}{f(\theta)} x(\theta) \right] \text{ a.e.}
\]

(28)

Since \( t(\theta) - \theta x(\theta) \) has the derivative \(-x(\theta)\) almost everywhere,

\[
\frac{d}{d\theta} \left[ w(\theta) - \frac{F(\theta)}{f(\theta)} x(\theta) \right] = - \left[ \frac{d}{d\theta} \frac{F(\theta)}{f(\theta)} \right] x(\theta) \text{ a.e.}
\]

(29)
8.2 Proof of Lemma 3

Suppose \( \{x(\cdot), t(\cdot), w(\cdot)\} \) is collusion feasible. Existence of non-decreasing \( \gamma(\cdot) \) assuming values in \([0, 1]\) and satisfying (14) follows from Lemma 2. Following the proof of Lemma 1, (14) implies that

\[
- \left[ \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} - \hat{\theta} - \frac{\hat{F}(\hat{\theta}) - \gamma(\hat{\theta})}{\hat{f}(\hat{\theta})} \right] x(\theta) \geq s(\theta) - s(\hat{\theta}) \geq - \left[ \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} - \hat{\theta} - \frac{\hat{F}(\hat{\theta}) - \gamma(\hat{\theta})}{\hat{f}(\hat{\theta})} \right] x(\hat{\theta}).
\]

(29)

for all \( \hat{\theta} \) and \( \theta \), where \( s(\theta) = t(\theta) + w(\theta) - \theta x(\theta) - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x(\theta) \).

- There exists a non-decreasing \( \gamma(\cdot) \), which assumes values in \([0, 1]\) and which yields a non-decreasing \( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \) that satisfies (14).

Suppose \( \theta > \hat{\theta} \). If \( x(\theta) < x(\hat{\theta}) \), the inequalities in (29) imply that

\[
\theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \geq \hat{\theta} + \frac{\hat{F}(\hat{\theta}) - \gamma(\hat{\theta})}{\hat{f}(\hat{\theta})}.
\]

(30)

On the other hand, if \( x(\hat{\theta}) = x(\theta) = \hat{x} \), then (30) is not a requirement for \( \gamma(\cdot) \) to satisfy (14). However, even if \( \gamma(\cdot) \) does not satisfy (30) for all \( \theta \) and \( \hat{\theta} \), another function on \([\underline{\theta}, \overline{\theta}]\) (the function \( \gamma^*(\cdot) \)) that would satisfy (30) can be constructed as follows:

\[
\theta + \frac{F(\theta) - \gamma^*(\theta)}{f(\theta)} = \sup_{\theta' \leq \theta} \left\{ \theta' + \frac{F(\theta') - \gamma(\theta')}{f(\theta')} \right\}.
\]

(31)

By construction, \( \gamma^*(\theta) = \gamma(\theta) \) on the intervals, where \( x(\cdot) \) is strictly increasing. Moreover \( x(\theta) \) still satisfies (14), when \( \gamma(\cdot) \) is replaced by \( \gamma^*(\cdot) \) (since \( x(\cdot) \) is constant whenever \( \gamma^*(\theta) \neq \gamma(\theta) \)). To prove the claim, I have to show that \( \gamma^*(\cdot) \) is non-decreasing and assumes values in \([0, 1]\).

First, note that \( \theta + \frac{F(\theta) - \gamma^*(\theta)}{f(\theta)} \) is continuous since \( \gamma(\cdot) \) is non-decreasing. And since \( F(\cdot) \) and \( f(\cdot) \) are continuous functions, \( \gamma^*(\cdot) \) is continuous as well. For each interval, where \( \gamma^*(\theta) = \gamma(\theta) \), the function \( \gamma^*(\cdot) \) is non-decreasing and assumes values in \([0, 1]\). This follows from the properties of the function \( \gamma(\cdot) \).

For each interval, where \( \gamma^*(\theta) \neq \gamma(\theta) \), it follows from the definition of \( \gamma^*(\cdot) \) that \( \theta + \frac{F(\theta) - \gamma^*(\theta)}{f(\theta)} \) is constant. For these latter intervals, consider \( \theta + \frac{F(\theta) - \gamma^*(\theta)}{f(\theta)} \) as a function of \( \theta \) and \( \gamma^* \). Then totally differentiate it. This reveals

\[
d\theta + \frac{d}{d\theta} \frac{F(\theta) - \gamma^*(\theta)}{f(\theta)} d\theta - \frac{1}{f(\theta)} d\gamma^* = 0,
\]

(32)

and after rearranging

\[
\frac{d\gamma^*}{d\theta} = f(\theta) \left[ 1 + \frac{d}{d\theta} \frac{F(\theta) - \gamma^*}{f(\theta)} \right].
\]

(33)
This differential equation, together with an initial value, reveals \( \gamma^* (\cdot) \) for the relevant interval. Since \( \gamma^* (\cdot) \) is a continuous function, the initial value will be in \([0, 1]\). It follows from the monotone hazard rate condition that \( \frac{d}{d\theta} \frac{F(\theta) - \gamma^*}{f(\theta)} \) is positive for \( \gamma^* \in [0, 1] \). Definition (31) implies that \( \gamma^* (\theta) \leq \gamma(\theta) \leq 1 \) for all \( \theta \). This last condition guarantees that \( \gamma^* (\cdot) \) is an increasing function that assumes values between 0 and 1 whenever \( \gamma^* (\theta) \neq \gamma(\theta) \).

- (29) and incentive compatibility of \( \{x(\cdot), t(\cdot)\} \) imply continuity of \( w(\theta) - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x(\theta) \) and (15).

Since \( \gamma(\theta) \) and \( \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \) are both non-decreasing, the latter function must be continuous. Since \( \{x(\cdot), t(\cdot)\} \) is incentive compatible, \( x(\cdot) \) must be non-increasing and therefore continuous almost everywhere. After taking the limit as \( \hat{\theta} \) approaches \( \theta \), condition (29) implies

\[
s'(\theta) = \frac{d}{d\theta} \left[ \theta + \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \right] x(\theta)
\]

almost everywhere.

The function \( s(\cdot) \) is continuous, since it is a maximum value function. Also recall that \( t(\theta) - \theta x(\theta) \) is continuous and differentiable almost everywhere. Therefore \( w(\theta) - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x(\theta) \) is also continuous and differentiable almost everywhere. Another expression for \( s'(\cdot) \) can be constructed from the definition of \( s(\theta) \):

\[
s'(\theta) = \frac{d}{d\theta} [t(\theta) - \theta x(\theta)] + \frac{d}{d\theta} \left[ w(\theta) - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x(\theta) \right] \quad \text{a.e.}
\]

Since \( t(\theta) - \theta x(\theta) \) has the derivative \( -x(\theta) \) almost everywhere,

\[
\frac{d}{d\theta} \left[ w(\theta) - \frac{F(\theta) - \gamma(\theta)}{f(\theta)} x(\theta) \right] = - \left[ \frac{d}{d\theta} \frac{F(\theta) - \gamma(\theta)}{f(\theta)} \right] x(\theta) \quad \text{a.e.}
\]

References


Figure 1: The optimal no-supervision transfer profile

![Graph showing the optimal no-supervision transfer profile.](image)

Figure 2: The optimal delegation implementable transfer – wage profile

![Graph showing the optimal delegation implementable transfer – wage profile.](image)
Figure 3: The optimal implementable transfer – wage profile

Figure 4: An implementable transfer – wage profile with a partly informed supervisor

$t''(\theta)$ = no-supervision transfer profile