Mechanism Design and Direct Communication in Collusion

Tetsuya Maruyama[1,2]
University of Pennsylvania

November 16, 2005

[1]Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, Email: tetsuya3@econ.upenn.edu

[2]I am thankful to my main advisor professor Steven Matthews for very detailed feedback and patient support. I am grateful to my dissertation committee members professor George Mailath and Nicola Persico for incisive comments. I also thank Jing Li, and Ichiro Obara, and audiences at the Economic Theory workshop at University of Pennsylvania. All errors are my own responsibility.

Abstract

This paper studies a principal-supervisor-agent model in which a privately informed supervisor is susceptible to collusion. We model collusion as a side mechanism proposed by the supervisor who observes a private signal correlated with the agent's type. We explore how the supervisory information helps the principal to extract information rent from the agent when the informational asymmetry between the supervisor and agent results in inefficient collusion on their part.

We first analyze an informed principal problem with correlated signals in our setting, and show conditions under which the existence of an informed principal problem does not affect the optimal design of the grand mechanism. Under those conditions, the optimal mechanism is the same as when collusion is formed by an uninformed benevolent outsider.

The second part of this chapter focuses on the supervisor's participation decisions once she obtains new information in collusion. The principal can take advantage of the implicit information revealed by the supervisor's exit decision when the coalition has limited ability to control it. If the coalition cannot write a contract based on the exit decision, or the principal can distinguish the timing of the supervisor's exit decisions, the principal using a simple mechanism can attain the outcome of the direct supervision benchmark in which she can directly observes the supervisor's signal.

Keywords: collusion, monitoring, asymmetric information, mechanism design

1 Introduction

This paper concerns optimal contracts for an organization in which an employee coalition may form to collude against the principal. The focus is on situations in which the ability of the coalition members to collude is restricted because they each have potentially useful private information that they may want to misrepresent. Our specific model concerns an agent who produces output at a privately known cost, and a supervisor who privately observes an imperfect signal of that cost. Suppose the principal attempts to give the agent high-powered incentives by basing his compensation on the signal the supervisor reports. Then the agent may want to bribe the supervisor to report whatever signal maximizes his compensation from the principal – paying the bribe from the resulting increase in his compensation. More generally, in order to maximize their total compensation the agent and supervisor may want to collude in both the output the agent chooses and the signal the supervisor reports. However, their ability to collude in this way is generally imperfect since neither knows the private information of the other. The purpose of this paper is to determine how much the possibility of this "communicative collusion" restricts the ability of a principal to induce effort and extract rents from her employees, and in what way this possibility affects the nature of optimal contracts in the organization.

Collusion between monitoring supervisors and productive agents is an issue in a variety of real arenas. For example:

Regulation: A government uses the reports of an investigatory agency to set the rates of a regulated public utility with privately known costs. Collusion between the utility firm and the agency is a manifestation of "regulatory capture."

Corporate Governance: The shareholders of a firm with privately informed management may hire an auditor to obtain independent information about the firm. But the firm's managers have an incentive to persuade the auditor to misreport (Kofman and Lawarrée, 1993).

Tax Auditing: Taxpayers are audited by auditors hired by governments. The taxpayer and auditor may have an incentive to collude by having the auditor misreport. This problem may be alleviated if the government compensates the auditor on the basis of how much tax he collects.

Job Applications: A recommendation letter to a potential employer is written for an academic job candidate by his Ph.D. supervisor. The contents of the letter may

be the result of a collusive arrangement between the job candidate and the supervisor (although this is very unlikely.)

This paper builds on the work on collusion under asymmetric information by Laffont and Martimort (1997, 2000) and Faure-Grimaud, Laffont, and Martimort (2003), which in turn built on the seminal work of Tirole (1986) on collusion under symmetric information. The central idea is that the organizational form can be understood as the result of an optimal "grand contract" that the principal offers to maximize profits subject to certain constraints. These constraints are incentive constraints, participation constraints, and now collusion-proof constraints, which insure that the employees are unable to gain by collusion. In Laffont and Martimort (1997), the colluding agents have independent information, and hence their model does not apply to the case of a monitoring supervisor. Laffont and Martimort's (2000) public good provision model has homogeneous agents with correlated types drawn from a symmetric joint distribution, while our model concerns an asymmetric environment. By removing the payoff relevant component from the supervisor's signal, we can focus on how asymmetric information within the coalition affects the values of the supervisory information in helping the principal to extract rent from the agent.

Our model departs from these works in that collusion mechanism is proposed by the supervisor rather than by a benevolent uninformed outsider. The main purpose of this modification is to examine how the information exchanged through direct communication between colluding parties affects the effectiveness of collusion formation and the optimal contract the principal proposes to her employees.[1]

Specifically, we will study two issues. One is an informed principal problem within the coalition. When the proposer of a contract has private information, there is additional source of inefficiency due to signaling problem from the proposer's side. Another issue is enforceability of side mechanism when the coalition members' beliefs are updated through direct communication. Through direct communication colluding parties exchange information among each other, and they might not want to follow the collusive agreement given their updated beliefs. The principal may find it profitable to renegotiate the grand mechanism when the implicit information revelation occurs through the breach of the side mechanism. In other words, the principal may benefit from the information that is beyond control of the coalition.

The first result shows that the optimal collusion-proof mechanism entails the same equilibrium outcome as when the collusion is proposed by an outsider. The supervisor tries to maximize the joint payoff for the coalition subject to the incentive constraints

[1] Quesada (2004) addresses the question of collusion in mechanisms under asymmetric information. She analyzes collusion as an informed principal problem, in which one of the agents makes a side mechanism offer. Unlike her model, we consider correlated type cases, and we focus on information leakage through collusion process.

and make transfer payments to the agent to satisfy the agent's participation constraints. The informed principal problem only affects the virtual valuations for the coalition, and the principal cannot design a grand mechanism to exploit the additional distortion generated by the informed principal problem.

The second result concerns the situations in which the principal cannot prevent the supervisor to leave after the collusion stage. The principal can achieve the direct supervision benchmark output and profit level if the side mechanism is incomplete in the sense it cannot specify monetary transfer rules based on the supervisor's exit decision. We show that the principal can attain this goal with a combination of simple contract and renegotiation. The principal can use the implicit information revelation through the supervisor's action to renegotiate the grand mechanism. We also do informal analysis on whether the principal can achieve the same outcome with this simple contract when the supervisor can include in the side mechanism monetary transfer rules based on the supervisor's exit decision. The coalition members may be able to manage to minimize the information revelation through the supervisor's actions.

In the next section the model is laid out. We then show some benchmark non-collusive equilibrium outcomes. In Section 3 we first describe the collusion mechanism as an informed principal problem with correlated signals. We then characterize the set of weakly collusion-proof mechanism and compare it to the standard cases in which collusion is formed by an uninformed benevolent outsider. In Section 4, we consider the cases in which the supervisor cannot commit to stay in the grand mechanism after she obtains information during collusion stage. Proofs missing in the text are relegated to the Appendix.

2 The Model

2.1 Players, Preferences, and Information

There are three players, a principal, an agent, and a supervisor. The principal receives utility from a level of output, $x \geq 0$, and monetary transfers to the agent ($t \in \Re$) and the supervisor ($w \in \Re$). The principal has a utility function $R(x) - w - t$, where $R(\cdot)$ is a twice differentiable strictly concave increasing function.

The agent is a productive unit who provides x. The marginal cost of production, c, is drawn from a type space $C = \{c_1, c_2\}$, where $0 < c_1 < c_2$. His utility function is given by $t - cx$. Let $\Delta c \equiv c_2 - c_1$.

The agent's type c is private information. The supervisor obtains a signal, σ, which is correlated with c. We assume that the signal space consists of two values, so let $\Sigma = \{\sigma_1, \sigma_2\}$. The cost of obtaining a signal is independent of c and normalized

to zero. The supervisor's utility depends only on w, the monetary rewards from the principal. The supervisor is risk neutral and has a utility function $V(w) = w$.

Let $p_{ij} = \Pr(\sigma_i, c_j)$. To exclude trivial cases we assume that $p_{ij} > 0 \ \forall i, j$. Let $\rho \equiv p_{11}p_{22} - p_{12}p_{21}$. Without loss of generality we assume that $\rho > 0$, or $\frac{p_{22}}{p_{21}} > \frac{p_{12}}{p_{11}}$, which is satisfied if the signal $\sigma = \sigma_k$ indicates that $c = c_k$ more likely than $c = c_l$, $l \neq k$. The higher is ρ, the more accurate information the signal σ conveys about c.

The level of output x and monetary transfers, w and t, are all verifiable, so the principal can base compensations w and t on x and enforce them. The signal, σ, is not hard evidence of c nor verifiable. The supervisor and the agent can observe their own private information but not the other's.

Since the supervisor's signal is payoff irrelevant, the following result is immediate.

The First-Best Allocations: Let $(\underline{x}^{fb}, \bar{x}^{fb})$ be the first-best efficient allocation, where $\underline{x}^{fb}$ ($\bar{x}^{fb}$) is the output level when $c = c_1$ (c_2). This allocation is independent of the supervisor's signal, and given by

$$\begin{aligned} R'(\underline{x}^{fb}) &= c_1, \\ R'(\bar{x}^{fb}) &= c_2. \end{aligned}$$

The three tier structure describes organizations in which the principal does not have the time, knowledge, or resources required to monitor the agent. It also describes situations in which verifiable signals (such as stock prices and accounting records) are not effective measures to discipline the agent, either because they fail to reflect relevant aspects of the agent's private information or they are influenced by other external factors. It is a general observation that shareholders of a firm do not act as a board of auditors which monitors the financial discipline of the firm. Auditing roles are commonly taken by a third party which works for the principal.

2.2 Organization and Communication

We describe the centralized organization, in which the principal has a decree over both employees, i.e., the supervisor and the agent. The principal offers a grand mechanism to the employees. When the principal designs this grand mechanism, she takes into account the possibility that the employees collude. We will later elaborate on the details about the grand mechanism and collusion. The timing of the entire game is as follows.

1. The agent observes his cost c.

2. The principal publicly offers a grand mechanism to the supervisor and agent.

3. The agent and the supervisor simultaneously decide whether to accept or reject the mechanism. If either of them rejects, the game ends with no further actions and no monetary transfers. The reservation values of all three parties are normalized to be zero. If both employees accept, the game continues to the next stage.

4. The supervisor observes a signal σ and decides whether to stay in the relationship or quit. If she quits, the game ends.[2] Otherwise, the game continues to the next stage.

5. Collusion takes place. If the employees fail to form a collusion, they will play the grand mechanism noncooperatively and the game will end. If they successfully form a collusion, the game continues to the next stage.

6. The grand mechanism is played: the supervisor and agent report messages, and monetary transfers are made by the principal. Finally, side payments, if any, are made according to the agreement in the collusion contract.

The basic structure of this model is that of a standard screening model, with the addition of a supervisor who may form a collusion with the agent.[3]

2.3 Grand Mechanism

The grand mechanism proposed by the principal is of the form

$$\Gamma = (M_A, M_S, \{(x(m), t(m), w(m))\}_{m \in M}),$$

where M_i is the finite set of messages that $i = A, S$ send to the principal simultaneously, and $x(\cdot), t(\cdot), w(\cdot)$ are decision rules when the messages are $m \in M \equiv M_A \times M_S$.[4] Since all three parties are risk-neutral in terms of money, there is no loss of generality in only considering deterministic monetary transfer rules. For x, the principal strictly prefers a deterministic rule, since $R''(\cdot) < 0$ and randomness in choosing x does not affect incentive compatibility constraints.

[2] The principal may continue interactions only with the agent if the supervisor has rejected the offer. We will discuss this possibility in detail in the chapter 3.

[3] In some real life situations, the principal may be able to offer a binding contract before the supervisor learns the signal. This possibility will certainly increase the set of implementable outcomes for the principal, since the participation constraint need be satisfied only ex-ante.

[4] You may consider the decision rules as being induced by designing a more specific transfer rule, $(T(x), W(x))$. The employees, given their private information, optimally choose x to maximize their utilities, and their decisions induce the decision rule.

2.4 Non-Collusion Benchmark Cases

In this subsection, we characterize the equilibrium outcomes of three benchmark cases. In all benchmarks, application of the revelation principle allows us to focus only on incentive compatible direct mechanisms in which the employees report truthfully. Let (x_{ij}, w_{ij}, t_{ij}) be the allocation rules that correspond to messages (σ_i, c_j).

2.4.1 Direct Supervision with Public Signals

Suppose that σ is verifiable evidence about c, the agent's cost,[5] and both the principal and the agent can obtain σ before contract offer is made.

Given the realization of signal σ_i, the principal proposes a direct mechanism $\{(x_{ij}^d, t_{ij}^d)\}_{j=1,2}$. We solve the principal's problem in the standard way. We first solve for a relaxed problem with only incentive compatibility constraints for the efficient type and participation constraints for the inefficient type. The solution will satisfy the monotonicity condition, $x_{i2} \leq x_{i1}$, which is sufficient for the solution to satisfy the remaining constraints. Let $\Pr(c_j|\sigma_i) = \frac{p_{ij}}{p_{i1}+p_{i2}}$ be the probability of $c = c_j$ conditional on $\sigma = \sigma_i$ (similarly, $\Pr(\sigma_i|c_j) = \frac{p_{ij}}{p_{1j}+p_{2j}}$ is the probability of $\sigma = \sigma_i$ conditional on $c = c_j$). The principal solves:

$$(P_{Di}) \qquad \max_{\{(x_{ij},t_{ij})\}_{j=1,2}} \sum_{j=1,2} \Pr(c_j|\sigma_i)(R(x_{ij}) - t_{ij})$$

subject to

$$t_{i1} - c_1 x_{i1} \geq t_{i2} - c_1 x_{i2}, \tag{1}$$

and

$$t_{i2} - c_2 x_{i2} \geq 0. \tag{2}$$

Solving this problem for $i = 1, 2$, we have:

$$x_{21}^d = x_{11}^d = \underline{x}^{fb}$$

and $x_{12}^d < x_{22}^d < \bar{x}^{fb}$ given by

$$\begin{aligned} R'(x_{22}^d) &= c_2 + \frac{p_{21}}{p_{22}}\Delta c, \\ R'(x_{12}^d) &= c_2 + \frac{p_{11}}{p_{12}}\Delta c. \end{aligned} \tag{3}$$

These equalities characterize optimal output levels, as the monotonicity conditions are satisfied.

[5] Assumption of verifiability of σ is for notational ease only, and is not necessary, since mechanism design under symmetric information makes it possible to costlessly elicit σ. See Maskin (1999) for the detail.

The agent's rents are

$$t_{i1}^d - c_1 x_{i1}^d = \Delta c x_{i2}^d,$$
$$t_{i2}^d - c_2 x_{i2}^d = 0,$$

for $i = 1, 2$.

The output level is efficient when $c = c_1$. The low cost type gets strictly positive information rent. Since $\frac{p_{11}}{p_{12}} > \frac{p_{21}}{p_{22}}$, the incentive cost of inducing the truth telling from the type c_1 agent is higher when $\sigma = \sigma_1$ than when $\sigma = \sigma_2$. Hence it is optimal for the principal to set $x_{12}^d < x_{22}^d$.

When the correlation ρ is high, $x_{12}^d = 0$. The principal commits to zero production when she obtains relatively accurate information that the agent is of type c_1 but the agent turns out to be of type c_2. When the principal observes σ_1, she knows that the agent is likely to be of type c_1. It is then costly to increase x_{12} and attract type c_1's misreport, while giving up x_{12} occurs with small probability and the expected loss from such an occurrence is low. The first-best efficiency is approximated in nearly perfectly correlated states.

2.4.2 No Supervisory Information

Suppose that there is no supervisory information available. Then the principal's problem is a standard screening model. The optimal outcome can be characterized by an incentive compatible direct mechanism $\{(x_i^n, t_i^n)\}_{i=1,2}$, where (x_i, t_i) is output and transfer when the agent's type is c_i. At the optimal solution, the downward incentive compatibility constraint and the participation constraint for the inefficient type are binding. The optimal solution entails $x_1^n = \underline{x}^{fb}$ and

$$R'(x_2^n) = c_2 + \frac{\Pr(c_1)}{\Pr(c_2)} \Delta c,$$
$$t_1^n - c_1 x_1^n = \Delta c x_2^n,$$
$$t_2^n - c_2 x_2^n = 0.$$

In contrast to the output levels in the direct supervision case, we have

$$x_{12}^d < x_2^n < x_{22}^d.$$

2.4.3 Supervision with Private Signal (No Collusion)

Now consider the case in which the supervisor and agent noncooperatively report to the grand mechanism. It is known from the literature of Bayesian implementation with correlated types that the principal can design a mechanism that extracts all

information rent from risk-neutral agents when they are not protected by limited liability (Cremer and McLean (1988)). The principal can implement outcomes as if she possesses perfect information. The principal can attain unconstrained optimal outcome and therefore specifies the first-best outputs levels:

$$\begin{aligned} x_{21}^{nc} &= x_{11}^{nc} = \underline{x}^{fb}, \\ x_{22}^{nc} &= x_{12}^{nc} = \bar{x}^{fb}. \end{aligned}$$

The principal can elicit the agent's type with no incentive cost by offering a transfer scheme that consists of two parts, $t_{ij}^{nc} = t_j^n + t(\sigma_i)$.[6] The first part, t_j^n, is the transfer in the no-supervisory information case specified above. This portion is independent of the supervisor's signal, and therefore the direct mechanism (x^{fb}, t^n) induces truth telling as a dominant strategy. It leaves the agent zero information rent when $c = c_2$, but when $c = c_1$ the agent gets $\Delta c \bar{x}^{fb}$. The principal can extract this surplus without violating the incentive constraints by adding to this mechanism a lottery $t(\sigma_i)$ that depends only on the supervisor's signal. This is given by

$$\begin{pmatrix} \Pr(\sigma_1|c_1) & \Pr(\sigma_2|c_1) \\ \Pr(\sigma_1|c_2) & \Pr(\sigma_2|c_2) \end{pmatrix} \begin{pmatrix} t(\sigma_1) \\ t(\sigma_2) \end{pmatrix} = \begin{pmatrix} -\Delta c \bar{x}^{fb} \\ 0 \end{pmatrix}.$$

Equivalently,

$$\begin{aligned} t(\sigma_1) &= -\frac{1}{\rho} \Pr(c_1) p_{22} \Delta c \bar{x}^{fb} < 0, \\ t(\sigma_2) &= \frac{1}{\rho} \Pr(c_2) p_{12} \Delta c \underline{x}^{fb} > 0. \end{aligned}$$

The direct mechanism (x^{fb}, t^{nc}) is ex-post efficient, incentive compatible, and extracts full information rent.[7]

The key idea behind this result is that an agent's private signal is likely to provide some information about supervisor's private information. As a result, the agent's beliefs and hence the expected return from the transfer are type dependent. The principal can use the supervisor's truthful report as the basis of the transfer to the agent.[8] She sets this transfer arbitrarily large enough to outweigh the incentive cost to implement arbitrary levels of x.

The normative implication of this result is quite striking: agents' private information is irrelevant for the optimal mechanism design unless it is completely independent. However, the level of monetary transfer required to achieve the full extraction

[6] For the supervisor, the principal sets $w_{ij} = 0$ for all i, j.

[7] There are other transfer schemes that attain ex-post efficiency and full rent extraction. The incentive constraints need not be binding.

[8] As we will see later, the fact that truth telling must be self-enforcing has a substantial effect on the outcome when the supervisor is in the position of the (informed) principal.

increases as the supervisor's signal becomes less correlated with the agent's type. In a nearly independent environment, the transfer is very high.[9]

The mechanism (x^{fb}, t^{nc}), derived above, may induce the supervisor and agent to form a collusion. Since $t(\sigma_1) < t(\sigma_2)$ and the supervisor's wages are constant across all state realizations, the agent may bribe the supervisor for always reporting σ_2.

3 Collusion

In the following analyses, the principal can no longer prevent collusion. We adopt the standard approach in the literature motivated by Tirole (1992) and model collusion as an enforceable side-mechanism among colluding parties. The colluding parties agree on manipulations of messages sent to the principal along with monetary transfers to each other. The principal designs a grand mechanism Γ in accordance with the collusion mechanism.[10] We denote the collusion mechanism as Λ.

3.1 Coalition's Objective

Before outlining the collusion mechanism proposed by the supervisor, we first assume that the collusion mechanism is proposed by a benevolent outsider who does not know the true values of types or signals. The outsider's objective function places arbitrary weights on the supervisor and agent's payoffs: the outsider places a welfare weight $\alpha \in [0, 1]$ on the supervisor's utility and $1 - \alpha$ on the agent's. The revelation principle applies to this sub-contracting stage. The outsider proposes the following direct side-mechanism:

$$\Lambda = (\Sigma, C, \{\phi(\sigma, c), y_S(\sigma, c), y_A(\sigma, c)\}_{(\sigma,c)\in\Sigma\times C}).$$

This side-mechanism consists of the message spaces Σ and C, which are the players' type spaces, and decision rules $\phi(\cdot, \cdot)$, $y_i(\cdot, \cdot)$. The manipulation rule $\phi(\cdot, \cdot)$ maps reports from the supervisor and agent into the set ΔM of possibly random messages sent to the grand mechanism. The monetary transfer rule, $y_i(\cdot, \cdot)$ is a mapping into $\Re$ for each $i = A, S$, and must satisfy the budget balance, $y_S(\cdot) + y_A(\cdot) = 0$.

The proposal of the side-mechanism following the grand mechanism offer induces a two-stage game $G(\Gamma, \Lambda)$. In the first *ratification* stage, the supervisor and the agent

[9] Several articles question the premises of the full extraction result. Robert (1991) shows that with limited liability or with risk averse agents, the full extraction result fails to hold for low correlation of types. More recently, Neeman (2004) allows the possibility that agents with the same beliefs may have different preferences. The agents in effect have multi-dimensional private information (beliefs and types). It is then generically impossible to extract full surplus from all types of the agent who have the same belief.

[10] We use the terms side-mechanism and collusion mechanism interchangeably.

have an opportunity to veto Λ. The second stage is an *implementation* stage in which the employees send messages directly to Γ if Λ has been rejected, or to Λ otherwise. The outsider then tells the employees a recommendation of messages sent to Γ. When the principal makes the grand mechanism proposal, she takes into account the continuation game that consists of the proposal of Λ and $G(\Gamma, \Lambda)$.

We are interested in the set of perfect Bayesian equilibria of the entire game. We ignore the subset of equilibria in which the coalition is never formed because each agent expects the other to refuse to collude. Such strategies are weakly dominated.

Let $q^0 = (q^0_\sigma, q^0_c)$ be the set of the prior beliefs held by the employees at the beginning of $G(\Gamma, \Lambda)$, where q^0_σ (q^0_c) is the prior belief about the signal σ (c) held by the agent (supervisor). Let $(\tilde{q}_\sigma, \tilde{q}_c)$ be beliefs following a deviation by the supervisor (agent), where $\tilde{q}_\sigma$ ($\tilde{q}_c$) is the belief about σ (c) held by the agent (supervisor). These beliefs $(\tilde{q}_\sigma, \tilde{q}_c)$ can be arbitrary. Let $G(\Gamma, \tilde{q}_\sigma, q^0_c)$ be the continuation game induced by the grand mechanism Γ following a veto against Λ by the supervisor. Let $E(\Gamma, \tilde{q}_\sigma, q^0_c)$ be the set of Bayesian Nash equilibria in such game ($G(\Gamma, \tilde{q}_c, q^0_\sigma)$ and $E(\Gamma, \tilde{q}_c, q^0_\sigma)$ are defined similarly.) Let also $U_S(\sigma, e_S)$ ($U_A(c, e_A)$) be the expected payoff of the supervisor (agent) of type σ (c) in $e_S \in E(\Gamma, \tilde{q}_\sigma, q^0_c)$ ($e_A \in E(\Gamma, \tilde{q}_c, q^0_\sigma)$). Note that even though $U_S(\sigma, e_S)$ and $U_A(c, e_A)$ are evaluated with the beliefs q^0_c and q^0_σ, equilibria e_S and e_A are affected by $\tilde{q}_\sigma$ and $\tilde{q}_c$.

We are interested in the class of equilibria in which the beliefs are unchanged after playing Λ. We define $G(\Gamma \cdot \Lambda, q^0)$ as the continuation collusion game following unanimous acceptance of Λ by the supervisor and the agent, and $E(\Gamma \cdot \Lambda, q^0)$ as the set of equilibria of this game.

We apply the following notion from Cramton and Palfrey (1995).

Definition 1 Let $U_S(\sigma)$ ($U_A(c)$) be the type σ (c) supervisor's (agent's) payoff from playing the truthful equilibrium $e^* \in E(\Gamma \cdot \Lambda, q^0)$. A side mechanism Λ is *unanimously ratified* for $(e_S, e_A, \tilde{q}_\sigma, \tilde{q}_c)$ if for all σ, c,

$$\begin{aligned} U_S(\sigma) &\geq U_S(\sigma, e_S), \\ U_A(c) &\geq U_A(c, e_A), \end{aligned}$$

where $e_S \in E(\Gamma, \tilde{q}_\sigma, q^0_c)$ ($e_A \in E(\Gamma, \tilde{q}_c, q^0_\sigma)$) is a noncooperative equilibrium of $G(\Gamma, \tilde{q}_\sigma, q^0_c)$ ($G(\Gamma, \tilde{q}_c, q^0_\sigma)$). .

We first characterize a set of equilibrium outcomes of the continuation game that starts from the proposal of a side mechanism. Given the grand mechanism Γ, an equilibrium of the continuation game consists of two elements:

- the set of beliefs $(\tilde{q}_\sigma, \tilde{q}_c)$, and associated noncooperative equilibria for the grand mechanism when the collusion fails to be formed, i.e., $e_S \in E(\Gamma, \tilde{q}_\sigma, q^0_c)$ and $e_A \in E(\Gamma, \tilde{q}_c, q^0_\sigma)$;

- collusive side mechanism Λ that maximizes the outsider's objective function and is unanimously ratified for $(e_S, e_A, \tilde{q}_\sigma, \tilde{q}_c)$.

Let $\phi_{ij} = \phi(\sigma_i, c_j)$ be the (possibly random) recommendation of messages when the supervisor and the agent report σ_i and c_j to the side-mechanism. We abuse notation and denote by $(x(\phi_{ij}), w(\phi_{ij}), t(\phi_{ij}))$ the expected values of outcomes derived over the distribution of ϕ_{ij}. When the side mechanism Λ is unanimously ratified along the equilibrium path, the beliefs of the players will not be updated after the acceptance decisions are made.

An equilibrium of the continuation game thus is $(\tilde{q}_\sigma, \tilde{q}_c)$, $e_S \in E(\Gamma, \tilde{q}_\sigma, q_c^0)$ and $e_A \in E(\Gamma, \tilde{q}_c, q_\sigma^0)$, and a side mechanism that is given by the solution to:

$$(S_\alpha) \qquad \max_{\phi(.),y(.)} \sum_{i,j} p_{ij}[\alpha(w(\phi_{ij}) + y_S(\sigma_i, c_j)) + (1-\alpha)(t(\phi_{ij}) - c_j x(\phi_{ij}) + y_A(\sigma_i, c_j))]$$

subject to budget balance:

$$y_S(\sigma_i, c_j) + y_A(\sigma_i, c_j) = 0$$

for all $(i,j) \in \{1,2\}^2$, the Bayesian incentive compatible constraints for the supervisor and agent:

$$U_S(\sigma_i) \geq \sum_{j=1,2} \Pr(c_j|\sigma_i)(w(\phi_{kj}) + y_S(\sigma_k, c_j)),$$
$$U_A(c_j) \geq \sum_{i=1,2} \Pr(\sigma_i|c_j)(t(\phi_{il}) + y_A(\sigma_i, c_l) - c_j x(\phi_{il}))$$

for all $(i,k) \in \{1,2\}^2$ and $(j,l) \in \{1,2\}^2$, and the participation constraints:

$$U_S(\sigma_i) \geq U_S(\sigma_i, e_S),$$
$$U_A(c_j) \geq U_A(c_j, e_A)$$

for all $(i,j) \in \{1,2\}^2$.

3.1.1 Weakly Collusion-Proof Mechanisms

Instead of tackling the complexity of strategies in the combined game, we want to characterize the set of equilibrium outcomes of the collusion-proof grand mechanism.

Definition 2 A collusion mechanism Λ_0 is *the null side mechanism* if

$$\Lambda_0 = \{\phi = Id, y_S = 0, y_A = 0\},$$

where Id is the identity matrix.

Definition 3 A grand mechanism Γ is *weakly collusion-proof* if and only if 1) it is an incentive compatible direct revelation mechanism; 2) the null side mechanism Λ_0 is unanimously ratified for $(e^*, e^*, q_\sigma^0, q_c^0)$, where e^* is the truthful equilibrium of $G(\Gamma, q_\sigma^0, q_c^0)$ played with passive beliefs; and 3) there does not exist Λ that gives a strictly higher expected joint payoff to the coalition.

Proposition 1 *Suppose that there exists a perfect Bayesian equilibrium of the entire game such that a side-contract Λ is unanimously ratified for some $(e_S, e_A, \tilde{q}_\sigma, \tilde{q}_c)$, and that Λ is the outsider's optimal choice. Then any outcome of such an equilibrium can be replicated by a perfect Bayesian equilibrium in which the principal offers a weakly collusion-proof grand mechanism.*

This proposition says that the outcome of any equilibrium of the overall game in which collusion occurs is also the outcome of an equilibrium in which the principal offers a weakly collusion-proof grand mechanism. A weakly collusion-proof mechanism is an incentive compatible direct grand mechanism, and it is optimal for the outsider to offer the null side mechanism.[11] Acceptance of the null side mechanism is supported by e^* together with the passive beliefs.

As will be confirmed in the next subsection, mechanisms of our interest are such that only the incentive constraints for the types c_1 and σ_1 may be binding. The next proposition characterizes the set of weakly collusion-proof mechanisms.

Proposition 2 *A grand mechanism Γ is weakly collusion-proof only if there exists $\varepsilon_\alpha \in [0,1]$ such that for all i, j,*

$$w_{11} + t_{11} - c_1 x_{11} \geq w_{ij} + t_{ij} - c_1 x_{ij}, \tag{4}$$

$$w_{21} + t_{21} - c_1 x_{21} \geq w_{ij} + t_{ij} - c_1 x_{ij}, \tag{5}$$

$$w_{12} + t_{12} - (c_2 + \varepsilon_\alpha \frac{p_{11}}{p_{12}} \Delta c) x_{12} \geq w_{ij} + t_{ij} - (c_2 + \varepsilon_\alpha \frac{p_{11}}{p_{12}} \Delta c) x_{ij}, \tag{6}$$

$$w_{22} + t_{22} - (c_2 + \frac{\varepsilon_\alpha p_{21}}{p_{22} + \frac{\varepsilon_\alpha \rho}{p_{12}}} \Delta c) x_{22} \geq w_{ij} + t_{ij} - (c_2 + \frac{\varepsilon_\alpha p_{21}}{p_{22} + \frac{\varepsilon_\alpha \rho}{p_{12}}} \Delta c) x_{ij}. \tag{7}$$

When $\varepsilon_\alpha > 0$, the incentive compatibility constraint for the type c_1 agent in the program (S_α) must be binding.

In this lemma the subcase of ε_α indicates its dependency on α.[12] For different values of α, the necessary condition for a grand mechanism being weakly collusion-proof differs only in ε_α. This term measures inefficiency within the coalition, and is

[11] We use the term "weak" to indicate that there may exist other equilibria of $G(\Gamma, \Lambda_0)$ in which Λ_0 is not unanimously ratified, or non-truth-telling equilibrium plays occur.

[12] By solving the outsider's maximization problem (S_α) by Lagrangean method, we get $\varepsilon_\alpha = \frac{\underline{\delta}_A}{\alpha + \underline{\delta}_S + \underline{\nu}_S}$, where $\underline{\delta}_A$,$\underline{\delta}_S$, and $\underline{\nu}_S$ are the multipliers for the type σ_1 and type c_1's incentive constraints and σ_1 type's participation constraint respectively.

a choice variable of the principal. It satisfies $\varepsilon_\alpha \in [0,1)$ for all $\alpha > 0$, and $\varepsilon_\alpha \in [0,1]$ when $\alpha = 0$.

By applying revealed preference argument repeatedly to the inequalities given in Proposition 2, we obtain the following set of coalition constraints:

$$\begin{aligned}
\pi_{11} &= \pi_{21} (\equiv \pi_1), && (8)\\
\pi_1 &\geq \pi_{22} + \Delta c x_{22}, && (9)\\
\pi_{12} &\geq \pi_{22} - \varepsilon_\alpha \frac{p_{11}}{p_{12}} \Delta c (x_{22} - x_{12}),\\
x_{22} &\geq x_{12}. && (10)
\end{aligned}$$

We solve for the weakly collusion-proof mechanism for a given value of α that maximizes the principal's expected payoff. Ignoring the interim individual incentive compatibility constraints for the types c_2 and σ_2, which we ex-post can confirm to hold,[13] the principal's objective is to solve

$$(P_\alpha) \qquad \max_{\pi,w,x,\varepsilon} \textstyle\sum_{i,j} p_{ij} (R(x_{ij}) - c_j x_{ij} - \pi_{ij})$$

subject to the coalition incentive constraints (8)-(10), the individual Bayesian incentive constraints:

$$\sum_{i=1,2} \Pr(\sigma_i | c_1)(\pi_{i1} - w_{i1}) \geq \sum_{i=1,2} \Pr(\sigma_i | c_1)(\pi_{i2} - w_{i2} + \Delta c x_{i2}), \tag{11}$$

$$\sum_{j=1,2} \Pr(c_j | \sigma_1) w_{1j} \geq \sum_{j=1,2} \Pr(c_j | \sigma_1) w_{2j}, \tag{12}$$

and the interim participation constraints:

$$\begin{aligned}
&\sum_{i=1,2} \Pr(\sigma_i | c_1)(\pi_{i1} - w_{i1}) \geq 0, && (13)\\
&\sum_{i=1,2} \Pr(\sigma_i | c_2)(\pi_{i2} - w_{i2}) \geq 0, && (14)\\
&\sum_{j=1,2} \Pr(c_j | \sigma_1) w_{1j} \geq 0,\\
&\sum_{j=1,2} \Pr(c_j | \sigma_2) w_{2j} \geq 0.
\end{aligned}$$

Proposition 3 *The optimal weakly collusion-proof mechanism is independent of the welfare weight α.*

[13] Because of the correlation only σ_1 type supervisor's incentive constraint may be binding.

The principal's optimal choice of mechanism is independent of the welfare weight.[14][15] To understand this result, suppose that collusion takes place under symmetric information and that all the weight is placed on the supervisor, *i.e.*, $\alpha = 1$. Given the grand mechanism, the outsider recommends reports so as to maximize the (equally weighted) sum of the agent and supervisor's utilities, and monetary transfers that gives the agent his reservation value and the supervisor receives the rest of the surplus. Absent income effects in the coalition members' utilities, this is optimal for the outsider, since α only affects the outsider's choice on how to distribute the rent. When collusion takes place under asymmetric information, the agent and supervisor should receive information rent in order to tell the truth. Although the welfare weight α affects the value of the rent, the principal cannot exploit any additional distortion.

Now we consider collusion formation initiated by the supervisor.

3.2 No Participation Decisions after Collusion

First, we modify the timing of the game as follows.

1. The agent privately observes his cost c.

2. The principal publicly offers a grand mechanism to the supervisor and agent.

3. The agent and the supervisor simultaneously decide whether to accept or reject the mechanism. If either of them rejects, the game ends with no further actions and no monetary transfers. Then all three parties receive zero. If both employees accept, the game continues to the next stage.

4. The supervisor privately observes σ and decides whether to stay in or quit. If she quits, the game ends. Otherwise, the game moves on to the next stage.

5'. The supervisor offers a side mechanism If the agent rejects, they will play the grand mechanism noncooperatively and the game ends. If he accepts, the game continues to the next stage.

6'. The agent and supervisor simultaneously report to each other messages chosen from the message spaces specified in the side mechanism.

[14] This implies that the principal does not need to know the weight in order to solve her problem.

[15] We can show by solving the principal's maximization problem that the principal can always attain the same output level and payoff as if the principal can directly observe the supervisor's signal. This outcome can be also attained when the principal cannot observe the supervisor's signal but the signal is observed by the agent.

7'. The grand mechanism is played, and side payments, if any, are made according to the rule specified in the side mechanism.

At the collusion stage, the supervisor's objective is to maximize her expected payoff by proposing a side mechanism that may depend on her type σ_i :

$$\Lambda_{\sigma_i} = (M_S, M_A, \{(\phi(m), y(m))\}_{m \in M_S \times M_A})_{\sigma_i},$$

where $\phi(\cdot)$ is the rule of announcement made to the grand mechanism and $y(\cdot)$ is the monetary transfer rule from the supervisor to the agent.[16]

We are interested in perfect Bayesian equilibrium in which the supervisor's side mechanism offers are accepted by both types of the agents. Before we start formal analysis, we need to look at the flow of information exchanged in the collusion stage.

Communication Technology Absent an outsider, direct communication between the supervisor and agent would likely to affect their beliefs about each other's private information. Firstly, the supervisor's proposal of side mechanism may reveal her private signal. As we will see later, both types of the supervisor propose the same contract, and no information leakage occurs through the proposal. Secondly, the supervisor and agent send messages to each other rather than to an outsider before they play the grand mechanism. This would provide another source of information to learn about the other party's type.

If their beliefs are updated, it might not be optimal for them to report to the grand mechanism according to the agreement. Likewise, with updated beliefs, some types of the employees might find it optimal not to participate in the grand mechanism. Throughout this paper, we assume that the side mechanism is enforceable so that the coalition members have to report as promised.

With regard to the participation decisions after the collusion stage, we do the following analyses. In this subsection we consider the case in which the both coalition members can commit not to leave after the collusion stage. In the next section we focus on the cases in which the supervisor is free to leave if she finds it optimal to do so, and examine how this possibility affects the optimal design of the grand mechanism. We only consider the supervisor's decision because we are interested in constructing a simple grand mechanism such that the supervisor exits on the equilibrium path, and the principal can attain the direct supervision benchmark outcome.

Suppose that the coalition members can commit not to leave after the collusion stage. Then no action by the supervisor and agent depends on the information they

[16] As in Chapter 2, we abuse notations by using the same representations of these functions as well as message spaces, which might be different for each σ_i.

might have obtained during the collusion stage. Let $G(\Gamma, \Lambda_{\sigma_i})$ be the game induced by the proposal of the side-mechanism following the grand mechanism offer. The revelation principle applies to $G(\Gamma, \Lambda_{\sigma_i})$.

We first solve for the supervisor's optimal side mechanism choice given a grand mechanism Γ. Let $G(\Gamma)$ be the continuation game that consists of the proposal of Λ_{σ_i} and $G(\Gamma, \Lambda_{\sigma_i})$. An informed principal problem exists when we analyze the proposal of Λ_{σ_i} by the supervisor; the agent's belief about the supervisor's type may be updated by observing the proposal.

Let $\tilde{q}_c$ be the belief about c held by the supervisor when the agent rejects the supervisor's offer. Let $E(\Gamma, \tilde{q}_c, q_\sigma^\Lambda)$ be the set of associated noncooperative equilibria for the grand mechanism, where q_σ^Λ is the belief about σ held by the agent at the beginning of $G(\Gamma, \Lambda_{\sigma_i})$. Note that q_σ^Λ may be updated from the prior after observing the side mechanism offer. Let e_A be an element of $E(\Gamma, \tilde{q}_c, q_\sigma^\Lambda)$, and $U_A(c_i, e_A)$ be the agent's expected payoff in e_A calculated with q_σ^Λ.

Similarly, define $G(\Gamma \cdot \Lambda_{\sigma_i}, q_c^0, q_\sigma^\Lambda)$ as the continuation game following the acceptance of Λ_{σ_i} by the agent. We say that a side mechanism Λ_{σ_i} is *unanimously ratified* for $(e_A, \tilde{q}_c)$ if for all c_j,

$$U_A(c_j) \geq U_A(c_j, e_A),$$

where $U_A(c_j)$ is the type c_j agent's payoff in the truth telling equilibrium e^* of $G(\Gamma \cdot \Lambda_{\sigma_i}, q_c^0, q_\sigma^\Lambda)$.

Let $\phi_{ij} = \phi(\sigma_i, c_j)$ be the recommendation of messages when the supervisor and the agent report σ_i and c_j to the side-mechanism. For $\lambda \in [0, 1]$, and for given Γ, consider the following program:

$$(S_{IP}) \qquad \max_{\phi(\cdot), y(\cdot)} \sum_j [\lambda p_{1j}(w(\phi_{1j}) - y_{1j}) + (1 - \lambda) p_{2j}(w(\phi_{2j}) - y_{2j})]$$

subject to interim incentive and participation constraints for the agent evaluated with belief q_σ :

$$\sum_{i=1,2} q(\sigma_i|c_j)(t(\phi_{ij}) + y_{ij} - c_j x(\phi_{ij})) \geq \sum_{i=1,2} q(\sigma_i|c_j)(t(\phi_{il}) + y_{il} - c_j x(\phi_{il})),$$

$$\sum_{i=1,2} q(\sigma_i|c_j)(t(\phi_{ij}) + y_{ij} - c_j x(\phi_{ij})) \geq U_A(c_j, e_A)$$

for all $(j, l) \in \{1, 2\}^2$, and for some belief $\tilde{q}_c$ and some $e_A \in E(\Gamma, \tilde{q}_c, q_\sigma)$ for $j = 1, 2$, where $U_A(c_j, e_A)$ is the type c_j agent's expected payoff in a noncooperative equilibrium e_A of the continuation game following his rejection of the supervisor's proposal supported with $\tilde{q}_c$, and the interim incentive and participation constraints for the

supervisor evaluated with her prior belief q_c^0:

$$\begin{aligned}\sum_{j=1,2} \Pr(c_j|\sigma_i)(w(\phi_{ij}) - y_{ij}) &\geq \sum_{j=1,2} \Pr(c_j|\sigma_i)(w(\phi_{kj}) - y_{kj}) \\ \sum_{j=1,2} \Pr(c_j|\sigma_i)(w(\phi_{ij}) - y_{ij}) &\geq U_S(\sigma_i, e^0)\end{aligned}$$

for all $(i,k) \in \{1,2\}^2$, and for all $i = 1,2$, where $U_S(\sigma_i, e^0)$ is the type σ_i supervisor's expected payoff in the equilibrium e^0 of the game without collusion.

Following the literature of mechanism design with collusion, we impose the passive belief assumption: $\tilde{q}_c = q_c^0$.

Definition 4 An allocation (x, w, t, y) *Pareto dominates* (x', w', t', y') if (x, w, t, y) gives both types of the supervisor at least as high payoff as (x', w', t', y'), and gives at least one type strictly higher payoff. An allocation (x, w, t, y) *strictly Pareto dominates* (x', w', t', y') if it gives both types strictly higher payoffs.

Definition 5 An allocation (x, w, t, y) is *Pareto optimal* if it arises from the solution of (S_{IP}) for some weight λ and beliefs (q_c^0, q_σ).

This definition of Pareto optimality is analogous to that of Maskin and Tirole (1990). It differs from their definition in that our definition requires an allocation to satisfy the supervisor's incentive constraints.

Definition 6 An allocation (x, w, t, y) is *strongly Pareto optimal for belief* q_σ if 1) there exists λ such that it is Pareto optimal for (q_c^0, q_σ), and 2) there is no belief q'_σ and corresponding Pareto optimal allocation $(x', w', t'.y')$ that Pareto dominates (x, w, t, y) if q'_σ is strongly positive and strictly Pareto dominates (x, w, t, y) if q'_σ is not strongly positive.

When q'_σ does not have full support, there are a continuum of Pareto optimal allocations for the belief, each of which has different payoff for the type that q'_σ places probability zero. For this reason, we require strictly Pareto domination when q'_σ places zero on one type so that strongly Pareto optimality is well defined. Let $Y^*(q_\sigma)$ be the set of strongly Pareto optimal allocations for belief q_σ.

Proposition 4 *Suppose that no participation decision is made after the collusion stage. Suppose also that $Y^*(q_\sigma) \neq \emptyset$ for all q_σ. Then the set of equilibrium outcomes of the continuation game $G(\Gamma)$, which consists of the proposal of Λ_{σ_i} and $G(\Gamma, \Lambda_{\sigma_i})$, coincides with the set of strongly Pareto optimal allocation for the prior belief q_σ^0, $Y^*(q_\sigma^0)$.*

Moreover, $\Lambda_{\sigma_1} = \Lambda_{\sigma_2}$ in every equilibrium of the continuation game $G(\Gamma)$.

This proposition is an extension of Proposition 2 in Maruyama (2005) to a game in which the agent's reservation values are determined by playing another mechanism. In Maruyama, the reservation values are exogenously fixed and independent of the agent's type. In the game we are facing now, the information revealed by the supervisor's side mechanism offer not only affects the on-equilibrium path plays but also the agent's reservation values, as it affects how the grand mechanism would be played noncooperatively should the agent reject the offer.

Collusion-Proofness Principle Weak collusion-proofness principle applies. To see this, consider a perfect Bayesian equilibrium of the entire game in which the principal's optimal choice of grand mechanism is Γ^*, all types of the employees accept the offer, and the supervisor's optimal choice of side mechanism is Λ^*, which is without loss of generality incentive compatible direct mechanism. This side mechanism gives the agent equilibrium expected payoff $\bar{U}_A(c)$ that is at least as high as his reservation value $\bar{U}_A(c, \bar{e}_A)$, where $\bar{e}_A$ is some continuation equilibrium should the agent reject the offer. Then there exists another equilibrium of the entire game sustained with the passive beliefs, in which the principal proposes a direct mechanism $\tilde{\Gamma} = \Gamma^* \cdot \Lambda^*$, the optimal choice for the supervisor is to propose the null contract, and the agent's equilibrium payoff is $\bar{U}_A(c)$. Suppose that the null contract is not the optimal response to $\tilde{\Gamma}$, then there exists a side contract $\tilde{\Lambda}$ that gives strictly higher payoff to the supervisor and at least $\bar{U}_A(c) \geq \bar{U}_A(c, \bar{e}_A)$ to the agent, contradicting the fact that Λ^* rather than $\Lambda^* \cdot \tilde{\Lambda}$ is the supervisor's optimal response to Γ^*.

We need to analyze the class of equilibria in which a side mechanism is accepted by the both types of the agent on the equilibrium path. Let $\phi_{ij} = \phi(\sigma_i, c_j)$ be the recommendation of messages when the supervisor and the agent report σ_i and c_j to the side-mechanism.

An equilibrium of the continuation game after a grand mechanism Γ^* is accepted by the both employees consists of:

- the belief $\tilde{q}_c$, and an associated noncooperative equilibria for the grand mechanism $e_A \in E(\Gamma, \tilde{q}_c, q_\sigma^0)$;
- the set of collusive side mechanisms $(\Lambda^*_{\sigma_1}, \Lambda^*_{\sigma_2})$, where $\Lambda^*_{\sigma_i}$ maximizes the type σ_i supervisor's expected payoff, and is accepted by both types of the agent;
- the belief about the supervisor's signal $\tilde{q}_\sigma^\Lambda$ held by the agent when the supervisor proposes a side mechanism Λ other than Λ^*

For all $\lambda \in [0, 1]$, and for given Γ, consider the following program:

$$(S_\lambda) \qquad \max_{\phi(\cdot), y(\cdot)} \sum_j [\lambda p_{1j}(w(\phi_{1j}) - y_{1j}) + (1 - \lambda) p_{2j}(w(\phi_{2j}) - y_{2j})]$$

subject to the Bayesian incentive compatible constraints for the supervisor and agent:

$$\sum_{j=1,2} \Pr(c_j|\sigma_i)(w(\phi_{ij}) - y_{ij}) \geq \sum_{j=1,2} \Pr(c_j|\sigma_i)(w(\phi_{kj}) - y_{kj}),$$

$$\sum_{i=1,2} \Pr(\sigma_i|c_j)(t(\phi_{ij}) + y_{ij} - c_j x(\phi_{ij})) \geq \sum_{i=1,2} \Pr(\sigma_i|c_j)(t(\phi_{il}) + y_{il} - c_j x(\phi_{il}))$$

for all $(i,k) \in \{1,2\}^2$ and $(j,l) \in \{1,2\}^2$, and the interim participation constraints:

$$\begin{aligned} \sum_{i=1,2} \Pr(\sigma_i|c_j)(t(\phi_{ij}) + y_{ij} - c_j x(\phi_{ij})) &\geq U_A(c_j, e_A) \\ \sum_{j=1,2} \Pr(c_j|\sigma_i)(w(\phi_{ij}) - y_{ij}) &\geq U_S(\sigma_i, e^0) \end{aligned}$$

for some $e_A \in E(\Gamma, q_c^0, q_\sigma^0)$ for $j = 1, 2$, and for $i = 1, 2$, where $U_A(c_j, e_A)$ is the type c_j agent's expected payoff in a noncooperative equilibrium e_A of the continuation game following his rejection of the supervisor's proposal supported with q_c^0.

From the Proposition 4, the supervisor's optimal mechanism proposal is given by solving (S_λ) for some $\lambda \in [0,1]$. This problem can be regarded as the outsider's problem (S_α) with $\alpha = 1$ and an arbitrary weight λ $(1-\lambda)$ on the utility of the type σ_1 (σ_2) supervisor. Since the supervisor does not reveal her signal through the offer of side mechanism, the constraints for the agent are interim.

Proposition 5 *Suppose that conditions in Proposition 4 are satisfied. Suppose that the supervisor proposes a side mechanism. The optimal collusion-proof mechanism entails the same allocations as when an outsider proposes a side mechanism.*

Whether a benevolent uninformed outsider or the supervisor with private signal offers a side contract does not affect the optimal collusion-proof mechanism. The potential signaling problem when the supervisor proposes a side mechanism affects the efficiency within the coalition just as the incentive constraints for the supervisor does in the outsider's problem. Since the supervisor can use monetary transfers to adjust the share of surplus that accrues to the coalition, it is of her interest to suggest reports that maximize the total virtual payoff for the coalition.[17] Even though the supervisor knows her own signal when she proposes a side mechanism, the optimal collusion-proof mechanism is as if the side mechanism is proposed by an uninformed outsider.

[17] Risk neutrality is crucial assumption for this result.

4 Supervisor's Participation Decisions

In this section, we consider the situations in which the supervisor is free to leave the grand mechanism any time before the employees send messages to it. In particular, she can leave after collusion agreement is made. When the supervisor's ex-post payoff is strictly negative in some state realizations, since the grand mechanism allows the supervisor to leave after she learns the signal, she is willing to leave and guarantee non-negative ex-post payoff unless she can credibly commit in the side mechanism to stay in the grand mechanism.

The timing of the entire game is as follows.

1. The agent privately observes c.
2. The principal publicly offers a grand mechanism to the supervisor and agent.
3. The agent and the supervisor simultaneously decide whether to accept or reject the mechanism. If either of them rejects, the game ends with no further actions and no monetary transfers. Then all three parties receive zero. If both employees accept, the game continues to the next stage.
4. Signal σ is realized. The supervisor decides whether to stay in or quit. If she quits, the game ends. Otherwise, the game moves on to the next stage.

5'. The supervisor offers a side mechanism. If the agent rejects, they will play the grand mechanism noncooperatively and the game ends. If he accepts, the game continues to the next stage.

6'. The agent and supervisor simultaneously make announcements to each other.

6.5'. The supervisor decides whether to stay in or quit. If she quits, the game ends. Otherwise, the game continues.

7'. The grand mechanism is played. Finally, side payments, if any, are made according to the rule specified in the side mechanism.

First we consider the set of grand mechanisms that satisfy all of the supervisor's ex-post participation constraints. The weak collusion-proofness principle applies to these cases, and so we analyze the class of direct mechanisms to which the supervisor's optimal response is to propose the null mechanism.

The principal proposes a direct mechanism:

$$\Gamma = (\Sigma, C, \{x(\sigma, c), t(\sigma, c), w(\sigma, c)\}_{(\sigma,c)\in\Sigma\times C}).$$

Given the grand mechanism, the supervisor proposes a side mechanism that solves (S_λ). We then obtains the set of coalition constraints, (8), (9), (10), and

$$\pi_{12} \geq \pi_{22} - \varepsilon_\lambda \frac{p_{11}}{p_{12}} \Delta c (x_{22} - x_{12}). \tag{15}$$

The principal's problem is to solve:

$$(P_I) \qquad \max_{\pi,w,x} \sum_{i,j} p_{ij}(R(x_{ij}) - c_j x_{ij} - \pi_{ij})$$

subject to the coalition incentive constraints; (8), (9), (10), and (15), interim incentive compatibility constraints for the supervisor and the agent; (11) and (12), interim participation constraints for the agent; (13) and (14), and ex-post participation constraints for the supervisor:

$$w_{ij} \geq 0$$

for $(i, j) \in \{1, 2\}^2$.

Lemma 1 *The solution to the program (P_I) involves non-supervisory information benchmark output level x^n. All four ex-post participation constraints for the supervisor are binding. The principal's payoffs are the same as when there is no supervisory signals.*

If the principal wants the supervisor to stay in all state realizations, she has to propose a rule that is independent of supervisor's report. Otherwise, at least one of the coalition constraints would be violated. The possibility of collusion renders supervisory information completely useless. The principal gains from the signal σ by reducing the type c_1 agent's rent depending on σ. Since the coalition constraints require that the coalition receive the same total ex-post rent when $c = c_1$, the supervisor's wages should be adjusted to make up for the reduction of the agent's rent. If the supervisor's wages cannot be negative, however, the only way the principal can use σ to extract rents is to reduce the type c_2 agent's payoff. This is clearly impossible because the participation constraint for the type c_2 agent would be violated. If the principal cannot reduce the rent at all, there is no point of setting $x_{22} > x_{12}$. Then the optimal contract is independent of the supervisor's signal.

4.1 Learning from the Supervisor's Exit Decisions

The question is whether the principal can improve the situation by designing a mechanism that permits the supervisor to leave at one point for the purpose of soliciting information. A difficulty of analyzing such cases is that we can no longer apply the collusion-proofness principle, since the supervisor's exit decision is based on her

updated beliefs. Consider an equilibrium in which the following occurs on the equilibrium path; the principal proposes $\tilde{\Gamma}$, the supervisor proposes $\tilde{\Lambda}$, and the supervisor exits when she receives one message from the agent and stays in when she receives other messages. Consider another grand mechanism $\hat{\Gamma}$ that replicates the equilibrium payoffs from playing this entire game. Even though $\hat{\Gamma}$ may exist, it is unlikely to be collusion-proof.

Instead of characterizing the optimal collusion-proof mechanism, we will examine whether certain outcome would be achieved by some grand mechanisms. For now we assume that the side mechanism is incomplete in the sense that it cannot specify the monetary transfer rule based on the supervisor's exit decision.

Consider the following simple direct grand mechanism Γ_1:

$$\begin{aligned} w_{11} &= w_{12} = 0, \\ w_{21} &= w_{22} = D < -\Delta c(x_{22} - x_{12}), \\ (t, x) &= (t^d, x^d), \end{aligned}$$

where (t^d, x^d) is the direct supervision benchmark outcome that is obtained as the optimal solution for the principal when she directly and publicly obtains the supervisory signal at the beginning of the game. Since $x^d_{22} > x^d_{12}$, $D < 0$. Consider also the following renegotiation contract that the principal proposes to the agent when the supervisor exits:

$$(t^e(c_j), x^e(c_j)) = (t^d_{2j}, x^d_{2j}) \text{ for } j = 1, 2,$$

where $(t^e(\cdot), x^e(\cdot))$ is a direct mechanism.

Proposition 6 *The principal can implement the same output level and receives the same payoffs as in the direct supervision benchmark by proposing* Γ_1 *together with the renegotiation contract* $(t^e(\cdot), x^e(\cdot))$.

Since the supervisory information is fully screened in this equilibrium, the choice of output level is optimal for the principal.[18] Since only the agent has a production technology, there is a room for renegotiation between the principal and the agent after the supervisor rejects the grand mechanism. The principal and agent renegotiate Γ_1 and agree on $(t^e(\cdot), x^e(\cdot))$. Given the principal's updated belief about the agent, it is optimal to propose $(t^e(\cdot), x^e(\cdot))$, which is incentive compatible and yields outcome (t^d_{2j}, x^d_{2j}). From the same argument as above, no profitable collusion arrangement can be made by the supervisor.

[18] This continuation game has other equilibria in which the type σ_2 supervisor stays in and misreports her signal as σ_1 with a positive probability. In those equilibria, the supervisory information is not fully screened and hence the output levels specified in Γ_1 are not optimal for the principal.

4.2 Contracting on Exit Decisions

The mechanism Γ_1 is susceptible to collusion if the coalition members can sign a contract based on the supervisor's exit decision. It is then possible to give the supervisor an incentive to exit by collusion. Consider the following strategies. The supervisor always stays after observing σ, and proposes the following side mechanism Λ_1.[19] This side mechanism specifies the following rules. If the agent tells the supervisor that his types is c_1, the supervisor promises to exit and requests transfer payment $\Delta c(x_{22}^d - x_{12}^d)$. Otherwise, the rules require (σ_1, c_2) to be reported to Γ_1 and no monetary transfer to be made. If the agent rejects the offer, the supervisor stays and reports σ_1 to Γ_1.

This proposal is accepted by both types of the agent.[20] When they play the side mechanism, the agent tells his true type to the supervisor, and the supervisor reports σ arbitrarily. After the collusion, the supervisor leaves if the agent has announced c_1. The agent reports the truth and supervisor (if she is still in the game) always reports σ_1 to the grand mechanism. Whenever the agent's type is c_1, the coalition receives a rent $\Delta c x_{22}^d$, and therefore the (σ_1, c_1) coalition gains by $\Delta c(x_{22}^d - x_{12}^d)$ from this side mechanism. The supervisor's strategies depend only on c. In fact, her exit decision fully reveals the agent's signal to the principal. The principal's output choice in the grand mechanism is clearly not optimal.

This collusion can be blocked, however, if the principal can distinguish the timing at which the supervisor exits. If the supervisor exits at interim stage, the principal proposes a renegotiation contract $(t^e(\cdot), x^e(\cdot))$ to the agent, but if the supervisor exits after the collusion stage, she proposes $x(\cdot) = \underline{x}^{fb}, t(\cdot) = c_1 \underline{x}^{fb}$ for all c_i. If the supervisor proposes the side mechanism Λ_1 and the coalition members follow the strategies described above, the principal infers that $c = c_1$ with probability one if the supervisor has exited after the collusion stage. The renegotiation rule is therefore the optimal reaction by the principal.

Proposition 7 *The mechanism Γ_1 is susceptible to collusion if the employees can collude on the supervisor's exit decision. However, if the principal can distinguish the timing at which the supervisor exits, the direct supervision benchmark output and profit levels can be implemented with Γ_1 and renegotiation.*

Facing Γ_1 and the renegotiation rules, the side mechanism should effectively enforce a clause like this; "If the agent reports certain message to the side mechanism, the supervisor, regardless of her message, should receive a reward and exit with probability q." From the coalition's point of view, its potential gain is from inducing the

[19] The type σ_2 supervisor stays because she can gain by telling the agent that she is of type σ_1 in the collusion stage.

[20] The type c_2 agent may decline the offer. The result will be the same in that case.

supervisor to exit when $c = c_1$, while the supervisor's actions should not reveal too much information before the renegotiation stage is reached. The supervisor's exit at the interim stage tells that her signal is σ_2. If the principal can distinguish the timing of the exit decisions, she can infer good amount of information so that the principal can take advantage of it when the grand mechanism is renegotiated.

5 Conclusion

We analyze collusion under asymmetric information when the principal relies on the private supervisory information in providing incentives to a productive agent and in reducing the information rent. This paper especially focuses on the effect of direct communication among coalition members on the effectiveness of collusion and on the form of the contract the principal proposes. When the coalition members can commit to staying in the grand mechanism and playing it according to the collusion agreement, it is irrelevant who proposes a side mechanism even though the informed principal problem within the coalition may render their communication less efficient.

Through direct communication between the supervisor and agent in the collusion stage, their beliefs about each other's type are updated. The principal may try to use the pieces of information they obtained through collusion and set up a more complex mechanism. The additional information in such a mechanism may not help the principal, however, if the coalition can manipulate those information. One way of circumventing this problem is to utilize information that cannot be stipulated in the collusion mechanism. We analyze the situations in which the side mechanism cannot be enforced when a member decides not to participate in the grand mechanism. When the coalition is moderated by an outsider, the outsider is an enforcer of a side contract as well as a communication device to coordinate coalition members' actions. When a side mechanism is proposed by one of the employees, it is unlikely that a proposer of a contract can credibly commit to taking an action that would punish her/himself without relying on external enforcement entities. The principal may exploit this opportunity by proposing a renegotiation contract based on the implicit information revelation.

As the simple example in the previous section suggests, analyzing collusion as a noncooperative game is very complex. It is a useful future work to see whether modified version of collusion-proofness would apply to mechanisms that are designed to induce the coalition members to exit.

There are a few aspects of collusion our model does not capture. We only focus on hidden information in this paper. When the agent takes a hidden action, there are at least two issues to consider. One is renegotiation. If the principal cannot commit to not renegotiating the original grand contract, the possibility of renegotiation will

impose additional constraints. Another issue is risk sharing, which will further reduce the set of feasible outcomes for the principal due to limitation to offer high-powered incentives. Finally, when there is more than one productive agent, the supervisor also has a role to facilitate coordination among agents with conflicting interests. Allowing the agents to form an efficient coalition might benefit the principal if the supervisor's role as a coordinator is vital to the organization.

Finally, this exercise is a step towards the analysis of decentralized organization, in which the principal delegates to the supervisor the right to contract with the agent.[21] Such delegation brings the supervisor into the position of an informed principal. Whether delegation would function as a device to deter collusion when a privately informed supervisor proposes a collusion mechanism is an important question to be answered in our future research.

Appendix

Proof of Proposition 1: Given a grand mechanism Γ, consider any perfect Bayesian equilibrium such that a side-mechanism Λ^* is unanimously ratified on the equilibrium path for some $(\bar{e}_S, \bar{e}_A, \tilde{q}_\sigma, \tilde{q}_c)$. There is no loss of generality to focus on a truth telling equilibrium of a direct mechanism that maps the report (σ, c) into (ϕ, y). The side-mechanism Λ^* solves the program (S) with the reservation values $U_S(\sigma, \bar{e}_S)$ and $U_A(c, \bar{e}_A)$, and the supervisor and agent receive equilibrium payoffs $\bar{U}_S(\sigma)$ and $\bar{U}_A(c)$.

Now consider a direct grand mechanism $\hat{\Gamma} = \Gamma \cdot \Lambda^*$. We want to show that there exists a perfect Bayesian equilibrium such that the principal proposes $\hat{\Gamma}$, the outsider proposes the null side-mechanism Λ_0, and this choice is supported by the passive beliefs.

Given the grand mechanism $\hat{\Gamma}$, Λ_0 is unanimously ratified for $(e^*, e^*, q^0_\sigma, q^0_c)$, since the players' payoffs from playing the truth telling equilibrium $e \in E(\hat{\Gamma} \cdot \Lambda_0, q^0)$ are $\bar{U}_S(\sigma)$ and $\bar{U}_A(c)$, while $U_S(\sigma, e^*) = \bar{U}_S(\sigma)$ and $U_A(c, e^*) = \bar{U}_A(c)$. It only remains to show that Λ_0 solves the program (S) with the reservation values $\bar{U}_S(\sigma)$ and $\bar{U}_A(c)$. Suppose in contrary that there is a side-mechanism offer $\tilde{\Lambda}$ that solves the program and gives the total payoff strictly higher than with Λ_0. Then, since $\bar{U}_S(\sigma) \geq U_S(\sigma, \bar{e}_S)$ and $\bar{U}_A(c) \geq U_A(c, \bar{e}_A)$, the outsider could have proposed $\Lambda^* \cdot \tilde{\Lambda}$ instead of Λ^* in response to Γ, and strictly increased the total payoff. This contradicts the fact that Λ^* is the optimal side-mechanism. ■

[21] Melumad, Mookherjee, and Reichelstein (1995), Faure-Grimaud, Laffont, and Martimort (2003), Celik (2003), and Mookherjee and Tsumagari address the issue of delegation in comparison to a centralized organization structure.

Proof of Proposition 2: We consider the class of grand mechanisms such that interim incentive compatibility constraints for the types c_1 and σ_1 as well as the participation constraints for the types c_2 and σ_2 may be binding at the optimal solution. Then the outsider's problem (S_α) can be written as below, and we can solve it using the standard Lagrangian multiplier maximization technique. The Lagrange multipliers are indicated in the parentheses that precede the constraints.

$$\max_{\phi(.),y(.)} \sum_{i,j} p_{ij}[\alpha(w(\phi_{ij}) + y_S(\sigma_i, c_j)) + (1-\alpha)(t(\phi_{ij}) - c_j x(\phi_{ij}) + y_A(\sigma_i, c_j))]$$

subject to the budget balance: (λ_{ij})

$$y_S(\sigma_i, c_j) + y_A(\sigma_i, c_j) = 0$$

for $i = 1, 2$ and $j = 1, 2$, the Bayesian incentive compatible constraint for the type σ_1 (c_1) supervisor (agent): $(\underline{\delta}_S, \underline{\delta}_A)$

$$\begin{aligned} &p_{11}(w(\phi_{11}) + y_S(\sigma_1, c_1)) + p_{12}(w(\phi_{12}) + y_S(\sigma_1, c_2)) \\ &\qquad \geq p_{11}(w(\phi_{21}) + y_S(\sigma_2, c_1)) + p_{12}(w(\phi_{22}) + y_S(\sigma_2, c_2)), \\ &p_{11}(t(\phi_{11}) + y_A(\sigma_1, c_1) - c_1 x(\phi_{11})) + p_{21}(t(\phi_{21}) + y_A(\sigma_2, c_1) - c_1 x(\phi_{21})) \\ &\qquad \geq p_{11}(t(\phi_{12}) + y_A(\sigma_1, c_2) - c_1 x(\phi_{12})) + p_{21}(t(\phi_{22}) + y_A(\sigma_2, c_2) - c_1 x(\phi_{22})), \end{aligned}$$

the participation constraints for the types σ_2 and c_2: $(\bar{\nu}_S, \bar{\nu}_A)$

$$\begin{aligned} &p_{21}(w(\phi_{21}) + y_S(\sigma_2, c_1)) + p_{22}(w(\phi_{22}) + y_S(\sigma_2, c_2)) \geq (p_{21} + p_{22}) U_S(\sigma_2, e_S), \\ &p_{12}(t(\phi_{12}) + y_A(\sigma_1, c_2) - c_2 x(\phi_{12})) + p_{22}(t(\phi_{22}) + y_A(\sigma_2, c_2) - c_2 x(\phi_{22})) \geq (p_{12} + p_{22}) U_A(c_2, e_A) \end{aligned}$$

for some $e_S \in E(\Gamma, q_\sigma^0, q_c^0)$ and $e_A \in E(\Gamma, q_c^0, q_\sigma^0)$, and the participation constraints for the types σ_1 and c_1: $(\underline{\nu}_S, \underline{\nu}_A)$

$$\begin{aligned} &p_{11}(w(\phi_{11}) + y_S(\sigma_1, c_1)) + p_{12}(w(\phi_{12}) + y_S(\sigma_1, c_2)) \geq (p_{11} + p_{12}) U_S(\sigma_1, e_S), \\ &p_{11}(t(\phi_{11}) + y_A(\sigma_1, c_1) - c_1 x(\phi_{11})) + p_{21}(t(\phi_{21}) + y_A(\sigma_2, c_1) - c_1 x(\phi_{21})) \geq (p_{11} + p_{21}) U_A(c_1, e_A) \end{aligned}$$

for some $e_S \in E(\Gamma, q_\sigma^0, q_c^0)$ and $e_A \in E(\Gamma, q_c^0, q_\sigma^0)$.

Optimizing with respect to $y_i(\sigma_1, c_1)$ yields

$$\begin{aligned} \lambda_{11} + p_{11}(\alpha + \underline{\delta}_S + \underline{\nu}_S) &= 0, \qquad (16) \\ \lambda_{11} + p_{11}(1 - \alpha + \underline{\delta}_A + \underline{\nu}_A) &= 0. \qquad (17) \end{aligned}$$

Optimizing with respect to $y_i(\sigma_1, c_2)$ yields

$$\lambda_{12} + p_{12}(\alpha + \underline{\delta}_S + \underline{v}_S) = 0, \quad (18)$$

$$\lambda_{12} - p_{11}\underline{\delta}_A + p_{12}(1 - \alpha + \bar{\nu}_A) = 0. \quad (19)$$

Optimizing with respect to $y_i(\sigma_2, c_1)$ yields

$$\lambda_{21} - p_{11}\underline{\delta}_S + p_{21}(\alpha + \bar{\nu}_S) = 0, \quad (20)$$

$$\lambda_{21} + p_{21}(1 - \alpha + \underline{\delta}_A + \underline{v}_A) = 0. \quad (21)$$

Optimizing with respect to $y_i(\sigma_2, c_2)$ yields

$$\lambda_{22} - p_{12}\underline{\delta}_S + p_{22}(\alpha + \bar{\nu}_S) = 0, \quad (22)$$

$$\lambda_{22} - p_{21}\underline{\delta}_A + p_{22}(1 - \alpha + \bar{\nu}_A) = 0. \quad (23)$$

From (16) and (17), $\alpha + \underline{\delta}_S + \underline{\nu}_S = 1 - \alpha + \underline{\delta}_A + \underline{\nu}_A$. This together with optimizing with respect to ϕ_{11} gives:

$$\phi_{11}^* \in \arg\max[w(\phi_{11}) + t(\phi_{11}) - c_1 x(\phi_{11})].$$

From (20) and (21), $1-\alpha+\underline{\delta}_A+\underline{\nu}_A = \alpha+\bar{\nu}_S-\frac{p_{11}}{p_{21}}\underline{\delta}_S$. This together with optimizing with respect to ϕ_{21} gives:

$$\phi_{21}^* \in \arg\max\left[w(\phi_{21}) + t(\phi_{21}) - c_1 x(\phi_{21})\right].$$

From (18) and (19), $1-\alpha+\bar{\nu}_A-\frac{p_{11}}{p_{12}}\underline{\delta}_A = \alpha+\underline{\delta}_S+\underline{\nu}_S$. This together with optimizing with respect to ϕ_{12} gives:

$$\phi_{12}^* \in \arg\max\left[w(\phi_{12}) + t(\phi_{12}) - \left(c_2 + \frac{(p_{11}/p_{12})\underline{\delta}_A}{\alpha + \underline{\delta}_S + \underline{\nu}_S}\Delta c\right) x(\phi_{12})\right]$$

or

$$\phi_{12}^* \in \arg\max\left[w(\phi_{12}) + t(\phi_{12}) - \left(c_2 + \varepsilon_a \frac{p_{11}}{p_{12}}\Delta c\right) x(\phi_{12})\right],$$

where $\varepsilon_\alpha = \frac{\underline{\delta}_A}{\alpha+\underline{\delta}_S+\underline{\nu}_S} \in [0, 1)$.

From (22) and (23), $1 - \alpha + \bar{\nu}_A - \frac{p_{21}}{p_{22}}\underline{\delta}_A = \alpha + \bar{\nu}_S - \frac{p_{12}}{p_{22}}\underline{\delta}_S$. This together with optimizing with respect to ϕ_{22} give:

$$\phi_{22}^* \in \arg\max\left[w(\phi_{22}) + t(\phi_{22}) - \left(c_2 + \frac{\varepsilon_\alpha p_{21}}{p_{22} + \varepsilon_\alpha \frac{\rho}{p_{12}}}\Delta c\right) x(\phi_{22})\right].$$

In a weakly collusion proof mechanism, all of the participation constraints to join the coalition specified above are binding so that the values of the multipliers are not uniquely determined. Therefore the principal has freedom to choose the values of ε. If $\varepsilon > 0$, then the incentive compatibility constraint for type c_1 must be binding in this outsider's problem. In a weakly collusion-proof mechanism, $\phi_{ij}^* = (\sigma_i, c_i)$ with probability one, and $y_S(\sigma_i, c_j) = y_A(\sigma_i, c_j) = 0$ for all i and j. We then obtain the Proposition. ■

Proof of Proposition 3: By solving the program (P_α) using the Lagrangian multiplier maximization method, we see that the principal's optimal choice involves $\varepsilon_\alpha \to 1$ for all α, and therefore the solution does not depend on α. ■

Proof of Proposition 4: ($\Longrightarrow$ sufficiency): A strong Pareto optimal allocation for q^0 is a perfect Bayesian equilibrium outcome.

Let $(x^*, w^*, t^*, y^*) \equiv \{(x^*_{ij}, w^*_{ij}, t^*_{ij}, y^*_{ij})\}_{(i,j)\in\{1,2\}^2}$ be a strongly Pareto optimal allocation for the prior q^0. Consider the following strategies on the equilibrium path: Both types of the supervisor offer the same truth-telling direct mechanism Λ^* with outcome (x^*, w^*, t^*, y^*). Both types of the agent accept the offer. Finally the supervisor and agent reveal their information truthfully. We first prove that these strategies are the best responses on the equilibrium path. We thereafter show that for any finite mechanism other than (x^*, w^*, t^*, y^*),[22] we can find beliefs and an equilibrium of the continuation game after the proposal such that no type of the supervisor is better off than when she has offered (x^*, w^*, t^*, y^*).

Since the supervisor's mechanism offer is pooling, and both types of the agent accept the offer, their beliefs will not be updated before the last stage is reached. The optimality of truth-telling hence is guaranteed by the incentive compatibility constraints, which are evaluated with the prior beliefs. Then the participation constraints for the agent imply that the acceptance decision is optimal.

It only remains to show that the mechanism offer is optimal. Given the principal's offer Γ, for an arbitrary alternative mechanism Λ, let $G(\Gamma \cdot \Lambda, q^0_c, q^\Lambda_\sigma)$ be the continuation game following the proposal of Λ (and before the agent's acceptance decision), where $\tilde{q}_\sigma$ is the agent's updated belief about σ and q^0_c is the supervisor's prior about c. In our setting, perfect Bayesian equilibrium of the continuation game is sequential equilibrium, and therefore there exists an equilibrium of the continuation game.

Suppose that there exists a mechanism Λ such that for any $\tilde{q}_\sigma$ and any equilibrium of $G(\Gamma\cdot\Lambda, q^0_c, q^\Lambda_\sigma)$ there is at least one type of the supervisor that receives strictly higher expected payoff than from (x^*, w^*, t^*, y^*).

Let $V_\Gamma(\tilde{q}^\Lambda_\sigma)$ be the set of the supervisor's equilibrium payoff of $G(\Gamma\cdot\Lambda, q^0_c, \tilde{q}^\Lambda_\sigma)$ when the agent's belief is $\tilde{q}^\Lambda_\sigma$. Since for any $\tilde{q}^\Lambda_\sigma$ the continuation equilibrium is a sequential equilibrium, $V_\Gamma(\tilde{q}^\Lambda_\sigma)$ is upper hemicontinuous. It is also convex-valued as mentioned in the text. Let V_Γ be a convex and compact set that contains $V_\Gamma(\tilde{q}^\Lambda_\sigma)$ for all $\tilde{q}^\Lambda_\sigma$.

Let $v^* = (v^*_1, v^*_2)$ be the vector of the supervisor's expected payoffs associated with (x^*, w^*, t^*, y^*). For $v \in V_\Gamma$ define the correspondence

$$q_i(v) \equiv \arg\max_{p_i} [p_i v_i + (1 - p_i)v^*_i]$$

[22] The mechanism needs not be a direct mechanism.

and

$$q(v) \equiv q_1(v) \times q_2(v).$$

For belief $\tilde{q}_\sigma^\Lambda$ and $v \in V_\Gamma$, consider the correspondence

$$(\tilde{q}_\sigma^\Lambda, v) \to q(v) \times V_\Gamma(\tilde{q}_\sigma^\Lambda).$$

Since either $v_1 > v_1^*$ or $v_2 > v_2^*$, this correspondence is non-empty at any $(\tilde{q}_\sigma^\Lambda, v)$. This correspondence is upper hemicontinuous and convex-valued, and hence it has a fixed point $(\bar{q}_\sigma, \bar{v})$.

Since v^* is strongly Pareto optimal, $v_i^* \geq \bar{v}_i$ for at least one type of the supervisor. Suppose that $\bar{v}_1 > v_1^*$. Then $v_2^* \geq \bar{v}_2$, and hence $\bar{q}_\sigma(\sigma_1|c_j) = 1$, $j = 1, 2$, by the construction of $\bar{q}_\sigma$. With this belief $\bar{v}_1$ cannot be greater than in the direct supervision benchmark. Since v_1^* is at least as large as in the direct supervision benchmark, it must be the case that $v_1^* \geq \bar{v}_1$, a contradiction. The case in which $\bar{v}_2 > v_2^*$ is the exact parallel.

($\Longleftarrow$ necessity): Any strongly Pareto optimal allocation for q^0 is a perfect Bayesian equilibrium allocation.

The proof of necessity is very lengthy and similar to Maskin and Tirole (1990), and is omitted. See Maruyama (2005) for the further details. ■

Proof of Proposition 5: The necessary conditions for the grand mechanism to be weakly collusion-proof are obtained by solving (S_λ) for all λ, and set $y_{ij} = 0$ and $(x(\phi_{ij}), t(\phi_{ij}), w(\phi_{ij})) = (x_{ij}, t_{ij}, w_{ij})$ for all i, j. Then we have the same conditions as (4)-(7) except for ε_α being replaced by $\varepsilon_\lambda = \frac{\underline{\delta}_A}{\lambda + \underline{\delta}_S + \underline{\nu}_S} \in [0, 1)$. Then the proposition follows by solving the principal's maximization problem. ■

Proof of Lemma 1: First note that the principal can set $w_{22} = w_{21} = 0$ without violating other constraints. Secondly we claim that $x_{22} = x_{12}$ at the optimal solution. Suppose not, *i.e.*, $x_{22} > x_{12}$. Then by solving the program it can be shown that the optimal solution entails $x_{12} > x_{22}$, a contradiction. It follows that $x_{22} = x_{12}$, which implies $\pi_{22} = \pi_{12}$ from the coalition constraints. Under these constraints, the optimal solution for the principal entails the non-supervisory information benchmark outcome; $x = x^n$. ■

Proof of Proposition 6: Consider the game followed by the principal's proposal of Γ_1. Consider the following strategies taken by the supervisor and agent.

- Both employees accept Γ_1. If the supervisor has rejected it, the agent tells the truth to Γ_1.

- After observing σ, the type σ_1 supervisor proposes the null contract Λ_0, and the type σ_2 leaves. If the supervisor has proposed a side mechanism other than Λ_0, the agent responds optimally given updated belief about the supervisor's signal.
- After the type σ_2 supervisor has left, the agent reports the truth to Γ_1.
- When $\sigma = \sigma_1$, both types of the agent accept Λ_0. If the agent has rejected Λ_0, the supervisor stays and tells the truth.
- Both coalition members tell the truth to each other when they play Λ_0.
- The agent and the type σ_1 supervisor tell the truth to Γ_1.

These strategies yield the following final payoffs:

state	supervisor	agent
(σ_1, c_1)	0	$\Delta c x_{12}$
(σ_2, c_1)	0 (exit)	$\Delta c x_{22}$
(σ_1, c_2)	0	0
(σ_2, c_2)	0 (exit)	0

We will show by backward induction that these strategies together with the passive beliefs form a perfect Bayesian equilibrium.

After the type σ_1 supervisor proposes Λ_0 and the agent accepts it, it is optimal for both employees to tell the truth to each other. For the supervisor, telling σ_2 will never give strictly positive payoffs. For the agent, truth telling is optimal since Γ_1 is ex-post incentive compatible. If the agent rejects Λ_0, it is optimal for the supervisor to stay, report σ_1, and receive zero payment from the principal. Since the rejection of Λ_0 will not affect the supervisor's actions, the agent accepts Λ_0.

After the supervisor leaves, the agent tells the truth to Γ_1, since $(t^e(\cdot), x^e(\cdot))$ is incentive compatible.

The supervisor cannot do better by proposing a side mechanism other than Λ_0. Note first that any reporting rule that involves reporting σ_2 with probability one will never result in the actual report of σ_2, since the supervisor would exit to avoid paying the penalty $-D$. The supervisor always has to make this payment to the principal if she decides to participate and reports σ_2, but this penalty is greater than the largest possible gain for the coalition by misreporting. Hence the supervisor cannot gain from arrangement of misreporting σ_1 to σ_2. For a similar reason, the supervisor cannot gain from any type of promise to mix reports between σ_2 and σ_1, since no transfer payment is feasible to induce the supervisor's participation regardless of her updated belief about the agent's type. Finally, reporting rules that involves misreporting σ_2 as

σ_1 will never be profitable, since neither the supervisor nor agent's payoff will increase by such a misreporting. But the supervisor does not have a gain to make a transfer payment to the agent.

Note that the argument just described above holds regardless of the agent's belief about the supervisor's signal. It then follows that the type σ_2 supervisor's participation cannot be induced by any collusion arrangement, since her signal is payoff irrelevant.

It is easy to check that three other acceptance decisions are optimal. ■

Proof of Proposition 7: Consider the renegotiation rules described in the text. We need to show that the supervisor cannot strictly increase her payoff by proposing any side mechanism. Suppose in contrary that there exists such a side mechanism Λ_1'. Then the following claim holds:

Claim 1 *The supervisor never exits at the interim stage.*

This follows from the facts that the supervisor's signal is payoff irrelevant, and that her interim belief about the agent's types has full support. Both type of the supervisor stays in at the interim stage, and pretends to be whichever type gives higher payoff in the collusion stage.

Claim 2 *The supervisor's payoff is zero if she stay in* Γ_1' *after the collusion stage.*

The supervisor would report σ_1 with probability one to Γ_1', since the penalty from reporting σ_2 exceeds the biggest gain for the coalition from collusion. Then the agent has no incentive to bribe the supervisor. Then the only possibility is to request transfer payment from the agent by the threat of exiting. The agent, however, would not participate in such a collusion.

Since none of the employees are better off by the supervisor's exit, the proposition follows from the two claims. ■

References

[1] Celik, G. (2003): "Mechanism Design with Collusive Supervision," University of British Columbia.

[2] Che, Y.K., and J. Kim (2004): "Collusion-Proof Implementation of Optimal Mechanisms," University of Wisconsin.

[3] Cramton, P., and T.R. Palfrey (1995): "Ratifiable Mechanisms: Learning from Disagreement," *Games and Economic Behavior,* 10, 255-283.

[4] Crémer, J., and R. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247-1258.

[5] Faure-Grimaud, A., J.J. Laffont, and D. Martimort (2003): "Collusion, Delegation and Supervision with Soft Information," *Review of Economic Studies,* 70, 253-279.

[6] Kofman, F., and J. Lawarrée (1993): "Collusion in Hierarchical Agency," *Econometrica,* 61, 629-656.

[7] Laffont, J.J., and D. Martimort (1997): "Collusion under Asymmetric Information," *Econometrica,* 65, 875-911.

[8] —— and —— (2000): "Mechanism Design with Collusion and Correlation," *Econometrica,* 68, 309-342.

[9] Maruyama, T. (2005): "Informed Principal Problem with Correlated Signals," University of Pennsylvania.

[10] Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies, 66, 3-23.*

[11] McAfee, P., and J. McMillan, (1992): "Bidding Rings," *American Economic Review,* 82, 579-599.

[12] Melumad, N., D. Mookherjee, and S. Reichelstein (1995): "Hierarchical Decentralization of Incentive Contracts," *Rand Journal of Economics,* 26, 654-692.

[13] Mookherjee, D., and M. Tsumagari (2004): "The Organization of Supplier Networks: Effects of Delegation and Intermediation," *Econometrica, 72, 1179-1219.*

[14] Neeman, Z. (2004): "The Relevance of Private Information in Mechanism Design," *Journal of Economic Theory,* 117, 55-77.

[15] Quesada, L. (2004): "Collusion as an Informed Principal Problem," University of Wisconsin.

[16] Robert, J. (1991): "Continuity in Auction Design," *Journal of Economic Theory,* 55, 169-179.

[17] Tirole, J. (1986): "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics and Organization,* 2, 181-214.

[18] —— (1992): "Collusion and the Theory of Organizations," in *Advances in Economic Theory, Vol. 2,* ed. by J.J. Laffont. Cambridge University Press, 151-206.