Department of Economics

Supervisory Efficiency and Collusion in a Multiple-Agent Hierarchy

Xiaogang Che[1]
City, University of London

Yangguang Huang
Hong Kong University of Science and Technology

Le Zhang
Macquarie University

Department of Economics
Discussion Paper Series
No. 20/01

[1] Corresponding author: Xiaogang Che, Department of Economics, City, University of London, Northampton Square, London EC1V 0HB, UK. Email: xiaogang.che@city.ac.uk

Supervisory Efficiency and Collusion in a Multiple-Agent Hierarchy*

Xiaogang Che† Yangguang Huang‡ Le Zhang§

July 2019

Abstract

We analyze a principal-supervisor-two-agent hierarchy with inefficient supervision. The supervisor may collects a wrong signal on each agent's unobservable effort level. When reporting to the principal, the supervisor can collude with one or both agents to manipulate the signal in exchange for a bribe. In contract design, we identify a new trade-off between the loss from supervisor-agent collusion and the risk from inefficient supervision: Although allowing collusion makes shirking more attractive to the agents, it brings in a benefit because it can "correct" an incorrect negative signal when the agent has exerted effort. Such collusive supervision saves risk premiums that the principal has to pay for incentive provision. We characterize the principal's optimal contract choice among no-supervision, collusion-proof, and collusive-supervision contracts. We show that the collusive-supervision contract dominates when the supervisory efficiency is at an intermediate level.

Keywords: Optimal contract, hierarchy, collusion, supervisory efficiency, multiple agents.
JEL codes: D73, D82, D86.

*We would like to thank Kunal Sengupta for helpful discussions and comments at the initial stage of the project. We thank Yongmin Chen, Angus Chu, Kim-sau Chung, Andrew Clausen, Damian Damianov, Xinyu Hua, Yiquan Gu, Yuan Ju, Rongzhu Ke, Fahad Khalil, Tilman Klumpp, Jacques Lawarrée, Yunan Li, Zhiyun Li, Jingfeng Lu, David Martimort, Bibhas Saha, Ron Siegel, Kathryn Spier, Guofu Tan, Birendra Rai, Leslie Reinhorn for helpful comments. The paper benefited from discussions at China Meeting of Econometric Society, HKUST Workshop on Law and Economics, York Game Theory Conference, Durham Micro-Workshop, Liverpool University, Xi'an Jiaotong Liverpool University, Hong Kong Baptist University, University of Washington, SPMiD Conference, EARIE Conference, Macquarie University, and International Conference on Game Theory.

†City, University of London, UK. E-mail: chexiaogang0925@gmail.com.
‡Hong Kong University of Science and Technology, Hong Kong. E-mail: huangyg@ust.hk.
§Macquarie University, Australia. E-mail: lyla.zhang@mgsm.edu.au.

1 Introduction

In many organizations, agents are rewarded based on both objective performance indicators and subjective evaluations made by their supervisors. For example, most private companies establish some key performance indicators (KPI)[1] to evaluate employees such as the amount of sales and project completion progress, but at the same time the employees also have their performance subjectively assessed by intermediate supervisors. In the public sector, the promotion of low-level government officials usually depends not only on their performance on assigned tasks but also the opinions of higher-ups. In such cases, supervisory efficiency – whether the supervisor can accurately and correctly assess performance – plays an important role in agents' incentive provisions.

To explore this importance, we examine the presence of supervisory efficiency in a multiple-agent environment, where the multiple agents work as a team to complete a task. A typical example is that a manager of a pharmaceutical company hires a research team that consists of multiple researchers (agents). The manager cannot observe each researcher's contribution but can observe whether the development of a new drug is successful. Part of the team working hard may be sufficient to achieve the goal; thus, a moral hazard problem arises: Some researchers may free ride on others. To provide proper incentive, an intermediate supervisor is sent to assess the agents' performances and obtains signals about the contribution of each teammate. Having this supervisor can be valuable in preventing free riding and incentivizing the whole team.[2] On the other hand, having the supervisor may lead to two problems: First, the supervisor may make mistakes in reviewing the performance of these researchers; second, the supervisor may collude with (some or all of) the researchers and always report that "everybody works hard," which, in turn, mitigates the effectiveness of the supervision. In this case, is the supervisory information still useful? Should collusion be allowed or completely prevented? Will the supervisor collude with one agent only or both agents? More generally, how should the principal design the optimal contract?

With these research questions, we study the contracting problem in a multiple-agent hierarchy with both the possibility of inefficient supervision and collusion. Specifically, we consider a three-level hierarchy with a principal, a supervisor (she), and two productive agents (he/they). With some probability, the supervisory technology can be either efficient or inefficient: The former can always observe an accurate signal of the agents' effort levels, but the latter may collect a wrong signal. The principal initiates a contract with the supervisor and the two agents for a production task. Each agent can choose to either work or shirk. After the output is realized, the supervisor

[1]See `en.wikipedia.org/wiki/Performance_indicator`.

[2]In such cases, it may be difficult for agents to conduct cross-checking (Baliga, 1999) because each agent is involved in only some components of the task and thus lacks sufficient knowledge and information to evaluate the performances of others.

is sent to collect a signal about the effort level of each agent and to report to the principal. Before reporting the signal, the supervisor may collude with one or both agents to forge a signal that favors them; formally, the information is soft for the supervisor-agent coalition (Khalil et al., 2010). Contingent on the realized output level and the supervisory report, the principal pays transfers to the agents and the supervisor according to the contract.

We first establish the collusion-proof contract (Tirole, 1986) as a benchmark, where the supervisor and the agent(s) have no incentive to form any coalition under the payment structure. We show that when the supervisory efficiency, i.e., the accuracy of the signals, is sufficiently high, the collusion-proof contract dominates the no-supervisor contract, indicating that the supervisor is useful to the principal. After establishing the usefulness of inefficient supervision in the hierarchy, we explore the role of collusion between the supervisor and the agents.

We show that there exists a novel trade-off between the loss from supervisor-agent collusion and the risk due to inefficient supervision. Because the supervisory technology is imperfect, an incorrect signal may be observed. If the principal adopts the collusion-proof contract that induces the supervisor to tell the truth, an agent may be mistakenly punished when he works but the signal is wrong. Therefore, the principal must pay risk premiums to the agents as compensation if the payments depend on the inaccurate signal. As a result, inefficient supervision increases the cost of preventing collusion. Technically speaking, the interaction between inefficiency supervision and collusion drives the interlink between the coalition incentive compatibility (CIC) constraints and the individual incentive compatibility (IC) constraint; satisfying the CIC constraints raises the cost of satisfying the IC constraint.

With this insight mentioned above, we show that tolerating supervisor-agent collusion makes the principal better off under certain conditions because collusion helps the principal "correct" wrong supervisory signals: If an agent exerts high effort but the supervisory signal is negative, then a collusive supervisor will manipulate the report that reflects the true effort level. As a result, such collusive supervision lowers the cost of providing an incentive to agents. However, one should also take into account the downside of allowing collusion: When collusion is allowed, an agent can free ride on the other agent and bribe the supervisor to forge a positive report for him. The optimal collusive-supervision contract balances the gain and loss from allowing collusion based on the trade-off described above. In the collusive-supervision contract, the principal should prevent the full-coalition, which involves the supervisor and both agents, but allow the sub-coalition with one agent only. We then characterize the optimal contract design given various levels of supervisory efficiency. When supervisory efficiency is at an intermediate level, the collusive-supervision contract dominates both the collusion-proof and the no-supervisor contracts. This new finding provides an explanation for why some managers do not rely solely on objective performance indicators to reward workers but also consider subjective evaluations made by intermediate supervisors (MacLeod, 2003).

This paper is closely related to the vast literature on supervision and collusion in organizations and the design of optimal contracts. The seminal works by Tirole (1986, 1992) examine the role of corruptible supervision and the issues of incentive provision in a three-tier hierarchy. In the hierarchy, the supervisor can collude with the agent based on a side contract and conceal a negative signal or make a favorable report.[3] Tirole's central findings include that information from the corruptible supervisor remains useful for the principal and, moreover, that the optimal contract implemented by the principal is collusion-proof.[4] Tirole (1986) also argues that as supervisory information becomes less verifiable, the supervisor is less useful. Thus, when information is soft, i.e., entirely unverifiable, supervision becomes completely useless in the hierarchy. Kessler (2000) shows that if the supervisor's signal is verifiable, the possibility of collusion does not impose any additional cost to the principal.

In a single-agent hierarchy with soft supervisory information, Kofman and Lawarrée (1996) and Khalil et al. (2010) find that it may be optimal to allow collusion. In Kofman and Lawarrée (1996), the supervisor (auditor) can be either honest or dishonest, so the principal must adopt a high-powered incentive scheme to induce truth-telling. Allowing collusion is less expensive when the proportion of honest supervisors is large. In contrast, we focus on dishonest supervision only. In Section 6.3, we characterize the contract when the supervisor is honest but possibly inefficient and show that the collusive-supervision contract would also dominate the honest-supervision contract under certain conditions.

Khalil et al. (2010) consider the case in which the supervisor uses the ability to conceal positive signal to extort the agent. They show that both bribery (supervisor-agent collusion) and extortion weaken the incentive scheme, but the latter is more severe; thus, the principal benefits by allowing collusion to attenuate the room for extortion. We also reach the result that collusion becomes "useful" in lowering the principal's total cost in the contract. However, the basic trade-off in our model is different; the imperfect supervisory technology drives the usefulness of collusion in correcting supervisory signals in a multiple-agent environment. In Section 6.2, we show that such a trade-off

[3]This modeling framework opens up an important strand of literature on collusion in hierarchical agency. See, for example, Laffont and Tirole (1991); Kofman and Lawarrée (1993); Mookherjee and Png (1995); Strausz (1997); Lambert-Mogiliansky (1998); Khalil et al. (2013, 2015); Burlando and Motta (2015).

[4]The principal can prevent collusion by rewarding the supervisor with a payment equivalent to the agent's wage when she reports negative information about the agent's effort. In some environments, the principal may be better off allowing a certain scope for collusion between the supervisor and the agent when information is hard. See, for example, Che (1995); Olsen and Torsvik (1998).

does not exist in the setting with one agent only.

All the previous studies mentioned above consider collusion in a three-tier hierarchy with a single productive agent. Studies on contract design in multiple-agent organizations with supervisor-agent collusion are very limited.[5] Laffont (1990) examines a hidden gaming in which the supervisor can extort an agent by producing a negative report on the agent's individual contribution to the multiple-agent team. Laffont (1990) finds that if information is hard, the optimal payment scheme should be purely personalized; if, however, information is soft, it may be optimal to utilize some of the aggregate information to design the incentive scheme. We study a similar setting with soft information only. Our analyses contribute to this strand of literature by eliciting the reasons and characterizing the conditions under which allowing collusion benefits the principal. More interestingly, in such a multiple-agent environment, the collusive-supervision contract can sometimes even dominate the collusion-free contract with honest supervision.

Compared to a single-agent setting, the design of incentive schemes with multiple agents is more complicated because the output level depends on the joint effort of all agents. Therefore, when tailoring the payment scheme to an agent, the principal must account for the linkage between the effort of this agent and that of the other agent. In other words, in designing the contract, the principal needs to decide whether an agent's payment scheme should be tied to the performance of other agent(s) and, if so, how. A large body of literature has been established to rationalize both arguments of individual performance and team performance. See, for example, Hart and Holmstrom (1987); Ishiguro (2004); Bag and Pepito (2012); Ryall and Sampson (2016); Biener et al. (2018).

The characterizations of the no-supervision contract and the collusion-proof contract are in the line of individual performance, whereby an agent's reward is based on his own signal but not the other agent's. Nevertheless, the characterization of the collusive-supervision contract shows a special feature that combines both performance measures: If the agent's signal is negative, his rewards depend on the other agent's signal; if, however, his signal is positive, his reward does not depend on the other agent's signal. This feature echoes several existing results when using aggregate information (Laffont, 1990) and relative-performance evaluation (Che and Yoo, 2001) in team production environments.

The remainder of the paper proceeds as follows: We provide the model setup in Section 2. Sections 3 and 4 present the no-supervision contract and the collusion-proof contract, respectively,

[5] Most previous studies on contract design in multiple-agent organizations focus on the possibility of collusion among the agents. See, for example, Holmström and Milgrom (1990); Itoh (1993); Laffont and Martimort (1997, 2000); Baliga and Sjostrom (1998); Mookherjee and Tsumagari (2004); Severinov (2008); Kvaløy and Olsen (2019). In contrast, we examine the issue of supervisor-agent collusion in the multiple-agent setting.

as two benchmark cases. Section 5 shows how allowing collusion improves the incentive provision and characterizes the collusive-supervision contract. In Section 6, we discuss the robustness of our result when relaxing some assumptions. Section 7 concludes. Proofs are presented in the Appendix A.

2 Model Setup

Players and actions. We consider a three-level hierarchy with a principal, a supervisor, and two symmetric agents indexed by $i = A, B$. The principal is the owner of a firm and hires two agents as the productive units in the firm. Agent i can choose to either shirk or work, denoted by effort levels $e_i = 0$ and $e_i = 1$, respectively. Let $e \equiv (e_A, e_B)$ denote the pair of the two agents' efforts. The principal cannot observe the effort levels of agents.

After production, the output, $y \in \{H, L\}$, is realized and publicly observed, where H and L denote high and low output, respectively. We assume that $H > L > 0$ and $(H - L)$ is sufficiently large such that the principal strictly prefers that both agents work, i.e., $e = (1, 1)$. Therefore, in designing the incentive schemes, the principal aims to minimize the expected total payments from implementing $e = (1, 1)$. The probability of obtaining output H depends on both agents' efforts. Let $p(e) \in [0, 1]$ denote the probability that output H is realized, given e. The production process is teamwork, and thus, there is no separable output from an individual agent. The expected level of output is $p(e)H + (1 - p(e))L$. If both agents work, the probability of obtaining H is one, i.e., $p(1, 1) = 1$. If one or both agents shirk, then the output may still be high with some probability, characterized by $1 = p(1, 1) > p(0, 1) > p(0, 0) > 0$. By assuming $p(1, 1) = 1$, the agents face no uncertainty from the production technology when they both work. This setting facilities our focusing on the uncertainty entailed by the supervisory technology and clearly identifying the trade-off between the loss from supervisor-agent collusion and the risk from inefficient supervision.[6] By the symmetry of agents, we have $p_1 \equiv p(0, 1) = p(1, 0)$. This parameter p_1 measures how easy it is for an agent to free ride on the other agent.

Agent i has a utility function $U(w_i, e_i) = u(w_i) - \varphi e_i$, where w_i is the payment agent i receives, and $\varphi > 0$ denotes the disutility level of working. $u(w_i)$ satisfies $u(0) = 0$, $u'(\cdot) > 0$ and $u''(\cdot) \leqslant 0$. To ensure the existence of equilibrium, we assume that $u(w_i)$ satisfies the Inada conditions: $u'(0) = +\infty$, and $u'(+\infty) = 0$. Each agent accepts the contract as long as zero reservation utility is satisfied.

The supervisor is risk neutral and has zero reservation utility. After production, she collects a signal θ of the agents' effort levels from the state space $\Theta \equiv \{(1, 1), (0, 1), (1, 0), (0, 0)\}$. For each

[6] In Section 6.1, we show that the main results hold in the presence of production uncertainty, i.e., $0 < p(1, 1) < 1$.

signal θ, the first element represents the signal of agent A's effort level, and the second represents the signal of agent B's. The agents can also observe the signal θ. The supervisory technology is imperfect, which means that the supervisor can be either *efficient* or *inefficient* with probability λ and $(1-\lambda)$, respectively. The parameter $\lambda \in [0,1]$ captures the supervisory efficiency, which reflects the supervisor's ability to collect an accurate signal. If the supervisor is efficient, then the observed signal is accurate, i.e., $\theta = e$. If the supervisor is inefficient, she observes a random signal, that is, each $\theta \in \Theta$ is randomly observed with an equal probability of $1/4$. When the supervisor observes an incorrect signal, i.e., $\theta \neq e$, we interpret this as a mistake by the supervisor.[7] After collecting the signal, the supervisor sends a report $r \in \Theta$ about both agents' effort levels to the principal.

In the hierarchy, the principal contracts with the two agents and the supervisor before production. The contract specifies the conditions under which the supervisory information will be used and stipulates wage transfers $w_A^y(r) \geqslant 0$ and $w_B^y(r) \geqslant 0$ to the agents and a wage transfer $s^y(r) \geqslant 0$ to the supervisor according to output y and report r. After production, the principal collects the realized output y, and the supervisor observes a signal $\theta \in \Theta$ and strategically chooses a report r to maximize her payoff (in that she may collude with one or both of the agents to manipulate the signal). The principal then pays the transfers $w_i^y(r)$ to agent i and $s^y(r)$ to the supervisor following the contract.

Signal manipulation and side contract. After observing the signal but before reporting to the principal, the supervisor and the agents can choose to collude and manipulate the signal to favor their effort levels, i.e., reporting 1 with a signal of 0. Formally, information is soft for the supervisor-agent coalition in the sense that (i) the observed signal θ is not verifiable and can be manipulated costlessly, and (ii) the supervisor needs to collaborate with agent i to report $r \neq \theta$. In other words, the supervisor cannot forge information by herself, or it is too costly to do so without the agents' cooperation.

We model the collusion process as a side contract between the agent(s) and the supervisor that is assumed to be fully enforceable and unobservable by the principal. The side contract stipulates monetary transfers according to the realization of output y, signal θ, and report r. Denote the final payments to agent i and the supervisor in the coalition as $w_i^y(r|\theta)$ and $s^y(r|\theta)$, respectively. After signal manipulation, the members in the coalition divide the total payment by a Pareto efficient bargaining procedure that requires each party in the coalition to receive more than from choosing not to collude.[8]

[7] The overall probabilities that the supervisor collects a wrong and a correct signal are given by $\frac{3}{4}(1-\lambda)$ and $\lambda + \frac{1}{4}(1-\lambda)$, respectively.

[8] In Section 5.3, for the characterization of the collusive-supervision contract in the hierarchy, we assume that the side

The supervisor and the agents will not collude when they are indifferent between colluding and not colluding. In this case, the supervisor reports truthfully ($r = \theta$), and we have $w_i^y(r) = w_i^y(\theta)$ and $s^y(r) = s^y(\theta)$. A coalition is formed if and only if it strictly benefits all members. Hence, for a coalition between agent i and the supervisor, $w_i^y(r|\theta)$ and $s^y(r|\theta)$ must satisfy

(1) $$w_i^y(r|\theta) + s^y(r|\theta) = w_i^y(r) + s^y(r) \text{ for } r \neq \theta,$$

(2) $$w_i^y(r|\theta) > w_i^y(\theta) \text{ and } s^y(r|\theta) > s^y(\theta) \text{ for } r \neq \theta.$$

We say that a contract *allows collusion* if there exists some combination of θ and r such that (1) and (2) hold. For example, if the contract has the feature that $w_A^H(1,1) + s^H(1,1) > w_A^H(0,1) + s^H(0,1)$, when observing output $y = H$ and $\theta = (0,1)$, the supervisor can cooperate with agent A and report $r = (1,1)$ and then split the total payment $w_A^H(1,1) + s^H(1,1)$. Following the same definition, if the supervisor colludes with both agents, then it must be that

(3) $$w_A^y(r|\theta) + w_B^y(r|\theta) + s^y(r|\theta) = w_A^y(r) + w_B^y(r) + s^y(r) \text{ for } r \neq \theta,$$

(4) $$w_A^y(r|\theta) > w_A^y(\theta), w_B^y(r|\theta) > w_B^y(\theta), \text{ and } s^y(r|\theta) > s^y(\theta) \text{ for } r \neq \theta.$$

The objective of the supervisor-agent coalition is to forge a report r that maximizes the total payment from the principal.

Timing. Given the setup above, the timing of the game is as follows:

(1) The principal offers a contract specifying payments $\{w_A^y(r), w_B^y(r), s^y(r)\}$.
(2) The two agents and the supervisor decide whether to accept or reject the contract. If any of them rejects the contract, the game ends, and all parties receive their respective reservation utilities.
(3) If the contract has been accepted, the two agents simultaneously decide whether to work ($e_i = 1$) or to shirk ($e_i = 0$).
(4) The output y is realized and observed by all parties.
(5) The signal θ is realized and observed by the supervisor and the two agents.
(6) The supervisor and the agent(s) choose whether to collude and make a side contract. If the side contract is rejected, the supervisor will play non-cooperatively.
(7) The supervisor makes the report r to the principal.
(8) Transfers are paid according to the contract (and the side contract if necessary).

contract is implemented through a Nash bargaining game.

3 No-Supervision Contract

In the hierarchy, the supervisor may be inefficient and provide an incorrect signal. In addition, she may collude with the agent(s) against the principal's interests. The easiest way to avoid both problems is not to use the supervisory information. In this case, $s^y(r) = 0$, for all r and y. Agents' payments are based solely on output y, i.e., $w_i^y(r) = w_i^y$ for all r. We refer to this contract as the *no-supervision (no) contract*.

The principal's objective is to implement the effort choice $e = (1, 1)$ with the minimum total payments $C_{no} = p(1,1)(w_A^H + w_B^H) + (1-p(1,1))(w_A^L + w_B^L) = w_A^H + w_B^H$. We focus on the symmetric contract and take agent A as the representative. Given that agent B chooses to work ($e_B = 1$), the participation (IR) and the incentive compatibility (IC) constraints for agent A are

$$(IR_{no}) \qquad u(w_A^H) - \varphi \geqslant 0,$$

$$(IC_{no}) \qquad u(w_A^H) - \varphi \geqslant p_1 u(w_A^H) + (1-p_1)u(w_A^L).$$

Given that the payments to agents are symmetric, the principal's cost-minimization problem is

$$(P_{no}) \qquad \begin{aligned} &\min C_{no} = 2w_A^H \\ &\text{subject to } (IR_{no}),\ (IC_{no}),\ \text{and } w_i^y \geqslant 0,\ \forall\ i = A, B \text{ and } y = H, L. \end{aligned}$$

Note that (IR_{no}) is satisfied when (IC_{no}) is satisfied. The solution to (P_{no}) is the no-supervision contract as follows.

Proposition 1. *In the no-supervision contract, $w_i^L = 0$ and $w_i^H = \hat{w}_{no}^H$ for $i = A, B$, where $\hat{w}_{no}^H$ is determined by equation*

$$(\widehat{IC}_{no}) \qquad (1-p_1)u(\hat{w}_{no}^H) = \varphi.$$

The total payment of the principal is $\hat{C}_{no} = 2\hat{w}_{no}^H$.

In the absence of supervision, the principal compensates the agents only when the output is high.

4 Collusion-Proof Supervision

Next, we characterize the *collusion-proof (cp) contract* that leaves no incentive for the supervisor and the agents to collude. Perusing the collusion-proof contract helps us see the value of imperfect supervisory information, and we show that the collusion-proof contract indeed dominates the no-supervision contract when supervisory efficiency is sufficiently high.

We define collusion between the supervisor and both agents as a *full-coalition*, i.e., given $\theta = (0,0)$, the supervisor colluding with both agents reports $r = (1,1)$. We define collusion between the supervisor and only one of the two agents as a *sub-coalition*, i.e., given $\theta = (0,1)$ ($\theta = (1,0)$), the supervisor colludes with agent A (B) and reports $r = (1,1)$. Let $T^y(r) \equiv w_A^y(r) + w_B^y(r) + s^y(r)$ denote the aggregate transfer made by the principal to the two agents and the supervisor. $w_i^y(r) + s^y(r)$ represents the aggregate transfers to agent i and the supervisor. To prevent the full-coalition, the principal must ensure that reporting the truth does not result in strictly less joint payments for all possible coalitions, that is, satisfying the following coalition incentive compatibility (CIC) constraints (Tirole, 1992):

$$T^y(\theta) \geqslant T^y(r), \ \forall \ \theta, r \in \Theta.$$

Given output y, to satisfy constraints $T^y(\theta) \geqslant T^y(r)$, the aggregate payments to the two agents and the supervisor must be exactly the same across the four signal states:

$$(CIC_f) \qquad T^y(1,1) = T^y(1,0) = T^y(0,1) = T^y(0,0).$$

Similarly, to prevent sub-coalitions, given the other agent's signal, the payments to agent i and the supervisor must be the same:

$$(CIC_s) \qquad \begin{aligned} w_A^y(1,1) + s^y(1,1) &= w_A^y(0,1) + s^y(0,1), \\ w_A^y(1,0) + s^y(1,0) &= w_A^y(0,0) + s^y(0,0), \\ w_B^y(1,1) + s^y(1,1) &= w_B^y(1,0) + s^y(1,0), \\ w_B^y(0,1) + s^y(0,1) &= w_B^y(0,0) + s^y(0,0). \end{aligned}$$

From (CIC_f) and (CIC_s), we can easily derive the following lemma.

Lemma 1. *Collusion-proofness implies the following payment features to the agents:*

(a) $w_A^y(1,0) = w_A^y(1,1)$ *and* $w_A^y(0,1) = w_A^y(0,0)$ *for* $y = L, H$*;*

(b) $w_B^y(0,1) = w_B^y(1,1)$ *and* $w_B^y(1,0) = w_B^y(0,0)$ *for* $y = L, H$*.*

Therefore, to fully deter both types of coalitions, the incentive scheme for an agent should depend only on the signal of this agent but not on the other agent's. This feature will help us derive the collusion-proof contract. Taking agent A as the representative, given that $e_B = 1$ and the supervisor reports truthfully, the IR constraint is

$$(IR_{cp})\ \ \lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(1,1)) + u(w_A^H(1,0)) + u(w_A^H(0,1)) + u(w_A^H(0,0))\Big] - \varphi \geqslant 0,$$

and the IC constraint is

$$\begin{aligned}
&(IC_{cp})\\
&\ \lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(1,1)) + u(w_A^H(1,0)) + u(w_A^H(0,1)) + u(w_A^H(0,0))\Big] - \varphi\\
&\ \geqslant p_1\left\{\lambda u(w_A^H(0,1)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(1,1)) + u(w_A^H(1,0)) + u(w_A^H(0,1)) + u(w_A^H(0,0))\Big]\right\}\\
&\ + (1-p_1)\left\{\lambda u(w_A^L(0,1)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^L(1,1)) + u(w_A^L(1,0)) + u(w_A^L(0,1)) + u(w_A^L(0,0))\Big]\right\}.
\end{aligned}$$

where the left-hand side of (IC_{cp}) is the expected payoff when agent A works, and the two terms following λ and $1-\lambda$ represent the payoffs when the supervisor is efficient and inefficient, respectively. The right-hand side of (IC_{cp}) is the payoff when agent A shirks, and the two terms following p_1 and $(1-p_1)$ are the payoffs when the output is realized to be high and low, respectively.

The principal solves the following cost-minimization problem for the collusion-proof contract:

$$(P_{cp})\quad \begin{aligned}
&\min C_{cp} = \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(1,0) + T^H(0,1) + T^H(0,0)\Big]\\
&\qquad\quad = 2w_A^H(1,1) + s^H(1,1),\\
&\text{subject to } (CIC_f),\ (CIC_s),\ (IR_{cp}),\ (IC_{cp}),\\
&\qquad\qquad w^y(r) \geqslant 0,\ s^y(r) \geqslant 0,\ \forall\, r \in \Theta,\ i = A, B \text{ and } y = H, L.
\end{aligned}$$

The solution to (P_{cp}) is described in the following proposition:

Proposition 2. *Define* $\lambda^* = \frac{1-p_1}{1+p_1}$.

(a) *If* $\lambda < \lambda^*$, *the principal adopts the no-supervision contract in Proposition* 1.

(b) *If* $\lambda \geqslant \lambda^*$, *the collusion-proof contract is as follows:*

(b1) *For* $y = L$, *the agents and the supervisor do not obtain any rewards, i.e.,* $w_i^L(r) = s^L(r) = 0$, $\forall\, r \in \Theta,\ \ i = A, B$.

(b2) *For* $y = H$, *the payment structure is*

Report r	*Agent A*	*Agent B*	*Supervisor S*
$(1,1)$	$\tilde{w}_{cp}^H$	$\tilde{w}_{cp}^H$	0
$(1,0)$	$\tilde{w}_{cp}^H$	0	$\tilde{w}_{cp}^H$
$(0,1)$	0	$\tilde{w}_{cp}^H$	$\tilde{w}_{cp}^H$
$(0,0)$	0	0	$2\tilde{w}_{cp}^H$

where $\tilde{w}_{cp}^H$ is determined by equation

$$(\widetilde{IC}_{cp}) \qquad \lambda u(\tilde{w}_{cp}^H) + \frac{1}{2}(1-p_1)(1-\lambda)u(\tilde{w}_{cp}^H) = \varphi.$$

The principal incurs a total cost of $\tilde{C}_{cp} = 2\tilde{w}_{cp}^H$.

Proposition 2 shows how supervisory efficiency, measured by the parameter λ, affects the principal's contract choices. The collusion-proof contract is favorable only when supervisory information is sufficiently accurate, i.e., $\lambda \geqslant \lambda^*$. If the supervisor's signal is not accurate enough, the principal will prefer the no-supervision contract, in which the supervisory signal is ignored.

To prevent all possibilities of collusion, (IC_{cp}), (CIC_f), and (CIC_s) must be satisfied in the principal's cost-minimization problem. However, it is possible that an agent works but is nevertheless punished by the principal because the supervisor observes an incorrect signal. This potential mistake by the supervisor creates a conflict between truthful reporting and incentive provision. In other words, deterring collusion entails a risk of punishing a working agent due to incorrect supervisory signals. As a result, the agents require risk premiums to work, which raises the cost of implementing the collusion-proof contract.

5 Collusive Supervision

We have shown that the inefficient and manipulable supervisory technology is still valuable to the principal. However, when the collusion-proof contract prevents signal manipulation, it gives rise to the possibility that an inefficient supervisor reports a mistaken signal and harms the agents. Thus, in what follows, we explore the possibility of striking a balance in the trade-off between the loss from supervisor-agent collusion and the risk from inefficient supervision.

5.1 Incentive improvement by allowing collusion

When collusion is tolerated, the (CIC_f) and (CIC_s) constraints are removed from the principal's optimization problem. In designing the contract, the principal needs to consider signal manipulation and the payoffs resulting from the supervisor-agent collusion. Given that the principal always wants to implement $e = (1,1)$, the aggregate transfers to the two agents and the supervisor under signal $(1, 1)$ should be no less than under other signals, that is,[9]

$$T^y(1,1) \geqslant T^y(r) \; \forall \; r = \{(1,0),(0,1),(0,0)\}, y = L, H. \tag{5}$$

Hence, we can restrict our attention to the cases in which the supervisor-agent coalition make a positive adjustment of the signal instead of a negative adjustment. Taking agent A as the representative, there are three relevant cases of collusion. First, given $\theta = (0,0)$, form a sub-coalition and report $r = (1,0)$. Second, given $\theta = (0,0)$, form a full-coalition and report $r = (1,1)$. Third, given $\theta = (0,1)$, form a sub-coalition and report $r = (1,1)$.

Further, (5) implies that, when $\theta = (1,1)$, the supervisor always truthfully reports $r = (1,1)$. Hence, there is no need to reward the supervisor when she reports $r = (1,1)$, that is,

$$s^y(1,1) = 0 \text{ for } y = L, H. \tag{6}$$

When the signal is $\theta = (1,0)$, $(0,1)$, and $(0,0)$, there may be upward adjustment of the report. We use the notations $w_i^H(r|10)$, $w_i^H(r|01)$, and $w_i^H(r|00)$ to denote agent i's final payment after possible signal manipulation.

With the above analysis, we write the IC constraint of the representative agent A as follows

(IC_{cs})

$$\begin{aligned}
&\lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(1,1)) + u(w_A^H(r|10)) + u(w_A^H(r|01)) + u(w_A^H(r|00))\Big] - \varphi \\
&\geqslant p_1\Big\{\lambda u(w_A^H(r|01)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(1,1)) + u(w_A^H(r|10)) + u(w_A^H(r|01)) + u(w_A^H(r|00))\Big]\Big\} \\
&+ (1-p_1)\Big\{\lambda u(w_A^L(r|01)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^L(1,1)) + u(w_A^L(r|10)) + u(w_A^L(r|01)) + u(w_A^L(r|00))\Big]\Big\}.
\end{aligned}$$

We now focus on discussing (IC_{cs}) for collusive supervision. On the left-hand side of (IC_{cs}), the terms $u(w_A^H(r|10))$, $u(w_A^H(r|01))$, and $u(w_A^H(r|00))$ capture the potential benefit of allowing collusion. Because the contract implements $e = (1,1)$, if the signal is $\theta = (1,0)$, $(0,1)$, or $(0,0)$, this must be the supervisor's mistake due to inefficient supervision. Hence, if the supervisor makes a positive adjustment of the signal in her report, this corrective signal manipulation enlarges the left-hand side of (IC_{cs}) and thus increases the agent's incentive to work. However, if collusion for signal manipulation is allowed, the right-hand side of (IC_{cs}) will also increase because an agent

[9]Otherwise, for example, if $T^y(1,1) < T^y(1,0)$, then agents will always choose $e = (1,0)$ over $e = (1,1)$. The principal cannot implement $e = (1,1)$.

can shirk and bribe the agent to manipulate his signal. Given this trade-off between the risk from inefficient supervision and the loss from collusion, we now attempt to identify whether and under what conditions allowing supervisor-agent collusion would (i) generate a higher incentive for the agents to work and (ii) lead to an improvement over the no-supervision and the collusion-proof implementations.

We first rewrite (IC_{cs}) in the form of $Z_{cs} \geqslant 0$, where

$$\begin{aligned} Z_{cs} \equiv & \underbrace{\Big[\lambda + \frac{1}{4}(1-p_1)(1-\lambda)\Big] u(w_A^H(1,1))}_{\text{Term 1}} + \underbrace{\Big[\frac{1}{4}(1-p_1)(1-\lambda)\Big] u(w_A^H(r|10))}_{\text{Term 2}} \\ & + \underbrace{\Big[\frac{1}{4}(1-p_1)(1-\lambda)\Big] u(w_A^H(r|00))}_{\text{Term 3}} + \underbrace{\Big[\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1\Big] u(w_A^H(r|01))}_{\text{Term 4}} \\ & - \underbrace{(1-p_1)\Big\{\lambda u(w_A^L(r|01)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^L(1,1)) + u(w_A^L(r|10)) + u(w_A^L(r|01)) + u(w_A^L(r|00))\Big]\Big\}}_{\text{Term 5}} - \varphi. \end{aligned}$$

We then examine the maximization of Z term-by-term by choosing the payments. The analysis helps us to determine whether (certain types of) collusion would generate a higher incentive for the agents to work. The results are presented sequentially in Lemmas 2-6.

It is clear that the principal should minimize Term 5 by setting all payments associated with output L to zero. In this case, the supervisory information is useless, so the supervisor should also be rewarded with zero payment.

Lemma 2. *In the collusive-supervision contract, it is optimal for the principal not to reward the agents and the supervisor for $y = L$, i.e., $w_i^L(r) = s^L(r) = 0, \ \ \forall\, r \in \Theta, \ \ i = A, B$.*

When the output is low, allowing any supervisor-agent coalition to manipulate the signal would not improve the agents' incentive to work but have the opposite effect. Thus, offering zero payoffs to the agents and the supervisor across all possible signals is optimal.

Next, for Term 1, because $\lambda + \frac{1}{4}(1-p_1)(1-\lambda) \geqslant 0$, when $y = H$ and $r = (1,1)$, raising the payment $w_A^H(1,1)$ increases the agent's incentive to work. Denote $\check{w}_{cs}^H$ as the payment to agent A that implements $e = (1,1)$, which will be determined by the principal's cost minimization problem or, specifically, the binding IC constraint (see the analysis in the following). In this case, by (6), the supervisor should be rewarded with zero payment, i.e., $s^H(1,1) = 0$.

Lemma 3. *In the collusive-supervision contract, the principal should set $w_A^H(1,1) = w_B^H(1,1) = \check{w}_{cs}^H$ and $s^H(1,1) = 0$, where $\check{w}_{cs}^H$ is the equilibrium payment level that satisfies (IC_{cs}) for both agents.*

Now, consider Terms 2 associated with signal $\theta = (1,0)$. Obtaining this signal must result from an inefficient supervisor. Apparently, $\frac{1}{4}(1-p_1)(1-\lambda) \geqslant 0$, and thus, raising $w_A^H(r|10)$ provides a greater incentive for agent A to work. Therefore, to maximize Z_{cs}, when the signal is $\theta = (1,0)$, the principal should reward agent A with $w_A^H(r|10) = \check{w}_{cs}^H$ for $r = (1,1)$ and $(1,0)$; by symmetry, $w_B^H(r|01) = \check{w}_{cs}^H$ for $r = (1,1)$ and $(0,1)$. This implies the following result.

Lemma 4. *In the collusive-supervision contract, it is optimal to reward an agent if his own signal is positive, regardless of the other agent's signal, i.e.,* $w_A^H(1,1) = w_A^H(1,0) = \check{w}_{cs}^H$, *and* $w_B^H(1,1) = w_B^H(0,1) = \check{w}_{cs}^H$.

Therefore, agent A will not have the incentive to collude given $\theta = (1,0)$. However, note that by (5), $T(1,0) \leqslant T(1,1)$, and thus, $w_B^H(1,0) + s^H(1,0) \leqslant w_B^H(1,1) + s^H(1,1) = \check{w}_{cs}^H$. Hence, given $\theta = (1,0)$, depending on the reward schemes, the supervisor may collude with agent B and report $r = (1,1)$.

We now turn to Term 3, which requires more discussion. Similar to Term 2, obtaining a signal $\theta = (0,0)$ indicates that the supervisor is inefficient. Because $\frac{1}{4}(1-p_1)(1-\lambda) \geqslant 0$, a higher $w_A^H(r|00)$ provides a greater incentive for agent A to work. Given $\theta = (0,0)$, there are three possible reports: $r = (1,1)$, $(1,0)$, and $(0,0)$, corresponding to a full-coalition, a sub-coalition, and truthful reporting, respectively.

First, consider the full-coalition case. If the full-coalition is allowed by setting $T(1,1) > T(0,0)$, the report is $r = (1,1)$, and the total payment is $T^H(1,1) = w_A^H(1,1) + w_B^H(1,1) = 2\check{w}_{cs}^H$. Two agents and the supervisor divide this payment $T^H(1,1) = w_A^H(11|00) + w_B^H(11|00) + s^H(11|00)$. The supervisor must receive a strictly positive payoff ($s^H(11|00) > 0$) to manipulate the signal, so $w_A^H(11|00) < \check{w}_{cs}^H$. Alternatively, the full-coalition can be prevented by setting $T^H(1,1) = T^H(0,0)$ with $w_A^H(0,0) = w_B^H(0,0) = \check{w}_{cs}^H$ and $s^H(0,0) = 0$. This provides agent A with a greater incentive because $w_A^H(11|00) < w_A^H(0,0) = \check{w}_{cs}^H$. Therefore, instead of allowing the full-coalition, rewarding the agents directly as though they were working provides a higher incentive. In other words, full-coalition between the agents and the supervisor should be prevented in the collusive-supervision contract. In summary,

$$T^H(1,1) = T^H(0,0) = 2\check{w}_{cs}^H \text{ and } s(0,0) = 0. \tag{7}$$

Second, consider the sub-coalition case. If sub-coalition with agent A is implemented, then $w_A^H(1,0) + s(1,0) > w_A^H(0,0) + s(0,0) = \check{w}_{cs}^H$. From (5) and (7), $T^H(1,0) \leqslant T^H(0,0) = 2\check{w}_{cs}^H$. These two inequalities yield the following

$$\begin{aligned} w_B^H(1,0) &= T^H(1,0) - \left[w_A^H(1,0) + s^H(1,0)\right] \\ &< T^H(1,0) - w_{cs}^H \\ &\leqslant 2\check{w}_{cs}^H - \check{w}_{cs}^H \\ &= \check{w}_{cs}^H, \end{aligned}$$

which means that agent B's incentive to work is jeopardized. Thus, to prevent agent A and the supervisor from manipulating $\theta = (0,0)$ into $r = (1,0)$, the principal should set $s^H(1,0) = 0$. By symmetry, $s^H(0,1) = 0$.

We now summarize the payment schemes and collusion issues when $\theta = (0,0)$ as follows.

Lemma 5. *In the collusive-supervision contract,*

(a) *the payment schemes to the agents and the supervisor under signals $(0,0)$ and $(1,1)$ are the same, i.e., $w_A^H(1,1) = w_A^H(0,0) = \check{w}_{cs}^H$ and $s^H(1,1) = s^H(0,0) = 0$; by symmetry, $w_B^H(1,1) = w_B^H(0,0) = \check{w}_{cs}^H$;*

(b) *a sub-coalition that manipulates $\theta = (0,0)$ into $r = (1,0)$ or $(0,1)$ should be prevented, and therefore, $s^H(1,0) = s^H(0,1) = 0$.*

Note that the payment schemes to prevent the full-coalition are different in the collusion-proof contract and the collusive-supervision contract. This difference reflects the fact that satisfying the CIC constraints raises the cost of satisfying the IC constraint. With the observation of signal $(0,0)$, Term 3 indicates that a higher payment induces a greater incentive for the agent to work, as $\frac{1}{4}(1-p_1)(1-\lambda) \geqslant 0$. However, the implementation of the CIC constraints requires the principal to reward the supervisor, and that in turn discourages the agents from working. Finally, we consider Term 4. With signal $\theta = (0,1)$, there are two possibilities: First, agent A works, but the signal is wrong; second, agent A shirks, and the signal is correct. Therefore, the coefficient in front of $w_A^H(r|01)$ can be positive or negative. Let

$$\underline{\lambda} \equiv \frac{1-p_1}{1+3p_1}. \tag{8}$$

Apparently, if $\lambda \leqslant \underline{\lambda}$, $\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1 \geqslant 0$, then a higher $w_A^H(r|01)$ provides a greater incentive for agent A to work. If, however, $\lambda > \underline{\lambda}$, a lower $w_A^H(r|01)$ provides a greater incentive for agent A to work. We now separately examine these two cases.

When $\lambda \leqslant \underline{\lambda}$, from Lemma 4, $w_B^H(0,1) = \check{w}_{cs}^H$. From Lemma 3, we know that agent A's payment from the sub-coalition, $w_A^H(11|01)$, cannot be more than $\check{w}_{cs}^H$. Alternatively, setting $w_A^H(01|01) =$

$\check{w}^H_{cs}$, which implies that $w^H_A(0,1) = \check{w}^H_{cs}$, provides a greater incentive for agent A to work. Furthermore, combining the results in Lemmas 2-5, we can simplify function Z_{cs}:

$$\begin{aligned} Z_{cs} = &\left[\lambda + \frac{1}{4}(1-p_1)(1-\lambda)\right]u(w^H_A(1,1)) + \left[\frac{1}{4}(1-p_1)(1-\lambda)\right]u(w^H_A(1,0)) \\ &+ \left[\frac{1}{4}(1-p_1)(1-\lambda)\right]u(w^H_A(0,0)) + \left[\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1\right]u(w^H_A(0,1)) - \varphi. \end{aligned}$$

Given that $w^H_A(1,1) = w^H_A(1,0) = w^H_A(0,1) = w^H_A(0,0) = \check{w}^H_{cs}$, we have $Z_{cs} = (1-p_1)u(\check{w}^H_{cs}) - \varphi$. The equilibrium IC constraint, $(1-p_1)u(\check{w}^H_{cs}) - \varphi = 0$, is the same as ($\widehat{IC}_{no}$) for the no-supervision contract, which implies that $\check{w}^H_{cs} = \hat{w}^H_{no}$. Thus, when $\lambda \leqslant \underline{\lambda}$, all payment variation in the collusive-supervision contract is due to y. Allowing collusion cannot improve the agents' incentives over the no-supervision contract.

Now, consider the case when $\lambda > \underline{\lambda}$, in which lowering $w^H_A(r|01)$ increases agent A's incentive to work. Under the sub-coalition, $w^H_A(11|01) + s^H(11|01) = w^H_A(1,1) + s^H(1,1) = \check{w}^H_{cs}$, where $s^H(1,1) = 0$ by (6). To form the coalition, the supervisor must receive $s^H(11|01) > 0$, so agent A receives $w^H_A(11|01) < \check{w}^H_{cs}$. As collusion can lower $w^H_A(r|01)$, agent A has a higher incentive to work. This result indicates that when $\theta = (0,1)$ (or $(1,0)$), the principal should allow agent A (B) and the supervisor to form a sub-coalition that forges a report $r = (1,1)$ and shares the total payment $\check{w}^H_{cs}$. However, when the supervisory efficiency increases, the risk from mistaken punishment becomes smaller, and therefore the benefit from allowing such a sub-coalition will become smaller, in contrast to the collusion-proof implementation.

Summarizing the discussion above yields the following result.

Lemma 6. *In the collusive-supervision contract,*

(a) *if $\lambda \leqslant \underline{\lambda}$, it is optimal to reward the agent regardless of the signal, i.e., $w^H_A(1,1) = w^H_A(1,0) = w^H_A(0,1) = w^H_A(0,0) = \check{w}^H_{cs}$ for agent A. The contract is equivalent to the no-supervision contract;*

(b) *if $\lambda > \underline{\lambda}$, it is optimal to allow a sub-coalition, i.e., given $w^H_A(1,1) = \check{w}^H_{cs}$ and $w^H_A(0,1) = 0$ and $s(1,1) = s(0,0) = 0$, agent A colludes with the supervisor under signal $(0,1)$ to report $r = (1,1)$. With the side contract, agent A receives $w^H_A(11|01)$ and the supervisor receives $s^H(11|01)$, where $w^H_A(11|01) + s^H(11|01) = w^H_A(1,1) = \check{w}^H_{cs}$.*

The following proposition summarizes the payment scheme in the collusive-supervision contract derived by Lemmas 2-6.

Proposition 3. *Given $\lambda > \underline{\lambda}$, the collusive-supervision contract has the following features:*

(a) *$w^L_A(r) = 0 \ \forall\, r \in \Theta$, and $w^H_A(1,1) = w^H_A(1,0) = w^H_A(0,0)$ and $w^H_A(0,1) = 0$;*

(b) $w_B^L(r) = 0 \; \forall \; r \in \Theta$, *and* $w_B^H(1,1) = w_B^H(0,1) = w_B^H(0,0)$ *and* $w_B^H(1,0) = 0$;

(c) $s^y(r) = 0 \; \forall \; r \in \Theta$, *and* $y = L, H$;

(d) *The full-coalition is prevented as* $w_i^H(1,1) = w_i^H(0,0)$;

(e) *A sub-coalition between agent A (B) and the supervisor occurs when* $\theta = (0,1)$ ($\theta = (1,0)$).

Proposition 3 shows that allowing a certain type of collusion provides higher incentives for the agents to work. Note that, in the collusion-proof contract, the principal directly rewards the supervisor in return for her truthfully reporting the signal. However, under collusive supervision, the principal does not directly reward the supervisor. The supervisor only obtains a positive payment from the sub-coalition allowed by the contract. Proposition 3 also shows that, in the collusive-supervision contract, an agent's compensation scheme is determined by both individual and team performance measures: If the agent's signal is positive, his reward does not depend on the other agent's signal; if, however, his signal is negative, his reward depends on the other agent's signal.

With these results, the IC constraint is simplified to

$$(IC_{cs})$$
$$\lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(1,1)) + u(w_A^H(1,1)) + u(w_A^H(11|01)) + u(w_A^H(1,1))\Big] - \varphi$$
$$\geqslant p_1\left\{\lambda u(w_A^H(11|01)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(1,1)) + u(w_A^H(1,1)) + u(w_A^H(11|01) + u(w_A^H(1,1))\Big]\right\}.$$

The corresponding IR constraint is

$$(IR_{cs})\;\; \lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(1,1)) + u(w_A^H(1,1)) + u(w_A^H(11|01)) + u(w_A^H(1,1))\Big] - \varphi \geqslant 0,$$

which will be satisfied as long as (IC_{cs}) holds.

From the analysis above, we can then write the principal's cost minimization problem to implement $e = (1,1)$ as follows

$$(P_{cs}) \quad \begin{aligned} &\min C_{cs} = \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(11|10) + T^H(11|01) + T^H(0,0)\Big] \\ &\text{subject to } (IR_{cs}),\; (IC_{cs}), \\ &\quad w_i^y(r) \geqslant 0,\; w_i^y(r|\theta) \geqslant 0,\; s^y(r) \geqslant 0,\; s^y(r|\theta) \geqslant 0, \forall\; r \in \Theta,\; \theta \in \Theta,\; i = A, B. \end{aligned}$$

The solution to (P_{cs}) can be directly obtained from Proposition 3. Let us use $\alpha \in (0,1)$ to capture the share of the equilibrium payment $\check{w}_{cs}^H$ that the collusive agent obtains as the Pareto efficient bargaining outcome in the sub-coalition when manipulating $\theta = (0,1)$ ($\theta = (1,0)$) to report $(1,1)$, i.e., $w_A^H(11|10) = \alpha \check{w}_{cs}^H$. In this case, the supervisor receives $(1-\alpha)\check{w}_{cs}^H$ from the agent as the side payment or bribe.

Proposition 4. *Given $\lambda \geqslant \underline{\lambda}$, the collusive-supervision contract is as follows:*

(a) *For $y = L$, the agents and the supervisor do not obtain any rewards.*

(b) *For $y = H$, the payment structure is*

Report r	*Agent A*	*Agent B*	*Supervisor S*
$(1,1)$	$\check{w}_{cs}^H$	$\check{w}_{cs}^H$	0
$(1,0)$	$\check{w}_{cs}^H$	0	0
$(0,1)$	0	$\check{w}_{cs}^H$	0
$(0,0)$	$\check{w}_{cs}^H$	$\check{w}_{cs}^H$	0

where $\check{w}_{cs}^H$ is determined by the equation

$$(\widecheck{IC}_{cs}) \qquad \lambda\big[u(\check{w}_{cs}^H) - p_1 u(\alpha \check{w}_{cs}^H)\big] + (1-\lambda)(1-p_1)\big[\frac{3}{4}u(\check{w}_{cs}^H) + \frac{1}{4}u(\alpha \check{w}_{cs}^H)\big] = \varphi.$$

The principal pays a total amount $\check{C}_{cs} = 2\check{w}_{cs}^H$.

5.2 Comparisons across contracts

Now, we are prepared to compare (IC_{cs}) with (IC_{no}) and (IC_{cp}) to identify the conditions under which implementing the collusive-supervision contract is better than implementing the no-supervision and the collusion-proof contracts. Propositions 1-3 show that all payments associated with low output are zero. Let us then rewrite these IC constraints in the form of $Z \geqslant 0$, where

$$\begin{aligned}
Z_{no} &= (1-p_1)u(w_A^H(1,1)) - \varphi, \\
Z_{cp} &= \lambda u(w_A^H(1,1)) + \frac{1}{2}(1-p_1)(1-\lambda)u(w_A^H(1,1)) - \varphi, \\
Z_{cs} &= \lambda u(w_A^H(1,1)) - p_1 \lambda u(w_A^H(11|01)) + (1-p_1)(1-\lambda)\frac{1}{4}\big[3u(w_A^H(1,1)) + u(w_A^H(11|01))\big] - \varphi.
\end{aligned}$$

We first compare the no-supervision contract and collusive-supervision contract. The difference between Z_{cs} and Z_{no} is

$$\begin{aligned}
Z_{cs} - Z_{no} &= \big[\lambda + \frac{3}{4}(1-p_1)(1-\lambda) - (1-p_1)\big]u(w_A^H(1,1)) + \big[\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1\big]u(w_A^H(11|01)) \\
&= \big[\lambda p_1 - \frac{1}{4}(1-p_1)(1-\lambda)\big]\big[u(w_A^H(1,1)) - u(w_A^H(11|01))\big].
\end{aligned}$$

Clearly, $Z_{cs} - Z_{no}$ is increasing in λ. From (1) and (2), $u(w_A^H(1,1)) - u(w_A^H(11|01)) > 0$. From (8), $\lambda p_1 - \frac{1}{4}(1-p_1)(1-\lambda) < 0$ if and only if $\lambda > \underline{\lambda}$. Therefore, $Z_{cs} - Z_{no} > 0$ if and only if $\lambda > \underline{\lambda}$. In this case, collusive supervision provides greater incentives for the agents to work. The cutoff value of supervisory efficiency, $\underline{\lambda} = \frac{1-p_1}{1+3p_1}$, decreases in p_1. Intuitively, when p_1 becomes large, it is easier for one agent to free ride on the other, so the no-supervision contract is less preferred. It is easy to see that $\underline{\lambda} < \lambda^* = \frac{1-p_1}{1+p_1}$.

Next, we compare the collusion-proof contract and collusive supervision. Similarly, we take the difference between Z_{cs} and Z_{cp}.

$$\begin{aligned} Z_{cs} - Z_{cp} &= \frac{1}{4}(1-p_1)(1-\lambda)u(w_A^H(1,1)) + \big[\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1\big]u(w_A^H(11|01)) \\ &= \frac{1}{4}(1-p_1)(1-\lambda)\big[u(w_A^H(1,1)) - u(w_A^H(11|01))\big] - \lambda p_1 u(w_A^H(11|01)). \end{aligned} \tag{9}$$

When $\lambda = 1$, $Z_{cs} - Z_{cp} < 0$, the collusion-proof contract provides a higher incentive. Thus, when there is no inefficient supervision, the principal should adopt the collusion-proof contract.

When $\lambda = \lambda^* = \frac{1-p_1}{1+p_1}$, we have $Z_{cs} - Z_{cp} > 0$. Moreover,

$$\frac{d(Z_{cs} - Z_{cp})}{d\lambda} = -\frac{1}{4}(1-p_1) - p_1 u(w_A^H(11|01)) < 0,$$

so $Z_{cs} - Z_{cp}$ is decreasing in $\lambda \in [\lambda^*, 1]$. Hence, there must exist a unique cutoff value, denoted by $\overline{\lambda} \in (\lambda^*, 1)$, such that $Z_{cs} - Z_{cp} = 0$. $Z_{cs} > Z_{cp}$ if and only if $\lambda \in [\lambda^*, \overline{\lambda})$. The cutoff value,

$$\overline{\lambda} \equiv \frac{(1-p_1)\big[u(w^H(1,1)) + u(w_A^H(11|01))\big]}{\big[(1-p_1)u(w^H(1,1)) + (1+3p_1)u(w_A^H(11|01))\big]}, \tag{10}$$

depends on the agent's utility function and the side contract. $\overline{\lambda}$ decreases in p_1. A large p_1 implies a stronger incentive to free ride, and thus, the collusion-proof contract becomes more useful.

From the analysis above, we have the following result:

Proposition 5. *There exist unique cutoffs of the supervisory efficiency parameter $\underline{\lambda} \in (0, \lambda^*)$ and $\overline{\lambda} \in (\lambda^*, 1)$ such that if $\underline{\lambda} < \lambda < \overline{\lambda}$, collusive supervision induces a higher incentive for the agents to work than do no-supervision and collusion-proof supervision.*

The no-supervision, collusion-proof, and collusive-supervision contracts share a common feature: The total payment is two times the equilibrium wage, that is, $\hat{C}_{no} = 2\hat{w}_{no}^H$, $\tilde{C}_{cp} = 2\tilde{w}_{cp}^H$, and $\check{C}_{cs} = 2\check{w}_{cs}^H$, where $\hat{w}_{no}^H$, $\tilde{w}_{cp}^H$, and $\check{w}_{cs}^H$ are obtained by making (IC_{no}), (IC_{cp}), and (IC_{cs}) binding, respectively. Therefore, from Proposition 5, we can conclude the following:

Proposition 6. *It is optimal for the principal to use*

(a) *the no-supervision contract if $\lambda \leqslant \underline{\lambda}$;*

(b) *the collusive-supervision contract if $\underline{\lambda} < \lambda < \overline{\lambda}$; and*

(c) *the collusion-proof contract if $\lambda \geqslant \overline{\lambda}$.*

To understand this result, we discuss the gains and losses from using the inaccurate supervisory signal. An agent is wrongly punished when he works but the signal is negative. To mitigate the occurrence of this undesirable event, the collusive-supervision contract allows the agent to bribe the supervisor (by the amount $s(11|01)$) and correct the wrong signal. As a result, the principal gains from saving the risk premium payment to the agents. However, allowing collusion also incurs a loss because it creates the possibility that an agent shirks and bribes the supervisor to manipulate his report. In the case in which an agent shirks and his signal is negative, the agent receives nothing under the collusion-proof contract but still receives a positive payment ($w(11|01)$) under the collusive-supervision contract. Hence, collusion-proof supervision imposes a stronger punishment for shirking.

When λ is small, the supervisory technology is inaccurate. Allowing the payments to depend on the signal will expose the agents to excessive risks that require very high compensation. Hence, it is better to adopt the no-supervision contract. When λ is large, the supervisory technology is reliable enough, so it is optimal to let the payments be contingent on truthfully reported signals obtained from collusion-proof implementation. When λ is the intermediate range $[\underline{\lambda}, \overline{\lambda}]$, collusive supervision balances the gains and losses from using the inefficient supervisory technology and becomes the optimal way to provide incentives. Therefore, Proposition 6 demonstrates how the principal should design the contract based on the trade-off between loss from collusion and the risk from inefficient supervision.

5.3 Nash bargaining in the side contract

Based on the previous analysis, under collusive supervision, the supervisor colludes with agent A (respectively, agent B) to report $r = (1,1)$ when the signal is $\theta = (0,1)$ (respectively, $(1,0)$). The side contract specifies payments $w_A^H(11|01)$ and $s^H(11|01)$ when the signal is $\theta = (0,1)$ and the supervisor reports $r = (1,1)$. We assume that, in the sub-coalition, they divide the gain from collusion by a Nash bargaining procedure. Taking agent A as the representative, $\alpha \in (0,1)$ and $1-\alpha$ capture the bargaining power of agent A and the supervisor, respectively. The corresponding

Nash bargaining problem that determines $w_A^H(11|01)$ and $s^H(11|01)$ is

$$\text{(11)} \qquad \begin{aligned} &\max\big[w_A^H(11|01) - w_A^H(0,1)\big]^{\alpha}\big[s^H(11|01) - s^H(0,1)\big]^{1-\alpha}, \\ &\text{subject to } \; w_A^H(11|01) + s^H(11|01) = w_A^H(1,1) + s^H(1,1). \end{aligned}$$

From Proposition 3, $s^H(1,1) = 0$, $s^H(0,1) = 0$, $w_A^H(11|01) = \check{w}_{cs}^H$, and $w_A^H(0,1) = 0$. The bargaining problem reduces to $\max\big[w_A^H(11|01)\big]^{\alpha}\big[s^H(11|01)\big]^{1-\alpha}$ s.t. $w_A^H(11|01) + s^H(11|01) = \check{w}_{cs}^H$, which has the solution $w_A^H(11|01) = \alpha\check{w}_{cs}^H$ and $s^H(11|01) = (1-\alpha)\check{w}_{cs}^H$. Furthermore, (10) can be written as follows

$$\text{(12)} \qquad \overline{\lambda} = \frac{(1-p_1)\big[u(\check{w}_{cs}^H) + u(\alpha\check{w}_{cs}^H)\big]}{\big[(1-p_1)u(\check{w}_{cs}^H) + (1+3p_1)u(\alpha\check{w}_{cs}^H)\big]},$$

Note that the principal's choice between the collusive-supervision and collusion-proof contracts depends on $\overline{\lambda}$ in (12) that varies with α. Hence, the bargaining powers affect the principal's contract choice and total payment in equilibrium. The following proposition summarize the result.

Proposition 7. *Regarding the agent's bargaining power α in the side contract, we find that*

(a) *$\overline{\lambda}$ is decreasing in α.*

(b) *For $\lambda \in [\lambda^*, 1]$, if $\alpha \to 0$, then $\overline{\lambda} \to 1$, and the collusive-supervision contract dominates the collusion-proof contract; if $\alpha \to 1$, then $\overline{\lambda} \to \lambda^*$, and the opposite dominance holds.*

(c) *For $\lambda \in [\lambda^*, 1]$, $\check{C}_{cs}$ is increasing in α.*

Intuitively, if the agent has less bargaining power, having the supervisor manipulate the signal requires a larger bribe. Hence, with a smaller α, the agent has less incentive to shirk even though collusion is allowed, which reduces the downside of allowing collusion. As a result, the collusive-supervision contract becomes more attractive than the collusion-proof contract. In contrast, with a larger α, a smaller bribe is required to manipulate the signal. In collusive supervision, as the agents pay a smaller cost for shirking, the principal must pay more to incentivize the agents.

Figure 1 illustrates Propositions 6 and 7 under different combinations of the parameters λ and α. The Figure is generated by assuming that $u(w) = \sqrt{w}$ with $p_1 = 0.3$ and 0.7. The no-supervision contract is optimal in the left-hand area of the blue dashed line, whereas the collusion-proof contract is optimal in the right-hand area of the red curve. The dominance of the collusive-supervision contract occurs in the area between the blue dashed line and the red curve, representing $\underline{\lambda}$ and $\overline{\lambda}$, respectively. $\overline{\lambda}$ is decreasing in α, as shown in Proposition 7. In addition, the dominant regions of the no-supervision contract and the collusive-supervision contract shrink as p_1 increases. The reason is that when p_1 increases, shirking (and colluding with the supervisor) is more attractive to the agents, and in this case, the principal is better off adopting the collusion-proof contract.

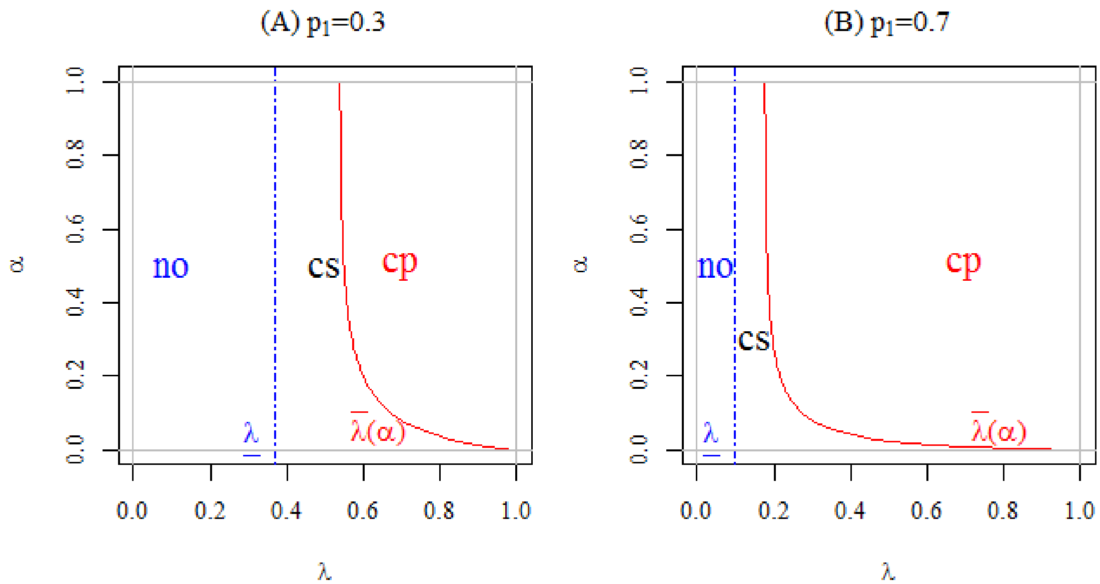

Figure 1: Illustration of the Optimal Contract Choice

6 Discussion

6.1 Production uncertainty in the multiple-agent setting

In the analysis above, we assume that $p(1,1) = 1$ to focus on the uncertainty entailed by the supervisory technology, which sets aside the uncertainty from production technology. As discussed in the literature review, most of previous studies on the three-level hierarchy with a single agent involve production uncertainty. Therefore, we examine the role of production uncertainty in the current setting with multiple agents.

When $0 < p(1,1) < 1$, there exists a possibility that both agents have exerted efforts but the production yields a lower output. Propositions 8 and 9 in Appendix B show that it is still beneficial to allow supervisor-agent collusion when the supervisory efficiency is at an intermediate level. In other words, the main results in Propositions 5 and 6 concerning optimal contract design in a multiple-agent hierarchy are robust to the setting of production uncertainty.

Next, let us consider an extreme case in which $\lambda = 1$ and $0 < p(1,1) < 1$, i.e., the supervisor always collects the correct signal. In this case, the problem of inefficient supervision disappears but uncertainty in production remains. The comparison among Z_{no}, Z_{cp}, and Z_{cs} shows that the collusion-proof contract strictly dominates the no-supervision contract and the collusive-supervision contract. Thus, it is optimal to prevent supervisor-agent collusion regardless of the level of production uncertainty. Therefore, this case clearly demonstrates that the failure of the collusion-proof principle in Tirole (1986) is rooted in inefficient supervision. The trade-off between inefficient supervision and collusion identified in this paper is novel.

6.2 Supervisory efficiency and collusion in a single-agent hierarchy

We have shown that a new trade-off between supervisory efficiency and collusion exists in the multiple-agent environment and that the principal would be better off allowing collusion. However, why does such a trade-off not exist in the single-agent setting? In this subsection, we examine the setting with one agent and show that the collusive-supervision contract is equivalent to the no-supervision contract.

We modify the model in the following way. There is one productive agent in the hierarchy. If the agent works, the probability of producing output H is one; in other words, there is no uncertainty regarding the production technology when the agent chooses to work. However, if the agent shirks, the probability of $y = H$ is $p \in (0, 1)$. As there is only one agent, the signal set is $\{0, 1\}$. We further assume that if the supervisor is inefficient (with probability $1 - \lambda$); she observes a signal of either 0 or 1 with equal probability, i.e., $1/2$. Let $w^y(1|0)$ denote the payoff of the agent when the signal is 0 but the supervisor and the agent collude to report 1. The IC constraint is as follows:

$$
\begin{aligned}
(IC_s) \qquad & \lambda u(w^H(1)) + (1-\lambda)\Big[\frac{1}{2}u(w^H(1)) + \frac{1}{2}u(w^H(1|0))\Big] - \varphi \\
& \geqslant p\Big\{\lambda u(w^H(1)) + (1-\lambda)\Big[\frac{1}{2}u(w^H(1)) + \frac{1}{2}u(w^H(1|0))\Big]\Big\} \\
& + (1-p)\Big\{\lambda u(w^L(1)) + (1-\lambda)\Big[\frac{1}{2}u(w^L(1)) + \frac{1}{2}u(w^L(1|0))\Big]\Big\},
\end{aligned}
$$

which can be rewritten as

$$
\begin{aligned}
& \Big\{\lambda u(w^H(1)) + (1-\lambda)\Big[\frac{1}{2}u(w^H(1)) + \frac{1}{2}u(w^H(1|0))\Big]\Big\} \\
& - \Big\{\lambda u(w^L(1)) + (1-\lambda)\Big[\frac{1}{2}u(w^L(1)) + \frac{1}{2}u(w^L(1|0))\Big]\Big\} \geqslant \frac{\varphi}{(1-p)}.
\end{aligned}
$$

Clearly, to provide incentives for the agent to work, it is optimal for the principal not to reward the agent when $y = L$. Furthermore, the principal should pay $w^H(1|0) = w^H(1)$ (and 0 to the supervisor). The equation can then be simplified to $(1-p)u(w^H(1)) \geqslant \varphi$, which is the same as ($IC_{no}$); the collusive-supervision contract is equivalent to the no-supervision contract. In other words, it is optimal for the principal to avoid both problems by not hiring a supervisor, and the trade-off does not exist in a single-agent hierarchy.

6.3 Honest supervision

A seemingly ideal contracting environment is that the supervisor is honest and always truthfully reports the observed signal ($r = \theta$). In this case, the optimization problem with an honest supervisor is the same as (P_{cp}) except that the (CIC_f) and (CIC_s) constraints are removed. We call this contract the *collusion-free (cf) contract*. The payment structure of the collusion-free contract is as follows. The principal pays zero to both agents when $y = L$. When $y = H$, each agent obtains $\tilde{w}^H_{cf}$ if the supervisory signal is positive, regardless of the signal about the other agent. By the equilibrium IC constraint, $\tilde{w}^H_{cf}$ is given by

$$\lambda u(\tilde{w}^H_{cf}) + \frac{1}{2}(1-p_1)(1-\lambda)u(\tilde{w}^H_{cf}) = \varphi,$$

which is the same as ($\widetilde{IC}_{cp}$), and thus, $\tilde{w}^H_{cf} = \tilde{w}^H_{cp}$. The supervisor does not need to be incentivized to tell the truth, so $s^y(r) = 0$ for all r. The total cost of the principal is $\tilde{C}_{cf} = 2\lambda\tilde{w}^H_{cf} + (1-\lambda)\tilde{w}^H_{cf} = (1+\lambda)\tilde{w}^H_{cf}$, which is clearly smaller than $\tilde{C}_{cp} = 2\tilde{w}^H_{cp}$. Therefore, the collusion-free contract (with an honest supervisor) always dominates the collusion-proof contract (with a corruptible supervisor).

Nevertheless, the collusion-free contract may not always dominate the no-supervision contract and the collusive-supervision contract. Example 1 demonstrates this possibility:

Example. Let $u(w) = \sqrt{w}$, $\varphi = 1$, $p_1 = 1/2$, $\lambda = 0.1$, and $\alpha = 1/2$. From ($\widehat{IC}_{no}$), $\hat{w}^H_{no} = 4$ and $\hat{C}_{no} = 8$. From ($\widetilde{IC}_{cp}$), $\tilde{w}^H_{cf} = \tilde{w}^H_{cp} \approx 9.467$, $\tilde{C}_{cp} \approx 18.935$ and $\tilde{C}_{cf} \approx 10.414$. From ($\breve{IC}_{cs}$), we have $\check{w}^H_{cs} \approx 4.310$ and $\check{C}_{cs} \approx 8.620$. Clearly, the collusion-free contract entails a higher cost than the no-supervision and collusive-supervision contracts. However, letting $u(w) = w$ and maintaining the same parameter values, we have $\hat{C}_{no} = 4$, $\tilde{C}_{cp} \approx 6.154$, $\check{C}_{cs} \approx 4.267$, and $\tilde{C}_{cf} \approx 3.385$. In this case, it is less costly to provide incentives to the agents using the collusion-free contract.

Therefore, surprisingly, having an honest but inaccurate supervisor may be more costly than having a corruptible and equally inaccurate supervisor. The reason is that it is possible to correct a wrong signal with a corruptible supervisor, but not with an honest one. Which scenario is better for the principal depends on the parameter values and the form of the utility function. We leave the general analysis of comparing honest and corruptible supervisors for future research.

7 Conclusion

We study a principal-supervisor-two-agent hierarchy in which the supervisor may be inefficient. The supervisor and the agents can collude to forge a supervisory report. We first demonstrate that, compared to the no-supervision contract, having the supervisor is still valuable for the principal. More importantly, we provide novel insights into collusion under inefficient supervision: There is a trade-off between the risk from inefficient supervision and the loss from collusion. When the supervisory efficiency is at an intermediate level, allowing a sub-coalition permits the supervisor to revise incorrect supervisory signals and reduce the risk rooted in inefficient supervisory technology. This key trade-off is essential for solving the optimal contract in a hierarchy with multiple agents and soft supervisory information. In the collusive-supervision contract, an agent's reward scheme is determined by both individual and team performance measures. We provide a full solution of the principal's contract design problem under different levels of supervisory efficiency. We obtain the novel finding that the collusive-supervision contract is optimal when the accuracy of the supervisory signal is at an intermediate level.

In most organizations including companies and institutions, the managerial hierarchy or system generally involves several layers from the top and intermediate managers to multiple productive units or agents. The agents' effort levels are typically unobservable, and thus, subjective and objective assessments are generally combined to evaluate their performances. Our analysis above provides a fundamental setting to model such complicated managerial environments and the related contract design problems. It sheds light on a novel impact jointly caused by supervisory efficiency and collusion in a multiple-agent hierarchy, which is not present in a single-agent framework.

Our results offer several managerial implications. First, when designing contracts, managers should consider the supervisors' capabilities on performance or promotion assessments. For example, as shown above, according to the level of supervisory inefficiency, the payment scheme in the optimal contract choice may or may not depend on the supervisory reports of other agents in the team. Second, our analysis not only answers the question of whether and when collusion should be tolerated but also examines to what extent. This serves as a guidance to managers on what types of collusion she should tolerate in practice. Finally, one should interpret our results in a caution that although allowing collusion is beneficial in some certain circumstances, in the long run it is more desirable to advance supervisory efficiency, i.e., rising accuracy and reliability of supervisory technology, to incentivize agents and solve the common complaint from the agents that "my supervisor does not evaluate my performance correctly."

References

Bag, P. K. and N. Pepito (2012): "Peer transparency in teams: Does it help or hinder incentives?" *International Economic Review*, 53, 1257–1286. 1

Baliga, S. (1999): "Monitoring and collusion with 'soft' information," *Journal of Law, Economics, and Organization*, 15, 434–440. 2

Baliga, S. and T. Sjostrom (1998): "Decentralization and collusion," *Journal of Economic Theory*, 83, 196 – 232. 5

Biener, C., M. Eling, A. Landmann, and S. Pradhan (2018): "Can group incentives alleviate moral hazard? The role of pro-social preferences," *European Economic Review*, 101, 230–249. 1

Burlando, A. and A. Motta (2015): "Collusion and the organization of the firm," *American Economic Journal: Microeconomics*, 7, 54–84. 3

Che, Y. K. (1995): "Revolving doors and the optimal tolerance for agency collusion," *RAND Journal of Economics*, 26, 378–397. 4

Che, Y.-K. and S.-W. Yoo (2001): "Optimal incentives for teams," *American Economic Review*, 91, 525–541. 1

Hart, O. and B. Holmstrom (1987): "The theory of contracts," *Advances in Economic Theory: Fifth World Congress*, 71–155. 1

Holmström, B. and P. Milgrom (1990): "Regulating trade among agents," *Journal of Institutional and Theoretical Economics (JITE)*, 85–105. 5

Ishiguro, S. (2004): "Collusion and discrimination in organizations," *Journal of Economic Theory*, 116, 357–369. 1

Itoh, H. (1993): "Coalitions, incentives, and risk sharing," *Journal of Economic Theory*, 60, 410–427. 5

Kessler, A. S. (2000): "On monitoring and collusion in hierarchies," *Journal of Economic Theory*, 91, 280–291. 1

Khalil, F., D. Kim, and J. Lawarrée (2013): "Contracts offered by bureaucrats," *RAND Journal of Economics*, 44, 686–711. 3

Khalil, F., J. Lawarrée, and T. J. Scott (2015): "Private monitoring, collusion, and the timing of information," *RAND Journal of Economics*, 46, 872–890. 3

Khalil, F., J. Lawarrée, and S. Yun (2010): "Bribery versus extortion: allowing the lesser of two evils," *RAND Journal of economics*, 41, 179–198. 1

Kofman, F. and J. Lawarrée (1993): "Collusion in hierarchical agency," *Econometrica*, 629–656. 3

——— (1996): "On the optimality of allowing collusion," *Journal of Public Economics*, 61, 383–407. 1

Kvaløy, O. and T. E. Olsen (2019): "Relational Contracts, Multiple Agents, and Correlated Outputs," *Management Science*, Forthcoming. 5

Laffont, J.-J. (1990): "Analysis of hidden gaming in a three-level hierarchy," *Journal of Law, Economics, and Organization*, 6, pp. 301–324. 1

Laffont, J.-J. and D. Martimort (1997): "Collusion under asymmetric information," *Econometrica*, 875–911. 5

——— (2000): "Mechanism design with collusion and correlation," *Econometrica*, 68, 309–342. 5

Laffont, J.-J. and J. Tirole (1991): "The politics of government decision-making: A theory of regulatory capture," *Quarterly Journal of Economics*, 106, pp. 1089–1127. 3

Lambert-Mogiliansky, A. (1998): "On optimality of illegal collusion in contracts," *Review of Economic Design*, 3, 303–328. 3

MacLeod, W. B. (2003): "Optimal contracting with subjective evaluation," *American Economic Review*, 93, 216–240. 1

Mookherjee, D. and I. P. L. Png (1995): "Corruptible law enforcers: How should they be compensated?" *Economic Journal*, 105, 145–59. 3

Mookherjee, D. and M. Tsumagari (2004): "The organization of supplier networks: effects of delegation and intermediation," *Econometrica*, 72, 1179–1219. 5

Olsen, T. E. and G. Torsvik (1998): "Collusion and renegotiation in hierarchies: A case of beneficial corruption," *International Economic Review*, 39, 413–38. 4

Ryall, M. D. and R. C. Sampson (2016): "Contract structure for joint production: risk and ambiguity under compensatory damages," *Management Science*, 63, 1232–1253. 1

Severinov, S. (2008): "The value of information and optimal organization," *RAND Journal of Economics*, 39, 238–265. 5

Strausz, R. (1997): "Delegation of monitoring in a principal-agent relationship," *Review of Economic Studies*, 64, pp. 337–357. 3

Tirole, J. (1986): "Hierarchies and bureaucracies: On the role of collusion in organizations," *Journal of Law, Economics, and Organization*, 2, 181. 1, 6.1

——— (1992): "Collusion and the Theory of Organizations," *In J.-J. Laffont, ed., Advances in Economic Theory, Sixth World Congress, Vol. 2. New York: Cambridge University Press*. 1, 4

Appendix A

Proof of Proposition 1. Rewrite (IC_{no}) as $(1-p_1)u(w_A^H)-(1-p_1)u(w_A^L)-\varphi \geqslant 0$. The Lagrangian for ($P_{no}$) is

$$\mathcal{L} = 2w_A^H - \delta\big[(1-p_1)u(w_A^H) - (1-p_1)u(w_A^L) - \varphi\big],$$

with the additional non-negativity constraints. The Kuhn-Tucker conditions for minimization are

$$\text{(A1)}: \quad \frac{\partial \mathcal{L}}{\partial w_A^H} = 2 - \delta(1-p_1)u'(w_A^H) \geqslant 0,\ w_A^H \geqslant 0,\ \text{ and }\ w_A^H \frac{\partial \mathcal{L}}{\partial w_A^H} = 0;$$

$$\text{(A2)}: \quad \frac{\partial \mathcal{L}}{\partial w_A^L} = \delta(1-p_1)u'(w_A^L) \geqslant 0,\ w_A^L \geqslant 0,\ \text{ and }\ w_A^L \frac{\partial \mathcal{L}}{\partial w_A^L} = 0;$$

plus the complementary slackness conditions for the constraints.

Step 1. It is impossible to have $\delta = 0$ because it implies that $w_A^H = 0$ in (A1), which violates (IC_{no}), yielding a contradiction. We therefore have $\delta > 0$ and $w_A^H > 0$, which further give $\partial\mathcal{L}/\partial w_A^H = 0$ and $\delta = 2/[(1-p_1)u'(w_A^H)]$.

Step 2. Given $\delta > 0$, (A2) indicates that $\partial\mathcal{L}/\partial w_A^L > 0$ and $w_A^L = 0$, which indicates no reward in the no-supervision contract when the output is low.

Step 3. Given $\hat{w}_A^L = 0$, when (IC_{no}) is binding, denoted by ($\widehat{IC}_{no}$), we have the value of $\hat{w}_{no}^H$:

$$(1-p_1)u(\hat{w}_{no}^H) = \varphi \iff \hat{w}_{no}^H = u^{-1}\left(\frac{\varphi}{1-p_1}\right).$$

The total payment of the principal is $\hat{C}_{no} = 2\hat{w}_{no}^H$. $\square$

Proof of Lemma 1. For part (a), we have $T^y(1,1) = T^y(1,0)$ from (CIC_f), that is,

$$w_A^y(1,1) + w_B^y(1,1) + s^y(1,1) = w_A^y(1,0) + w_B^y(1,0) + s^y(1,0).$$

Furthermore, (CIC_s) indicates that $w_B^y(1,1) + s^y(1,1) = w_B^y(1,0) + s^y(1,0)$. Therefore, $w_A^y(1,0) = w_A^y(1,1)$.

Similarly, (CIC_f) requires $T^y(0,1) = T^y(0,0)$, and therefore,

$$w_A^y(0,1) + w_B^y(0,1) + s^y(0,1) = w_A^y(0,0) + w_B^y(0,0) + s^y(0,0).$$

Again, (CIC_s) indicates that $w_B^y(0,1) + s^y(0,1) = w_B^y(0,0) + s^y(0,0)$. Hence, $w_A^y(0,0) = w_A^y(0,1)$.

Part (b) holds because the two agents are symmetric. □

Proof of Proposition 2. Given Lemma 1, rewrite (IC_{cp}) as

$$\begin{aligned}
&\lambda u(w_A^H(1,1)) + (1-\lambda)\big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{2}u(w_A^H(0,0))\big] - \varphi \\
&\geqslant p_1\Big\{\lambda u(w_A^H(0,0)) + (1-\lambda)\big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{2}u(w_A^H(0,0))\big]\Big\} \\
&+ (1-p_1)\Big\{\lambda u(w_A^L(0,0)) + (1-\lambda)\big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{2}u(w_A^L(0,0))\big]\Big\}.
\end{aligned}$$

For the principal's cost-minimization problem (P_{cp}), the Lagrangian is

$$\begin{aligned}
\mathcal{L} = {} & 2w_A^H(1,1) + s^H(1,1) \\
& - \delta\Big\{\lambda u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(0,0)) + (1-p_1)(1-\lambda)\big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{2}u(w_A^H(0,0))\big] \\
& - \lambda(1-p_1)u(w_A^L(0,0)) - (1-p_1)(1-\lambda)\big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{2}u(w_A^L(0,0))\big] - \varphi\Big\}.
\end{aligned}$$

with the additional non-negativity constraints. The Kuhn-Tucker conditions for minimization are

$$\text{(B1)}: \quad \frac{\partial \mathcal{L}}{\partial w_A^H(1,1)} = 2 - \delta\big[\lambda + \frac{1}{2}(1-\lambda)(1-p_1)\big]u'(w_A^H(1,1)) \geqslant 0,$$
$$w_A^H(1,1) \geqslant 0 \text{ and } w_A^H(1,1)\frac{\partial \mathcal{L}}{\partial w_A^H(1,1)} = 0;$$

$$\text{(B2)}: \quad \frac{\partial \mathcal{L}}{\partial w_A^L(1,1)} = \delta\big[\frac{1}{2}(1-\lambda)(1-p_1)\big]u'(w_A^L(1,1)) \geqslant 0,$$
$$w_A^L(1,1) \geqslant 0 \text{ and } w_A^L(1,1)\frac{\partial \mathcal{L}}{\partial w_A^L(1,1)} = 0;$$

$$\text{(B3)}: \quad \frac{\partial \mathcal{L}}{\partial w_A^H(0,0)} = -\delta\big[\frac{1}{2}(1-\lambda)(1-p_1) - \lambda p_1\big]u'(w_A^H(0,0)) \geqslant 0,$$
$$w_A^H(0,0) \geqslant 0 \text{ and } w_A^H(0,0)\frac{\partial \mathcal{L}}{\partial w_A^H(0,0)} = 0;$$

$$\text{(B4)}: \quad \frac{\partial \mathcal{L}}{\partial w_A^L(0,0)} = \delta\big[\lambda(1-p_1) + \frac{1}{2}(1-p_1)(1-\lambda)\big]u'(w_A^L(0,0)) \geqslant 0,$$
$$w_A^L(0,0) \geqslant 0 \text{ and } w_A^L(0,0)\frac{\partial \mathcal{L}}{\partial w_A^L(0,0)} = 0;$$

$$\text{(B5)}: \quad \frac{\partial \mathcal{L}}{\partial s^H(1,1)} = 1 \geqslant 0, \; s^H(1,1) \geqslant 0 \text{ and } s^H(1,1)\frac{\partial \mathcal{L}}{\partial s^H(1,1)} = 0;$$

plus the complementary slackness conditions for the constraints.

Step 1. It is impossible to have $\delta = 0$ because it implies that $\partial\mathcal{L}/\partial w_A^H(1,1) = 2 > 0$ and $w_A^H(1,1) = 0$ in (B1), which violates (IC_{no}), yielding a contradiction. Therefore, we should have $\delta > 0$ and $w_A^H(1,1) > 0$, implying that $\partial\mathcal{L}/\partial w_A^H(1,1) = 0$ and $\delta = 2/\left[\lambda + \frac{1}{2}(1-\lambda)(1-p_1))u'(w_A^H(1,1)\right]$. Furthermore, by Lemma 1, we have $w_A^L(1,1) = w_A^L(1,0) > 0$.

Step 2. Given $\delta > 0$, (B2) and (B4) show that $\partial\mathcal{L}/\partial w_A^L(1,1) > 0$ and $\partial\mathcal{L}/\partial w_A^L(0,0) > 0$, respectively, given that $w_A^L(1,1) = w_A^L(0,0) = 0$. By Lemma 1, we have $w_A^L(1,0) = w_A^L(0,1) = 0$. Furthermore, as $s^L(r)$ for all r do not appear in the Lagrangian, we are free to choose the minimum value ($s^L(r) = 0$) to satisfy the non-negativity constraints.

Step 3. From (B5), we clearly have $s^H(1,1) = 0$.

Step 4. From (B3), there is a unique cutoff, denoted by λ^*, that satisfies the following equation:

$$\frac{1}{2}(1-\lambda^*)(1-p_1) - \lambda^* p_1 = 0.$$[10]

This equation gives the following two cases: $\lambda < \lambda^*$ and $\lambda \geqslant \lambda^*$.
(a) When $\lambda < \lambda^*$, $\partial\mathcal{L}/\partial w_A^H(0,0)$ must be zero, implying that $w_A^H(0,0) = +\infty$. Therefore, there is no advantage for the principal in sending a supervisor, and he is better off implementing the no-supervision contract.
(b) When $\lambda \geqslant \lambda^*$, $\partial\mathcal{L}/\partial w_A^H(0,0)$ is strictly positive, which implies that $w_A^H(0,0) = 0$. By Lemma 1, we also have $w_A^H(0,1) = 0$. Furthermore, because (CIC_f) and (CIC_s) should hold, we then have $s^H(0,0) = 2w_A^H(1,1)$ and $s^H(1,0) = s^H(0,1) = w_A^H(1,1)$.

Finally, we denote $w_A^H(1,1)$ in equilibrium by $\tilde{w}_{cp}^H$, which is uniquely determined by ($\widetilde{IC}_{cp}$): $\lambda u(\tilde{w}_{cp}^H) + (1-p_1)(1-\lambda)\frac{1}{2}u(\tilde{w}_{cp}^H) = \varphi$.

By the symmetry of the two agents, we can easily establish the payment structure for agent B, with the principal incurring a total cost of $\tilde{C}_{cp} = 2\tilde{w}_{cp}^H$. □

[10] This condition can also be used to compare ($\widehat{IC}_{no}$) and ($\widetilde{IC}_{cp}$).

Proof of Proposition 7. Part (a): Differentiating $\overline{\lambda}$ with respect to α yields

$$\begin{aligned}\frac{\partial\overline{\lambda}}{\partial\alpha} &= \frac{(1-p_1)u'(\alpha\check{w}_{cs}^H)\check{w}_{cs}^H}{[(1-p_1)u(\check{w}_{cs}^H)+(1+3p_1)u(\alpha\check{w}_{cs}^H)]} - \frac{(1-p_1)(1+3p_1)[u(\check{w}_{cs}^H)+u(\alpha\check{w}_{cs}^H)]u'(\alpha\check{w}_{cs}^H)\check{w}_{cs}^H}{[(1-p_1)u(\check{w}_{cs}^H)+(1+3p_1)u(\alpha\check{w}_{cs}^H)]^2} \\ &= \frac{-4p_1(1-p_1)u(\check{w}_{cs}^H)u'(\alpha\check{w}_{cs}^H)\check{w}_{cs}^H}{[(1-p_1)u(\check{w}_{cs}^H)+(1+3p_1)u(\alpha\check{w}_{cs}^H)]^2} \\ &< 0.\end{aligned}$$

$\overline{\lambda}$ is decreasing in α.

Part (b): Plugging $\alpha = 0$ into (12) gives $\overline{\lambda} = 1$. With $\alpha = 0$, $(\breve{IC}_{cs})$ can be written as follows:

$$\lambda u(\check{w}_{cs}^H) + (1-\lambda)(1-p_1)\frac{3}{4}u(\check{w}_{cs}^H) = \varphi.$$

Comparing this equation with $(\widetilde{IC}_{cp})$ immediately indicates that $\check{w}_{cs}^H \leqslant \tilde{w}_{cp}^H$ for any $\lambda \in [\lambda^*, 1]$.

Plugging $\alpha = 1$ into (12) gives $\overline{\lambda} = \lambda^*$. With $\alpha = 1$, $(\breve{IC}_{cs})$ can be written as

$$\lambda u(\check{w}_{cs}^H) + \frac{1}{2}(1-p_1)(1-\lambda)u(\check{w}_{cs}^H) + \left[\frac{1}{2}(1-\lambda)(1-p_1) - \lambda p_1\right]u(\check{w}_{cs}^H) = \varphi.$$

Given that $\frac{1}{2}(1-\lambda)(1-p_1) - \lambda p_1 < 0$ for any $\lambda \in [\lambda^*, 1]$, comparing the equation above with $(\widetilde{IC}_{cp})$ immediately indicates that $\check{w}_{cs}^H \geqslant \tilde{w}_{cp}^H$.

Part (c): Let us define the terms associated with bargaining power α in $(\breve{IC}_{cs})$ as the function $\kappa \equiv (\frac{1}{4}(1-\lambda)(1-p_1) - \lambda p_1)u(\alpha\check{w}_{cs}^H)$. Clearly, if $\lambda > \underline{\lambda}$, then $\kappa < 0$, and therefore, an increase in bargaining power α leads to a higher $\check{w}_{cs}^H$ to satisfy $(\breve{IC}_{cs})$. □

Appendix B

(For Online Publication)

In this appendix, we examine the case in which $0 < p(1,1) < 1$ and show that the main result of the paper is robust. Denote $p_2 \equiv p(1,1)$ and assume that $0 < p_1 < p_2 < 1$, which means that having more agents working on production generates a higher probability of obtaining output H. In what follows, we provide the IR and IC constraints for the no-supervision contract, the collusion-proof contract, and the collusive-supervision contract. For the no-supervision contract, (IR_{no}) and (IC_{no}) are replaced by the following two constraints, respectively:

$$(IR'_{no}) \qquad p_2 u(w_A^H) + (1-p_2)u(w_A^L) - \varphi \geqslant 0.$$

$$(IC'_{no}) \qquad p_2 u(w_A^H) + (1-p_2)u(w_A^L) - \varphi \geqslant p_1 u(w_A^H) + (1-p_1)u(w_A^L).$$

For collusion-proof contract, (IR_{cp}) and (IC_{cp}) can be rewritten, respectively, as

$$(IR'_{cp}) \qquad \begin{aligned} & p_2\left\{\lambda u(w_A^H(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{2}u(w_A^H(0,0))\Big]\right\} \\ & + (1-p_2)\left\{\lambda u(w_A^L(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{2}u(w_A^L(0,0))\Big]\right\} - \varphi \geqslant 0. \end{aligned}$$

$$(IC'_{cp}) \qquad \begin{aligned} & p_2\left\{\lambda u(w_A^H(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{2}u(w_A^H(0,0))\Big]\right\} \\ & + (1-p_2)\left\{\lambda u(w_A^L(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{2}u(w_A^L(0,0))\Big]\right\} - \varphi \\ & \geqslant p_1\left\{\lambda u(w_A^H(0,0)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{2}u(w_A^H(0,0))\Big]\right\} \\ & + (1-p_1)\left\{\lambda u(w_A^L(0,0)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{2}u(w_A^L(0,0))\Big]\right\}. \end{aligned}$$

For collusive supervision, (IR_{cs}) and (IC_{cs}) can be rewritten, respectively, as

(IR'_{cs})

$$p_2\left\{\lambda u(w_A^H(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^H(1,1) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\Big]\right\}$$
$$+ (1-p_2)\left\{\lambda u(w_A^L(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00))\Big]\right\} - \varphi \geqslant 0.$$

(IC'_{cs})

$$p_2\left\{\lambda u(w_A^H(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^H(1,1) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\Big]\right\}$$
$$+ (1-p_2)\left\{\lambda u(w_A^L(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00)))\Big]\right\} - \varphi$$
$$\geqslant p_1\left\{\lambda u(w_A^H(11|01) + (1-\lambda)\Big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\Big]\right\}$$
$$+ (1-p_1)\left\{\lambda u(w_A^L(11|01)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00))\Big]\right\}.$$

Let us consider the case when $y = L$ and examine whether collusion would help the principal achieve a lower expected cost. We reach the following result.

Proposition 8. *Given $p_2 \in (0,1)$, with $y = L$, allowing collusion cannot improve the expected cost of the principal.*

Proof. Let us focus on the terms of the IC constraints associated with low output in the collusion-proof contract and in the collusive-supervision contract and compare them to determine which one induces lower payments to the agents.

We first examine the terms associated with low output in (IC'_{cp}). Define

$$X = \lambda(1-p_2)u(w_A^L(1,1)) - \lambda(1-p_1)u(w_A^L(0,0)) - (p_2-p_1)(1-\lambda)\left(\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{2}u(w_A^L(0,0))\right).$$

Since $w_A^L(0,0) = w_A^L(0,1)$ in the collusion-proof contract, we can then rewrite X as

$$X = \lambda(1-p_2)u(w_A^L(1,1)) - \lambda(1-p_1)u(w_A^L(0,1))$$
$$- (p_2-p_1)(1-\lambda)\left(\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(0,1)) + \frac{1}{4}u(w_A^L(0,0))\right).$$

Furthermore, because $u(w_A^L(0,1)) \leqslant u(w_A^L(11|01))$ and $u(w_A^L(0,0)) \leqslant u(w_A^L(11|00))$ in the side contract between the supervisor and the agents, we have the following inequality:

$$X \geqslant \lambda(1-p_2)u(w_A^L(1,1)) - \lambda(1-p_1)u(w_A^L(11|01)) \\ - (p_2-p_1)(1-\lambda)\left(\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00)\right).$$

The right-hand side of the inequality above are the terms associated with low output in (IC'_{cs}), which indicates that the collusion-proof contract induces lower expected payments to the principal when $y = L$. □

Next, we study the case of $y = H$. The result is stated in the following proposition.

Proposition 9. *Given $p_2 \in (0,1)$, define $\lambda^* \equiv \frac{p_2-p_1}{p_2+p_1}$. With $y = H$, there exist unique cutoffs $\underline{\lambda} \in (0, \lambda^*)$ and $\overline{\lambda} \in (\lambda^*, 1)$ such that*

(i) *if $\lambda < \underline{\lambda}$, the no-supervision contract is optimal;*

(ii) *if $\underline{\lambda} \leqslant \lambda \leqslant \overline{\lambda}$, it is beneficial for the principal to allow collusion; and*

(iii) *if $\lambda > \overline{\lambda}$, the collusion-proof contract is optimal.*[11]

Proof. Let us focus on the terms associated with high output in (IC'_{no}), (IC'_{cp}), and (IC'_{cs}). We compare them to determine which one induces lower payments to the agents.

We first consider the comparison between the no-supervision contract and the collusive-supervision contract when $\lambda \in [0, \lambda^*]$. Let us examine the terms associated with high output in (IC'_{cs}), which is denoted by F:

$$\begin{aligned} F(\lambda) &= \lambda p_2 u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(11|01)) \\ &\quad + (p_2-p_1)(1-\lambda)\left(\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) \\ &= \lambda p_2 u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(11|01)) + (p_2-p_1)(1-\lambda)\Big(u(w_A^H(1,1)) \\ &\quad - \frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\Big) \\ &> (p_2-p_1)u(w_A^H(1,1)) + (p_2-p_1)(1-\lambda)\left(\frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) \\ &\quad - \lambda p_1 u(w_A^H(11|01)). \end{aligned} \tag{13}$$

[11] Note that the values of $\underline{\lambda}$ and $\overline{\lambda}$ depend on the specific utility function form of the agents and the bargaining process between the supervisor and the agents.

Given that $w_A^y(11|00) \geqslant w_A^y(11|01)$ and $u(.)$ is increasing and concave, if $\lambda = \lambda^*$, it is easy to check that the following inequality should be true:

$$(p_2 - p_1)(1-\lambda^*)\left(\frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) - \lambda^* p_1 u(w_A^H(11|01)) > 0.$$

This implies that

$$F(\lambda = \lambda^*) > (p_2 - p_1)u(w_A^H(1,1)).$$

The right-hand side of the inequality above is the term associated with high output in the IC constraint (equation (IC'_{no})) of the no-supervision contract, indicating that the payment to the agents in the no-supervision contract is greater than the payment in the collusive contract, $\hat{w}^H > w_A^H(1,1)$, when $\lambda = \lambda^*$.

We then consider the case of $\lambda = 0$. Since $w_A^H(1,1) > w_A^H(11|01)$ and $w_A^H(1,1) > w_A^H(11|00)$,

$$F(\lambda = 0) = (p_2 - p_1)\left(\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) < (p_2 - p_1)u(w_A^H(1,1)).$$

The right-hand side of the inequality above is (IC'_{no}), which indicates that the payment to the agents in the no-supervision contract is less than the payment in the collusive-supervision contract. Hence, when $\lambda = 0$, $\hat{w}^H < w_A^H(1,1)$. Furthermore, it is easy to show that the derivative of F with respect to λ is positive:

$$\begin{aligned}
\frac{\partial F(\lambda)}{\partial \lambda} &= p_2 u(w_A^H(1,1)) - p_1 u(w_A^H(11|01)) \\
&\quad - (p_2 - p_1)\left(\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) \\
&> (p_2 - p_1)u(w_A^H(1,1)) \\
&\quad - (p_2 - p_1)\left(\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) > 0
\end{aligned}$$

Thus, by the continuity of $F(\lambda)$, there must exist a unique cutoff $\underline{\lambda} \in (0, \lambda^*)$ such that if $\lambda = \underline{\lambda}$, $w_A^H(1,1) = \hat{w}^H$.

Next, we compare the collusion-proof contract and the collusive-supervision contract when $\lambda \in [\lambda^*, 1]$. Let us examine the terms associated with high output in(IC'_{cp}), which is denoted by T:

$$T(\lambda) = \lambda p_2 u(w_A^H(1,1)) + (p_2 - p_1)(1-\lambda)\frac{1}{2}u(w_A^H(1,1)).$$

Since $w_A^H(0,1) = w_A^H(0,0) = 0$ in the collusion-proof contract, if $\lambda = \lambda^*$, the following equality should be true.

$$(p_2 - p_1)(1-\lambda^*)\left(\frac{1}{4}u(w_A^H(0,1)) + \frac{1}{4}u(w_A^H(0,0))\right) - \lambda^* p_1 u(w_A^H(0,1)) = 0.$$

This implies

$$\begin{aligned} T(\lambda = \lambda^*) = {} & \lambda^* p_2 u(w_A^H(1,1)) + (p_2 - p_1)(1-\lambda^*)\frac{1}{2}u(w_A^H(1,1)) \\ & + (p_2 - p_1)(1-\lambda^*)\left(\frac{1}{4}u(w_A^H(0,1)) + \frac{1}{4}u(w_A^H(0,0))\right) - \lambda^* p_1 u(w_A^H(0,1)) \\ < {} & \lambda^* p_2 u(w_A^H(1,1)) - \lambda^* p_1 u(w_A^H(11|01)) \\ & + (p_2 - p_1)(1-\lambda^*)\left(\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) \\ = {} & F(\lambda = \lambda^*). \end{aligned}$$

This indicates that the payment to the agents in the collusive-supervision contract is less than the payment in the collusion-proof contract, $w_A^H(1,1) < \tilde{w}^H$, when $\lambda = \lambda^*$.

When $\lambda = 1$, $F(\lambda = 1)$ can be written as follows:

$$F(\lambda = 1) = p_2 u(w_A^H(1,1)) - p_1 u(w_A^H(11|01)) < p_2 u(w_A^H(1,1)) - p_1 u(w_A^H(0,1)) = T(\lambda = 1).$$

The right-hand side of the inequality above are the terms associated with high output in (IC'_{cp}). Thus, the payment to the agents in the collusive-supervision contract is greater than the payment in the collusion-proof contract, $w_A^H(1,1) > \tilde{w}^H$, when $\lambda = 1$. We further check the derivative of T with respect to λ, that is,

$$\frac{\partial T(\lambda)}{\partial \lambda} = (p_2 + p_1)(1-\lambda)\frac{1}{2}u(w_A^H(1,1)) > 0.$$

Since the derivative of F with respect to λ is also positive, the functions T and F will only cross once. Denote the intersection between the two functions by $\overline{\lambda} \in (\lambda^*, 1)$ and we then have $\tilde{w}^H = w_A^H(1,1)$ when $\lambda = \overline{\lambda}$.

Summarizing the analysis above, we conclude that if $\lambda < \underline{\lambda}$, it is optimal to use the no-supervision contract; however, if $\underline{\lambda} \leqslant \lambda \leqslant \overline{\lambda}$, allowing collusion helps the principal lower the expected total payment; if $\overline{\lambda} < \lambda$, the collusion-proof implementation becomes optimal. □