Supervisory Efficiency, Collusion, and Contract Design*

Xiaogang Che† Yangguang Huang‡ Le Zhang§

May 2018

Abstract

We analyze a principal-supervisor-two-agent hierarchy with soft information. The supervisor may be inefficient such that a noisy signal on the agents' effort levels is observed. On the one hand, the agents require risk premiums to work due to the noisy signal. On the other hand, the supervisor and the agents may collude against the principal. We identify a new trade-off between inefficient supervision and supervisor-agent collusion, showing that under certain conditions tolerating collusion to take place helps to "correct" wrong supervisory signals and thus benefits the principal. Furthermore, the characterization of the collusive-supervision contract shows that collusion should be allowed with one agent only.

Keywords: Optimal contract, hierarchy agency, collusion, 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, and other participants for helpful discussions and comments. The article benefited from discussions at CMES (Hong Kong), 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.

†Durham University, UK. E-mail: `xiaogang.che@durham.ac.uk`.

‡Hong Kong University of Science and Technology, Hong Kong. E-mail: `huangyg@ust.hk`.

§Macquarie Graduate School of Management, Australia. E-mail: `lyla.zhang@mgsm.edu.au`.

1 Introduction

In many scenarios, a principal needs to hire multiple agents as a team to complete a task. For example, a manager of a pharmaceutical company may hire a research team that consists of multiple researchers. The manager cannot observe each researcher's contribution but can observe whether the development of a new drug is successful. Having part of the team work hard may be sufficient to achieve the goal; thus, a moral hazard problem arises: Some researchers may free ride on others. To provide proper incentives, an intermediate supervisor may be sent to assess the agents' performances based on some signals.[1] Having this supervisor can be valuable in preventing free riding and incentivizing the whole team.[2] However, it may lead to two problems: First, the supervisor may make mistakes in reviewing the performance of these researchers; second, the supervisor may collude with agents by always reporting that "everybody works hard," which, in turn, mitigates the effectiveness of the supervision.

In this article, we study the contracting problem in such a multiple-agent environment involving both inefficient supervision and collusion. We consider a three-level hierarchy with a principal, a supervisor (she), and two productive agents (he/they). 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 is sent to collect a signal about the effort level of each agent and to report to the principal. Contingent on the realized output level and the supervisory report, the principal pays transfers to the agents and the supervisor accordingly. The signal is both noisy and soft: On the one hand, with inefficient supervisor, the signal can be wrong; on the other hand, the supervisor may coordinate with one agent or both agents to forge a signal that favors them. In this case, is the supervisory information still helpful? Should collusion be allowed or completely prevented? If allowed, then should 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 first derive 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. 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 the principal tailors the payment scheme for an agent, he must account for the linkage between the effort of this agent and the other agent.[3] We show that when the super-

[1] One example is the key performance indicator (KPI). 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.

[3] Interestingly, our results show that regardless of whether collusion is allowed in the contract, the payment to an agent who has worked in production is independent of the other agent's effort level. However, this property does not

visory 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 relevance of supervision in the hierarchy, we explore the role of collusion between the supervisor and the agents. We show that there is a new trade-off between inefficient supervision and collusion. Because supervision technology is imperfect, an incorrect signal may be observed, and therefore, the principal must pay risk premiums to the agents as compensation to induce both agents to work. However, more payments, in turn, generate greater incentives for the agents to shirk and to collude with the supervisor. As a result, inefficient supervision increases the cost of preventing collusion. Technically speaking, as noted in Tirole (1992) and Khalil et al. (2010), satisfying the coalition incentive compatibility constraints raises the cost of satisfying the individual incentive compatibility constraint because they are interlinked.

With this insight, we show that tolerating supervisor-agent collusion makes the principal better off under certain circumstances because collusion helps the principal "correct" wrong supervisory signals: If an agent exerts high effort but the supervisory signal indicates that he shirks, then a collusive supervisor will help report the true effort level. As a result, collusion lowers the cost of providing incentives to risk-averse 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 principal to manipulate his report. The collusive-supervision contract balances the gain and loss from allowing collusion. With this trade-off, we solve the principal's problem again without the coalition incentive compatibility constraints. We find that when a high output is realized, the principal benefits from allowing collusion. More specifically, collusion should be allowed between the supervisor and one agent only, but not with both agents. The collusive-supervision contract dominates both the collusion-proof and the no-supervisor contracts when supervisory efficiency is at an intermediate level. This novel finding provides an explanation of why some managers do not reward workers based on objective performance indicators entirely but on based on subjective evaluations made by intermediate supervisors (MacLeod, 2003).

We further show that as an agent holds less bargaining power in the coalition, the collusive-supervision contract becomes more beneficial to the principal, in contrast to the collusion-proof contract. Moreover, the total cost of the collusive-supervision contract shows a non-monotone relationship with bargaining power: If supervisory efficiency is sufficiently large, then the total payment increases with the bargaining power; otherwise, it decreases with the bargaining power. Furthermore, in Section 7, we provide detailed discussions on (a) why the collusion-proofness principle fails; (b) why such a trade-off between inefficient supervision and collusion does not exist in the single-agent setting; and (c) a comparison with the honest-supervision contract.

always hold for the agent who has shirked during production. See the detailed discussion in Lemmas 1 and 2.

The remainder of the article proceeds as follows: In Section 2, we discuss the related literature. In Sections 3 and 4, we describe the model and provide the no-supervision contract. In Section 5, we characterize the collusion-proof contract and outline the conditions under which soft supervisory information is useful. Section 6 then discusses how allowing collusion improves the incentive provision and characterizes the collusive-supervision contract. We provide further discussions in Section 7 and then conclude in Section 8. All proofs are in the Appendix.

2 Related Literature

The present article 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 which information is hard (i.e., verifiable but concealable) and there is a threat of unobservable side contracting between the supervisor and the agent. One of Tirole's central findings is that information from the corruptible supervisor remains useful for the principal and, moreover, 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. However, some studies show that the principal may be better off allowing a certain scope for collusion between the supervisor and the agent when information is hard (e.g., Che, 1995).

Research on the optimal contract design in organizations with soft information has also received considerable attention. Baliga (1999) notes that the expected cost of a principal with soft information is equivalent to the cost with hard information if the principal takes reports from both the supervisor and the agent, namely, by cross-checking. Grimaud et al. (2003) consider a model in which the supervisory information is soft and the supervisor and the agent collude under asymmetric information, demonstrating that the collusion-proof contract can still be implemented by delegating the task of contracting with the agent to the supervisor. Celik (2009) considers a similar setting and shows that if the supervisor is partly informed of the agent's type, then a centralized organization dominates both no-supervision implementation and delegation.

[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. See extensions of Tirole's model by Cadot (1987); Laffont and Tirole (1991); Kofman and Lawarrée (1993, 1996); Mookherjee and Png (1995); Strausz (1997a,b); Laffont and Martimort (1998, 1999); Olsen and Torsvik (1998); Lambert-Mogiliansky (1998); Dewatripont and Tirole (2005); Vafaï (2010); Liang (2013); Khalil et al. (2013, 2015); Burlando and Motta (2015) for examples.

In a single-agent hierarchy with soft supervisory information, both 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 one of two types: honest and dishonest. To induce truth-telling, the principal must reward both types of supervisors, which can be more costly than allowing collusion. Khalil et al. (2010) consider the case that the supervisor can hide positive information and use this ability 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. Our study reach a qualitatively similar result that collusion becomes "useful" in lowering the principal's total cost in the contract. However, the basic trade-off in our model is completely different; the imperfect supervision technology drives the usefulness of collusion in correcting supervisory signals, especially in a multiple-agent environment.

All the previous studies mentioned above consider collusion in a three-tier hierarchy with a single productive agent. In contrast, our study attempts to provide insights regarding supervisor-agent collusion in a hierarchy with multiple agents.[5] Due to its analytical complexity, studies on this issue are very limited, and the design of the incentive scheme in such multiple-agent organizations remains unclear. Laffont (1990) examines a hidden gaming in which the supervisor can extort an agent by giving a negative report on the agent's performance in a multiple-agent environment, showing that if information is hard, the optimal incentives should be purely personalized; if, however, information is soft, it may be optimal to utilize some of the aggregate (non-personalized) information to design the incentive scheme. In contrast, our study focuses on a similar setting but 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.

3 The Model Setup

Players and actions. We consider a three-level hierarchy with a principal, a supervisor, and two symmetric agents ($i = A, B$). The principal is the owner of a firm and hires two agents as the productive units in the firm. The principal cannot observe the effort levels of the two agents.

Agent i can choose to either work or shirk, denoted as 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. After production, the output $y \in \{H, L\}$

[5] Most studies on the principal-multiple-agents problem focus on the possibility of collusion among the agents and the optimal choice of organizational structure. See, for example, Holmström and Milgrom (1990); Itoh (1993); Laffont and Martimort (1997, 2000); Baliga and Sjostrom (1998); Mookherjee and Tsumagari (2004); Severinov (2008).

is realized and is publicly observed, where H denotes high output, L denotes low output, and $H > L > 0$. We assume that $(H - L)$ 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 implement $e = (1, 1)$ by minimizing the expected total payments to the agents and the supervisor. 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 probabilities, characterized by $1 = p(1, 1) > p(0, 1) > p(0, 0) > 0$.[6] By the symmetry of agents, we have $p(0, 1) = p(1, 0) \equiv p_1$.

Agent i is risk averse with zero reservation utility, and has a separable 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. $u(w_i)$ is concave and satisfies $u(0) = 0$, $u'(\cdot) > 0$ and $u''(\cdot) < 0$. To ensure the existence of equilibrium, we assume that $u(w_i)$ also 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 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 supervision technology is imperfect, which means that the supervisor can be either "efficient" or "inefficient" with probability λ and $(1 - \lambda)$, respectively, where the parameter $\lambda \in [0, 1]$ captures the supervisory efficiency. If the supervisor is efficient, then the signal is accurate, i.e., $\theta = e$. If the supervisor is inefficient, the signal is noisy: Each $\theta \in \Theta$ is randomly observed with an equal probability of $1/4$. When the supervisor observes an incorrect signal, we interpret it here as a mistake by the supervisor. 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. The principal then pays monetary transfers $w_i^y(r)$ to agent i and $s^y(r)$ to the supervisor following the contract.

[6] By assuming $p(1, 1) = 1$, the agents face no uncertainty from the production technology when they both work. This setting allows us to focus on the uncertainty entailed by the supervision technology.

Side contracting. After observing the signal but before reporting to the principal, the supervisor and the agent(s) can choose to collude and manipulate the report without cost. The information for the supervisor-agent coalition is soft in the sense that the outcome of supervision is not verifiable, and by cooperating with the agent(s), the supervisor can make a report $r \neq \theta$. We assume here that the supervisor cannot forge information by herself or that it is too costly to do so without the agents' cooperation.[7] The objective of the supervisor-agent coalition is to forge a report r that maximizes the total payment from the principal. For example, 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 $w_A^H(1, 1) + s^H(1, 1)$. We assume that the supervisor and the agents will not collude when they are indifferent to colluding versus not colluding; moreover, the supervisor reports truthfully if she is indifferent.

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. The final payments to agent i and the supervisor in the coalition are denoted as $w_i^y(r|\theta) \geqslant 0$ and $s^y(r|\theta) \geqslant 0$, respectively. Instead of imposing a particular bargaining process in the side contract, it is sufficient to assume that the bargaining outcome is Pareto efficient and that each party in the coalition can guarantee a payoff no less than that of choosing not to collude. In Section 6.2, for the characterization of the collusive-supervision contract in the hierarchy, we assume that the side contract is implemented through a Nash bargaining game.

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

(1) The principal offers a contract specifying payments $\{w_A^y(r), w_B^y(r), s^y(r)\}$ according to y and 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) After 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 a report r to the principal.
(8) Transfers are paid according to the contract (and the side contract if necessary).

[7] This assumption is the same as that of Kofman and Lawarrée (1996) and Khalil et al. (2010), but Khalil et al. (2010) also consider the possibility that the supervisor can hide the information by claiming that she fails to collect a signal.

4 No-Supervision Contract

In the hierarchy, on the one hand, the supervisor may be inefficient and provide an incorrect signal; on the other hand, 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, $s^y(r) = 0, \forall r, y$ and agents' payments are based solely on output y, i.e., $w_i^y(r) = w_i^y, \forall r$. The three-level hierarchy reduces to a two-level hierarchy, and we refer to this contract as the *no-supervision (no) contract*.

The principal's objective is to implement the effort choice $e = (1,1)$ and to minimize total payments $C_{no} = p(1,1)(w_A^H + w_B^H) + (1 - p(1,1))(w_A^L + w_B^L) = w_A^H + w_B^H$ to the agents. 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 compatible (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 the agents are symmetric in the contract, the principal's cost-minimization problem is as follows:

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

Note that (IR_{no}) is satisfied when (IC_{no}) is satisfied. The solution for the optimization problem (P_{no}) gives the no-supervision contract, characterized 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 given by $\hat{C}_{no} = 2\hat{w}_{no}^H$.

In the absence of supervision, it is optimal for the principal to compensate the agents only when he observes high output.

5 Collusion-Proof Supervision

In this subsection, we study the *collusion-proof (cp) contract*, where the supervisor and the agents do not have incentives to collude but end up playing non-cooperatively. 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 $\theta = (0,1)$, the supervisor colludes with agent A and reports $r = (1,1)$. Let $T^y(r) \equiv w^y_A(r) + w^y_B(r) + s^y(r)$ denote the aggregate transfers made by the principal to the two agents and the supervisor. Let $K^y_i(r) \equiv w^y_i(r) + s^y(r)$ denote the aggregate transfers to agent i and the supervisor. To prevent both the full- and sub-coalitions mentioned above, the principal must ensure that the designed contract satisfies the following coalition incentive compatibility (CIC) constraints:

$$T^y(\theta) \geqslant T^y(r), \text{ and } K^y_i(\theta) \geqslant K^y_i(r), \ \forall\ \theta, r \in \Theta.$$

Given the 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, constraints $K^y_i(\theta) \geqslant K^y_i(r)$ imply that the payments to agent i and the supervisor must be exactly the same across the four signal states:

$$(CIC_s) \qquad K^y_i(1,1) = K^y_i(1,0) = K^y_i(0,1) = K^y_i(0,0), \text{ for } i = A, B.$$

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) $w^y_A(1,0) = w^y_A(1,1)$ *and* $w^y_A(0,1) = w^y_A(0,0)$ *for* $y = L, H$*;*

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

Lemma 1 demonstrates an interesting feature: To fully deter both types of coalitions, the incentive scheme for an agent should depend only on the signal of the agent's effort level. This feature will help us derive the collusion-proof contract in the following. We still take agent A as the representative. Given that the other agent chooses to work ($e_B = 1$) and the supervisor reports truthfully, agent A's IR and IC constraints are as follows:

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

(IC_{cp})

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

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:

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

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 can be stated 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 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 three-level hierarchy reduces to a standard principal-agent hierarchy.

To prevent collusion in which agents shirk and the supervisor reports positive signals on their effort levels, (IC_{cp}), (CIC_f), and (CIC_s) should 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 supervisory signal. This potential mistake by the supervisor creates a conflict between truthful reporting and incentive provision. In other words, although deterring collusion leads to a benefit from truthful supervisory information, it entails the cost of punishing a working agent due to incorrect supervisory signals. As a result, both agents require risk premiums to work, and it becomes costly to implement the collusion-proof contract.

6 Collusive Supervision

As discussed above, because of the trade-off between inefficient supervision and collusion, it becomes costly to implement the collusion-proof contract. Thus, it would be interesting to examine whether there is an improvement over the no-supervision and the collusion-proof contracts. In particular, we focus on whether the principal is better off tolerating the occurrence of collusion, and in the following analysis we refer to such a contract as the *collusive-supervision (cs) contract*.

6.1 Incentive improvement by allowing collusion

When collusion is allowed to occur, 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. As defined above, payoffs as a result of collusion between agent i and the supervisor are denoted by $w_i^y(r|\theta)$ and $s^y(r|\theta)$, where y and θ are realized and the supervisor reports r.

To make the analysis tractable, we first impose a restriction on the payment scheme: Agent i's reward is non-decreasing in the reported signal for him, given the reported signal of the other agent. That is, for $y = L, H$,

$$w_A^y(1,1) \geqslant w_A^y(0,1) \text{ and } w_A^y(1,0) \geqslant w_A^y(0,0) \\ w_B^y(1,1) \geqslant w_B^y(1,0) \text{ and } w_B^y(0,1) \geqslant w_B^y(0,0) \tag{1}$$

Under this restriction, the supervisor will make only an "upward" adjustment of the signal instead of a "downward" adjustment. Clearly, when the supervisory signal is $(0,1)$, the supervisor may collude with agent A and reports $(1,1)$. Then, from agent A's perspective, the IC constraint is

(IC'_{cs})

$$\lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{4}\big[u(w_A^H(1,1)) + u(w_A^H(?|10)) + u(w_A^H(11|01)) + u(w_A^H(?|00))\big] - \varphi$$
$$\geqslant p_1\Big\{\lambda u(w_A^H(11|01)) + (1-\lambda)\frac{1}{4}\big[u(w_A^H(1,1)) + u(w_A^H(?|10)) + u(w_A^H(11|01)) + u(w_A^H(?|00))\big]\Big\}$$
$$+ (1-p_1)\Big\{\lambda u(w_A^L(11|01)) + (1-\lambda)\frac{1}{4}\big[u(w_A^L(1,1)) + u(w_A^L(?|10)) + u(w_A^L(11|01)) + u(w_A^L(?|00))\big]\Big\},$$

The question marks in (IC'_{cs}) represent two issues that require further discussion: First, when the signal is $(1,0)$, does agent A still have the incentive to collude with the agent and report $(1,1)$? Second, when the signal is $(0,0)$, will the supervisor collude with both agents and report $(1,1)$, or will she collude with only one agent and report $(1,0)$ (or $(0,1)$)? To answer these questions, we derive the following lemma regarding the features of the payment scheme under collusive supervision:

Lemma 2. *When supervisor-agent collusion is allowed, payments to the agents have the following features: For $y = L, H$,*

(a) $w_A^y(1,1) = w_A^y(1,0)$, *and* $w_B^y(0,1) = w_B^y(1,1)$

(b) $w_A^y(11|00) \geqslant w_A^y(10|00)$, $s^y(11|00) \geqslant s^y(10|00)$, *and* $w_B^y(11|00) \geqslant w_B^y(1,0)$, *and* $w_B^y(11|00) \geqslant w_B^y(01|00)$, $s^y(11|00) \geqslant s^y(01|00)$, *and* $w_A^y(11|00) \geqslant w_A^y(0,1)$.

(c) $w_A^y(0,1) \leqslant w_A^y(0,0)$, *and* $w_B^y(1,0) \leqslant w_B^y(0,0)$.

Because $w_A^y(1,1) = w_A^y(1,0)$, when the signal is $(1,0)$, the supervisor has no incentive to collude with agent A. More importantly, given part (b) of Lemma 2, a full-coalition is always preferred to a sub-coalition when the signal is $(0,0)$.[8] Part (c) indicates that when a negative signal about an agent is reported, the agent would be punished if the other agent works. This result is consistent with that of Laffont (1990): When supervisory information is soft, the reward schemes should be based on the "aggregate information" of both agents' performances.

[8] Intuitively, suppose that agents A and B each obtain w under a favorable report and 0 otherwise. Provided that they equally divide the spoils, the supervisor and the colluding agent each obtain $\frac{1}{2}w$ under a sub-coalition; the supervisor and two agents each obtain $\frac{2}{3}w$ under a full-coalition.

With Lemma 2, the IR and IC constraints can then be stated as follows:

(IR_{cs})

$$\lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{4}\big[u(w_A^H(1,1)) + u(w_A^H(1,0)) + u(w_A^H(11|01)) + u(w_A^H(11|00))\big] - \varphi \geqslant 0.$$

(IC_{cs})

$$\begin{aligned}
&\lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{4}\big[u(w_A^H(1,1)) + u(w_A^H(1,0)) + u(w_A^H(11|01)) + u(w_A^H(11|00))\big] - \varphi \\
&\geqslant p_1\Big\{\lambda u(w_A^H(11|01)) + (1-\lambda)\frac{1}{4}\big[u(w_A^H(1,1)) + u(w_A^H(1,0)) + u(w_A^H(11|01)) + u(w_A^H(11|00))\big]\Big\} \\
&+ (1-p_1)\Big\{\lambda u(w_A^L(11|01)) + (1-\lambda)\frac{1}{4}\big[u(w_A^L(1,1)) + u(w_A^L(1,0)) + u(w_A^L(11|01)) + u(w_A^L(11|00))\big]\Big\}.
\end{aligned}$$

By comparing (IC_{cp}) and (IC_{cs}), we observe that $u(w_A^H(0,1))$ and $u(w_A^H(0,0))$ on the left-hand side of (IC_{cp}) are now replaced by $u(w_A^H(11|01))$ and $u(w_A^H(11|00))$. This captures the potential benefit of allowing collusion; that is, given $y = H$, when agent A works but the signal is incorrect, the corruptible supervisor helps correct the signal. This correction increases the agent's incentive to work. However, as shown on the right-hand side of (IC_{cs}), any negative supervisory report can be manipulated, which, in turn, induces agent A to have a higher incentive to shirk. Given this trade-off, we attempt to identify whether and under what conditions allowing supervisor-agent collusion would generate higher incentives for the agents to work. To do so, we first compare the terms in (IC_{no}) and (IC_{cs}) that are associated with low output.

Proposition 3. *For $y = L$, allowing supervisor-agent collusion cannot improve the agents' incentives to work.*

We then compare the terms associated with high output, which gives the following result:

Proposition 4. *For $y = H$, we have the following:*

(a) *A full-coalition cannot improve the agents' incentives to work. Moreover, it is optimal to reward the same payment to the agents with signals $(0,0)$ and $(1,1)$.*

(b) *There are unique cutoffs $\underline{\lambda} \in (0, \lambda^*)$ and $\overline{\lambda} \in (\lambda^*, 1)$ such that if $\underline{\lambda} < \lambda < \overline{\lambda}$, allowing a sub-coalition induces higher incentives for the agents to work.*

As discussed previously, because of the interlink between the IC and the CIC constraints, preventing supervisor-agent collusion raises the cost of preventing the agents from shirking. One way to restore some variation in the agents' compensation is to allow collusion to take place. Interestingly, Proposition 4 demonstrates that when λ is in a certain intermediate range, the supervisor-agent collusion in a sub-coalition leads to an improvement in the agents' incentives to work over the collusion-proof and no-supervision contracts. Another interesting feature is that the full-coalition should be fully prevented and, moreover, the same payment should be rewarded to the agent across the signals of $(0,0)$ and $(1,1)$. The reason is that when the IC constraint is satisfied in equilibrium, both agents will choose to work. If the supervisor reports $(0,0)$, it implies that an incorrect signal is observed, and therefore, instead of allowing collusion, continuing to reward the agents as though they were working provides higher incentives to the agents.

Note that the values of $\underline{\lambda}$ and $\overline{\lambda}$ depend on the specific utility function of the agents and the bargaining process between the supervisor and the agent(s) in the side contract. In the following subsection we characterize $\underline{\lambda}$, $\overline{\lambda}$, the collusive-supervision contract, and the principal's optimal contract choice over the contracts by imposing Nash bargaining in the side contract.

6.2 Characterization of the collusive-supervision contract

With the insight above, we now characterize the collusive-supervision contract and compare it with the no-supervision contract and the collusion-proof contract. From Proposition 4, we know that collusion-proofness holds between the signals of $(1,1)$ and $(0,0)$, and that the supervisor colludes with agent A (respectively, agent B) to report $(1,1)$ when the signal θ is $(0,1)$ (respectively, $(1,0)$). Taking agent A as the representative, let us use α and $1-\alpha$ to denote the bargaining powers of agent A and the supervisor in the sub-coalition, where $\alpha \in (0,1)$. The side contract specifies payments $w_A^H(11|01)$ and $s^H(11|01)$ when the signal is $(0,1)$ and the supervisor reports $(1,1)$. The corresponding Nash bargaining problem that determines $w_A^H(11|01)$ and $s^H(11|01)$ is

$$
\begin{aligned}
&\text{Max}\big[u(w_A^H(11|01)) - u(w_A^H(0,1))\big]^{\alpha}\big[s^H(11|01) - s^H(0,1)\big]^{1-\alpha}, \\
&\text{subject to } \; w_A^H(11|01) + s^H(11|01) = w_A^H(1,1) + s^H(1,1).
\end{aligned} \tag{2}
$$

The principal's minimization problem is as follows:

$$
(P_{cs}) \quad
\begin{aligned}
\text{Min } C_{cs} &= \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\big[T^H(1,1) + T^H(11|10) + T^H(11|01) + T^H(11|00)\big] \\
&= 2w_A^H(1,1) + s^H(1,1), \\
\text{subject to } &(IR_{cs}),\ (IC_{cs}), \\
&w_i^y(r) \geqslant 0,\ w_i^y(r|\theta) \geqslant 0,\ s^y(r) \geqslant 0,\ s^y(r|\theta) \geqslant 0, \forall\, r \in \Theta,\ \theta \in \Theta,\ i = A, B.
\end{aligned}
$$

Solving (P_{cs}) gives us the collusive-supervision contract:

Proposition 5. *The collusive-supervision contract has the following features:*

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

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

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

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

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

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

Proposition 5 indicates that the principal pays $\check{w}^H_{cs}$ to each agent and 0 to the supervisor with the signals of $(1,1)$ or $(0,0)$, which means that the full-coalition is fully prevented. However, when the signal is $(0,1)$ ($(1,0)$), the supervisor is allowed to collude with agent A (agent B) to report $(1,1)$ and then share the payment $\check{w}^H_{cs}$ from the principal.

6.3 Comparisons across contracts

After characterizing the no-supervision, collusion-proof, and collusive-supervision contracts, Propositions 1, 2, and 5 show that the total payment is zero when $y = L$ and $\hat{C}_{no} = 2\hat{w}^H_{no}$, $\tilde{C}_{cp} = 2\tilde{w}^H_{cp}$, and $\check{C}_{cs} = 2\check{w}^H_{cs}$, respectively, when $y = H$. We are now ready to examine the principal's choice of the optimal contract. Fixing the same wage across contracts, i.e., $C = 2w^H(1,1)$, we then need to focus only on the equilibrium IC constraints, ($\widehat{IC}_{no}$), ($\widetilde{IC}_{cp}$), and ($\widetilde{IC}_{cs}$).

We first compare the left-hand sides of ($\widetilde{IC}_{cs}$) and ($\widehat{IC}_{no}$). Define their difference as

$$\begin{aligned}\xi(\lambda) \equiv\ & \lambda\big[u(w^H(1,1)) - p_1 u(\alpha w^H(1,1))\big] + (1-\lambda)(1-p_1)\big[\frac{3}{4}u(w^H(1,1)) + \frac{1}{4}u(\alpha w^H(1,1))\big] \\ & - (1-p_1)u(w^H(1,1)), \\ =\ & \big[\lambda p_1 - \frac{1}{4}(1-\lambda)(1-p_1)\big]\big[u(w^H(1,1)) - u(\alpha w^H(1,1))\big].\end{aligned}$$

Because $\big[u(w^H(1,1)) - u(\alpha w^H(1,1))\big] > 0$, $\xi(\lambda) \geqslant 0$ implies that

$$(3) \qquad \lambda p_1 - \frac{1}{4}(1-\lambda)(1-p_1) \geqslant 0 \Longleftrightarrow \lambda \geqslant \frac{1-p_1}{1+3p_1} \equiv \underline{\lambda}.$$

When $\lambda = \underline{\lambda}$, the principal incurs the same cost across the no-supervision and collusive-supervision contracts, i.e., $\hat{w}_{no}^H = \breve{w}_{cs}^H$. If $\lambda > \underline{\lambda}$, the collusive-supervision contract is less costly than the no-supervision contract.

Next, we compare ($\breve{\widetilde{IC}}_{cs}$) and ($\widetilde{IC}_{cp}$) and define $\phi(\lambda)$ as the difference between their left-hand sides:

$$\begin{aligned}\phi(\lambda) &\equiv \lambda\big[u(w^H(1,1)) - p_1 u(\alpha w^H(1,1))\big] + (1-\lambda)(1-p_1)\big[\frac{3}{4}u(w^H(1,1)) + \frac{1}{4}u(\alpha w^H(1,1))\big] \\ &\quad - \big[\lambda u(w^H(1,1)) + \frac{1}{2}(1-p_1)(1-\lambda)u(w^H(1,1))\big], \\ &= \frac{1}{4}(1-\lambda)(1-p_1)u(w^H(1,1)) + \big[\frac{1}{4}(1-\lambda)(1-p_1) - \lambda p_1\big]u(\alpha w^H(1,1)).\end{aligned}$$

Having $\phi(\lambda) \geqslant 0$ implies that

$$\lambda \leqslant \frac{(1-p_1)[u(w^H(1,1)) + u(\alpha w^H(1,1))]}{[(1-p_1)u(w^H(1,1)) + (1+3p_1)u(\alpha w^H(1,1))]} \equiv \overline{\lambda}. \tag{4}$$

When $\lambda = \overline{\lambda}$, the collusion-proof and collusive-supervision contracts incur the same cost, i.e., $\hat{w}_{cp}^H = \breve{w}_{cs}^H$. If $\lambda < \overline{\lambda}$, the collusive-supervision contract is less costly than the collusion-proof contract.

In summary, the principal's optimal choice of contract is as follows:

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 $\lambda \geqslant \overline{\lambda}$.*

Example 1. We now provide a numerical example to illustrate the dominance of the collusive-supervision contract. Suppose that $u(w) = w$, $\varphi = 1$, $p_1 = 1/2$, $\lambda = 8/21$, and $\alpha = 1/2$. In the no-supervisor contract, plugging $p_1 = 1/2$ and $\varphi = 1$ into ($\widehat{IC}_{no}$) gives $\hat{w}_{no}^H = 2$, and thus, $\hat{C}_{no} = 4$. Recall Proposition 2, $\lambda^* = 1/3 < \lambda = 8/21$. From ($\widetilde{IC}_{cp}$), we have $\tilde{w}_{cp}^H = 28/15$, and $\tilde{C}_{cp} \approx 3.733$ in the collusion-proof contract. Because $\tilde{C}_{cp} < \hat{C}_{no}$, the principal benefits from hiring a supervisor. Equations (3) and (4) give us $\underline{\lambda} = 1/5$ and $\overline{\lambda} = 3/7$, respectively, and $\underline{\lambda} < \lambda = 8/21 < \overline{\lambda}$. In the collusive-supervision contract, it is easy to show from ($\breve{\widetilde{IC}}_{cs}$) that $\breve{w}_{cs}^H = 336/187$ and $\breve{C}_{cs} \approx 3.594$. Because $\breve{C}_{cs} < \tilde{C}_{cp}$, the collusive-supervision contract makes the principal better off.

In the analyses above, we show that allowing a sub-coalition would benefit the principal and, moreover, the bargaining power in the side contract becomes relevant in the design of the collusive-supervision contract. It would be interesting to observe how the bargaining power (in particular, taking the limits) affects the principal's contract choice and total payment in equilibrium. Note that $\underline{\lambda}$ is independent of α and is strictly lower than λ^*. However, given different levels of bargaining power, $\overline{\lambda}$ takes different values between λ^* and 1.

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 any $\lambda \in [\lambda^*, 1]$, 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) *If $\lambda > \underline{\lambda}$, $\check{C}_{cs}$ is increasing in α; otherwise, $\check{C}_{cs}$ is decreasing in α.*

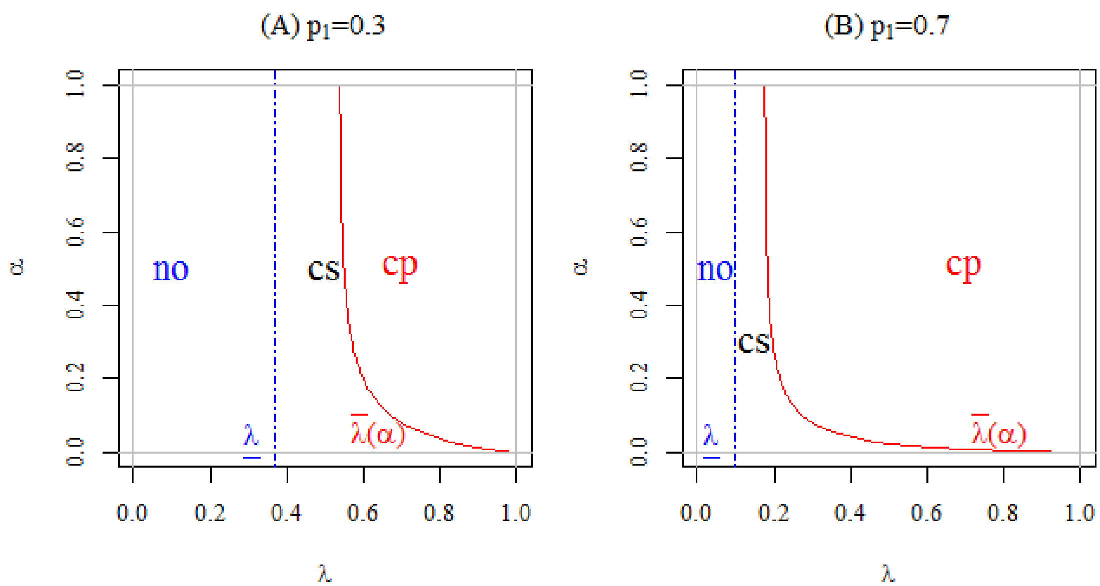

Figure 1: Illustration of the Optimal Contract Choice

Example 2. Figure 1 illustrates how $\overline{\lambda}$ and the principal's choice change in bargaining power α with a constant relative risk aversion (CRRA) utility function $u(w) = \sqrt{w}$ and parameter values of $p_1 \in \{0.3, 0.7\}$. We compute $\underline{\lambda}$ and $\overline{\lambda}$ by (3) and (4), respectively. The no-supervisor 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. $\overline{\lambda}$ is decreasing in α and converges to 1 when the supervisor has all the bargaining power in the side contract ($\alpha \to 0$). In addition, the dominant regions of the no-supervision contract and the collusive-supervisor 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.

7 Discussion

7.1 Failure of the collusion-proofness principle

Instead of allowing collusion, can the principal directly offer a contract with the *ex post* payments of the side contract and prevent collusion? The answer is No. To see this, let us consider a contract in which the *ex post* payments in the side contract are offered directly by the principal. In this contract, the payments with signals of $(0, 0)$ and $(1, 1)$ are the same as those in the collusive-supervision contract (characterized in Proposition 5), but the payments to the supervisor and the agent under the signals of $(1, 0)$ and $(0, 1)$ are replaced by the *ex post* bargaining payments, $\alpha \breve{w}_{cs}^H$ and $(1-\alpha)\breve{w}_{cs}^H$, respectively. The payment structure is as follows:

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

However, this payment structure results in a new collusion problem: Because $\breve{w}_{cs}^H + 0 < \breve{w}_{cs}^H + (1-\alpha)\breve{w}_{cs}^H$ for all $\alpha \in (0, 1)$, when the supervisor observes a signal $(0, 0)$, she has an incentive to collude with agent A and report $(1, 0)$. That is, by modifying the report for agent A, the supervisor extracts $(1-\alpha)\breve{w}_{cs}^H$ from agent B. Therefore, the collusion-proofness principle fails.

As mentioned in the introduction, bribery is a penalty of shirking because the agent has to pay it when getting a low signal. However, the principal cannot directly use it to provide incentives due to the existence of multiple agents. To use it, the principal has to rely on the collusion between the supervisor and the agent. This result echoes that of Khalil et al. (2010), where the possibility of extortion prevents the principal from using bribery to provide incentives.

7.2 Collusive supervision in a single-agent hierarchy

The analysis above shows 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, one could ask why such a trade-off does not exist in the single-agent setting. In this subsection, we examine the setting with one agent and show that the collusive-supervision contract generates the same cost as in 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. However, if the agent shirks, the probability of $y = H$ is $p_1 \in (0, 1)$. 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:

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

which can be rewritten as

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

Clearly, to provide an incentive for the agent to work, it is optimal for the principal not to reward the agent when $y = L$. Furthermore, the principal should reward $w^H(1|0) = w^H(1)$ (and 0 to the supervisor). The intuition is straightforward. When the agent chooses to work along the equilibrium path (in which the (IC_s) constraint is satisfied), if the reported signal is 0, it must be the supervisor's mistake. The equation can then be simplified as $(1 - p_1)u(w^H(1)) \geqslant \varphi$, which is the same as (IC_{no}). Therefore, the collusive-supervision contract is equivalent to the no-supervision contract. In other words, soft supervisory information is not useful in the single-agent case, as shown in Tirole (1986).

7.3 Honest supervision

A seemingly ideal contracting environment is that in which the supervisor is honest and always reports the observed signal truthfully ($r = \theta$). 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 the contract the *collusion-free (cf) contract*. The payment structure of the collusion-free contract is stated as follows. The principal pays zero to both agents when $y = L$. When $y = H$, each agent is paid by $\tilde{w}^H_{cf}$ if the supervisory signal is positive, regardless of the signal on the other agent. By the equilibrium IC constraint, $\tilde{w}_{cf}^H$ is given by

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

which is the same as ($\widetilde{IC}_{cp}$), and thus, $\tilde{w}_{cf}^H = \tilde{w}_{cp}^H$. 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 $\tilde{C}_{cf} = 2\lambda\tilde{w}_{cf}^H + (1-\lambda)\tilde{w}_{cf}^H = (1+\lambda)\tilde{w}_{cf}^H$, which is clearly smaller than $\tilde{C}_{cp} = 2\tilde{w}_{cp}^H$. 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. We provide numerical examples below to show that although the principal no longer needs to worry about the supervisor-agent collusion, having an honest but inaccurate supervisor may be more costly than having a corruptible and equally inaccurate supervisor. Which scenario is better for the principal depends on the parameter values and the form of the utility function.

Example 3. Let $u(w) = \sqrt{w}$, $\varphi = 1$, $p_1 = 1/2$, $\lambda = 0.1$, and $\alpha = 1/2$. From ($\widehat{IC}_{no}$), $\hat{w}_{no}^H = 4$ and $\hat{C}_{no} = 8$. From ($\widetilde{IC}_{cp}$), $\tilde{w}_{cf}^H = \tilde{w}_{cp}^H \approx 9.467$, $\tilde{C}_{cp} \approx 18.935$ and $\tilde{C}_{cf} \approx 10.414$. From ($\widetilde{IC}_{cs}$), we have $\check{w}_{cs}^H \approx 4.310$, 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.

8 Conclusion

In this article, we study a principal-supervisor-two-agent hierarchy in which the supervisory signal is both noisy and soft. 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 inefficient supervision and collusion. When the supervisory efficiency is at an intermediate level, allowing a sub-coalition permits the supervisor to “correct” wrong supervisory signals and saves the cost of incentive provision. Our results provide a justification for rewarding workers based on subjective evaluation made by intermediate supervisors when objective performance indicator are not fully accurate.

Appendix

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

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

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

plus the complementary slackness conditions for the constraints.

Step I. It is impossible to have $\delta = 0$ because it implies that $w_A^H = 0$ in (A1) and $w_A^L = 0$ in (A2), which violates (IC_{no}), yielding a contradiction. We therefore have $\delta > 0$ and $w_A^H > 0$, which further give $\partial L/\partial w_A^H = 0$ and $\delta = 2/[(1-p_1)u'(w_A^H)]$.

Step II. Given $\delta > 0$, (A2) indicates that $\partial L/\partial w_A^L > 0$ and $w_A^L = 0$, which indicates that $\hat{w}^L = 0$ in the no-supervision contract.

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

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

Further, the total payment of the principal is given by $\hat{C}_{no} = 2\hat{w}_{no}^H$. □

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

$$w_A^y(1,1) + \underbrace{w_B^y(1,1) + s^y(1,1)}_{K_B^y(1,1)} = w_A^y(1,0) + \underbrace{w_B^y(1,0) + s^y(1,0)}_{K_B^y(1,0)}.$$

Further, (CIC_s) indicates that $K_B^y(1,1) = K_B^y(1,0)$. 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,

$$w_A^y(0,1) + \underbrace{w_B^y(0,1) + s^y(0,1)}_{K_B^y(0,1)} = w_A^y(0,0) + \underbrace{w_B^y(0,0) + s^y(0,0)}_{K_B^y(0,0)}.$$

Again, (CIC_s) indicates that $K_B^y(0,1) = K_B^y(0,0)$. Hence, we have $w_A^y(0,0) = w_A^y(0,1)$.

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

$$\begin{aligned}
\text{(B1)}: \quad & \frac{\partial L}{\partial w_A^H(1,1)} = 2 - \delta\big[\lambda + \frac{1}{2}(1-\lambda)(1-p_1)\big]u'(w_A^H(1,1)) \geqslant 0, \\
& w_A^H(1,1) \geqslant 0 \ \text{ and } \ w_A^H(1,1)\frac{\partial L}{\partial w_A^H(1,1)} = 0; \\
\text{(B2)}: \quad & \frac{\partial L}{\partial w_A^L(1,1)} = \delta\big[\frac{1}{2}(1-\lambda)(1-p_1)\big]u'(w_A^L(1,1)) \geqslant 0, \\
& w_A^L(1,1) \geqslant 0 \ \text{ and } w_A^L(1,1)\frac{\partial L}{\partial w_A^L(1,1)} = 0; \\
\text{(B3)}: \quad & \frac{\partial L}{\partial w_A^H(0,0)} = -\delta\big[\frac{1}{2}(1-\lambda)(1-p_1) - \lambda p_1\big]u'(w_A^H(0,0)) \geqslant 0, \\
& w_A^H(0,0) \geqslant 0 \ \text{ and } w_A^H(0,0)\frac{\partial L}{\partial w_A^H(0,0)} = 0; \\
\text{(B4)}: \quad & \frac{\partial L}{\partial w_A^L(0,0)} = \delta\big[\lambda(1-p_1) + \frac{1}{2}(1-p_1)(1-\lambda)\big]u'(w_A^L(0,0)) \geqslant 0, \\
& w_A^L(0,0) \geqslant 0 \ \text{ and } w_A^L(0,0)\frac{\partial L}{\partial w_A^L(0,0)} = 0; \\
\text{(B5)}: \quad & \frac{\partial L}{\partial s^H(1,1)} = 1 \geqslant 0, \ s^H(1,1) \geqslant 0 \ \text{ and } \ s^H(1,1)\frac{\partial L}{\partial s^H(1,1)} = 0;
\end{aligned}$$

plus the complementary slackness conditions for the constraints.

Step I. It is impossible to have $\delta = 0$ because it implies that $\partial L/\partial w_A^H(1,1) = 2 > 0$ and $w_A^H(1,1) = 0$ in (B1), which violates the IC$_{cp}$ constraint, yielding a contradiction. Therefore, we should have $\delta > 0$ and $w_A^H(1,1) > 0$, implying that $\partial L/\partial w_A^H(1,1) = 0$ and $\delta = \frac{2}{(\lambda+\frac{1}{2}(1-\lambda)(1-p_1))u'(w_A^H(1,1))}$. Further, by Lemma 1, we have $w_A^L(1,1) = w_A^L(1,0) > 0$.

Step II. Given $\delta > 0$, (B2) and (B4) show that $\partial L/\partial w_A^L(1,1) > 0$ and $\partial L/\partial w_A^L(0,0) > 0$, respectively, given that $w_A^L(1,1) = w_A^L(0,0) = 0$. By Lemma 1, we have $w_A^L(1,0) = w_A^L(0,1) = 0$. Further, 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 III. From (B5), we clearly have $s^H(1,1) = 0$.

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

$$\frac{1}{2}(1-\lambda^*)(1-p_1) - \lambda^* p_1 = 0.$$[9]

This equation gives the following two cases: $\lambda < \lambda^*$ and $\lambda \geqslant \lambda^*$.

(a) When $\lambda < \lambda^*$, $\partial L/\partial w_A^H(0,0)$ 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 $\lambda \geqslant \lambda^*$, $\partial L/\partial w_A^H(0,0)$ is strictly positive, which implies that $w_A^H(0,0) = 0$. By Lemma 1, we also have $w_A^H(0,1) = 0$. Further, 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 $\tilde{C}_{cp} = 2\tilde{w}_{cp}^H$. □

Proof of Lemma 2. We show the lemma from the perspective of agent A. All results hold for agent B under a symmetric setting.

Part (a): If $w_A^y(1,1) > w_A^y(1,0)$, we can see that reducing $w_A^y(1,1)$ to $w_A^y(1,0)$ will not affect agent A's decision to work because agent A cannot directly affect the signal of agent B. Thus, the principal can reduce $w_A^y(1,1)$ to $w_A^y(1,0)$. If $w_A^y(1,1) < w_A^y(1,0)$, the same argument applies. Therefore, in the optimal contract, the principal will set $w_A^y(1,1) = w_A^y(1,0)$.

[9] This condition is also the one used to compare ($\widehat{IC}_{no}$) and ($\widetilde{IC}_{cp}$).

Part (b): Neither the IC nor IR constraints contains direct payments to the supervisor, and paying a positive amount to the supervisor does not improve the agents' incentive to work. As (CIC_f) and (CIC_s) constraints are removed, the principal will set $s^y(r) = 0$ for all r. So we have $s^y(1,1) = s^y(1,0)$. From part (a), we have $w_A^y(1,1) = w_A^y(1,0)$. We have $w_B^y(1,1) \geqslant w_B^y(1,0)$ by restriction (1).

With these three conditions, we have

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

In a full-coalition,

$$w_A^y(1,1) + w_B^y(1,1) + s^y(1,1) = w_A^y(11|00) + w_B^y(11|00) + s^y(11|00).$$

In a sub-coalition (with agent A), $w_A^y(1,0) + s^y(1,0) = w_A^y(10|00) + s^y(10|00)$. Adding $w_B^y(1,0)$ on both sides,

$$w_A^y(1,0) + w_B^y(1,0) + s^y(1,0) = w_A^y(10|00) + w_B^y(1,0) + s^y(10|00).$$

Therefore,

$$w_A^y(11|00) + w_B^y(11|00) + s^y(11|00) \geqslant w_A^y(10|00) + w_B^y(1,0) + s^y(10|00).$$

Under efficient bargaining, $w_A^y(11|00) \geqslant w_A^y(10|00)$, $w_B^y(11|00) \geqslant w_B^y(1,0)$, and $s^y(11|00) \geqslant s^y(10|00)$. Therefore, given $\theta = (0,0)$, all three parties prefer reporting $r = (1,1)$ to reporting $r = (1,0)$.

Part (c): If $w_A^y(0,1) > w_A^y(0,0)$, this relation implies that $w_A^y(11|01) > w_A^y(11|00)$ in the side contract. Then, the payoff for agent A in the sub-coalition is greater than the payoff in the full-coalition, which contradicts part (b). Therefore, we should have $w_A^y(0,1) \leqslant w_A^y(0,0)$. □

Proof of Proposition 3. Let us focus on the terms associated with low output in (IC_{cp}) and (IC_{cs}) and compare them to determine which one induces lower payments to the agents. In the collusion-proof contract, the expected payment when $y = L$ is

$$X \equiv -\lambda(1-p_1)u(w_A^L(0,0)) - (1-p_1)(1-\lambda)\big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{2}u(w_A^L(0,0))\big].$$

From Lemma 1, $w_A^L(0,0) = w_A^L(0,1)$ in the collusion-proof contract, and thus, we can rewrite X as

$$X = -\lambda(1-p_1)u(w_A^L(0,1)) - (1-p_1)(1-\lambda)\big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(0,1)) + \frac{1}{4}u(w_A^L(0,0))\big].$$

Further, given that $u(w_A^L(0,1)) < u(w_A^L(11|01))$ and $u(w_A^L(0,0)) < u(w_A^L(11|00))$ in the side contract between the supervisor and the agents, the following inequality should always hold:

$$X > -\lambda(1-p_1)u(w_A^L(11|01)) - (1-p_1)(1-\lambda)\big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00))\big].$$

The right-hand side of the inequality above contains terms associated with low output in (IC_{cs}), indicating that allowing collusion cannot improve upon the collusion-proof implementation when $y = L$. □

Proof of Proposition 4. Part (a): From Lemma 2, (IC_{cs}) can be rewritten as follows (denoted by Z):

$$\begin{aligned} Z \equiv\ & \lambda u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(11|01)) \\ & + (1-p_1)(1-\lambda)\big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\big] \\ & - (1-p_1)\Big\{\lambda u(w_A^L(11|01)) + (1-\lambda)\big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00))\big]\Big\} - \varphi. \end{aligned}$$

Differentiating Z with respect to $w_A^H(11|00)$ yields

$$\frac{\partial Z}{\partial w_A^H(11|00)} = (1-p_1)(1-\lambda)\frac{1}{4}u'(w_A^H(11|00)) > 0.$$

Clearly, Z is maximized if $w_A^H(11|00) = w_A^H(1,1)$. In other words, with $y = H$ and $\theta = (0,0)$, instead of allowing the full-coalition, rewarding the agents directly as though they were working provides higher incentives to the agents to work. Thus, it is optimal for the principal to set $w_A^H(0,0) = w_A^H(1,1)$, indicating that the full-coalition should be prevented.

Part (b): Let us focus on the terms associated with high output in (IC_{no}), (IC_{cp}), and (IC_{cs}), and 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 $\lambda \in [0, \lambda^*]$. Define F as the terms associated with high output in (IC_{cs}):

$$\begin{aligned}
F(\lambda) &\equiv \lambda u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(11|01)) \\
&\quad + (1-p_1)(1-\lambda)\big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\big] \\
&= \lambda u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(11|01)) + (1-p_1)(1-\lambda)\big[u(w_A^H(1,1)) \\
&\quad - \frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\big] \\
&> (1-p_1)u(w_A^H(1,1)) + (1-p_1)(1-\lambda)\big[\frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\big] \\
&\quad - \lambda p_1 u(w_A^H(11|01)).
\end{aligned}$$

Given that $w_A^H(11|00) = w_A^H(1,1) > w_A^H(11|01)$ and $u(\cdot)$ is increasing and concave, if $\lambda = \lambda^*$, it is easy to check that the following inequality holds:

$$(1-p_1)(1-\lambda^*)\big[\frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\big] - \lambda^* p_1 u(w_A^H(11|01)) > 0,$$

indicating that

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

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

Next, we take $\lambda = 0$. Because $w_A^H(1,1) > w_A^H(11|01)$ and $w_A^H(1,1) = w_A^H(11|00)$, we have

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

The right-hand side of the inequality above is the term associated with high output in (IC_{no}), indicating that the payment to the agents in the no-supervision contract is less than the payment in the collusive-supervision contract, $\hat{w}_{no}^H < w_A^H(1,1)$, when $\lambda = 0$. Further, it is easy to show that the derivative of F with respect to λ is positive;

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

By the continuity of $F(\lambda)$, there is a unique cutoff $\underline{\lambda} \in (0, \lambda^*)$ such that if $\lambda = \underline{\lambda}$, $w_A^H(1,1) = \hat{w}_{no}^H$.

Second, we consider the comparison between the collusion-proof contract and the collusive-supervision contract when $\lambda \in [\lambda^*, 1]$. Define G as the terms associated with high output in (IC_{cp}):

$$G(\lambda) \equiv \lambda u(w_A^H(1,1)) + (1-p_1)(1-\lambda)\frac{1}{2}u(w_A^H(1,1)).$$

Because $w_A^H(0,1) = w_A^H(0,0) = 0$ in the collusion-proof contract, if $\lambda = \lambda^*$, the following equality holds:

$$(1-p_1)(1-\lambda^*)\Big[\frac{1}{4}u(w_A^H(0,1)) + \frac{1}{4}u(w_A^H(0,0))\Big] - \lambda^* p_1 u(w_A^H(0,1)) = 0.$$

This implies

$$\begin{aligned}G(\lambda = \lambda^*) &= \lambda^* u(w_A^H(1,1)) + (1-p_1)(1-\lambda^*)\frac{1}{2}u(w_A^H(1,1)) \\ &\quad + (1-p_1)(1-\lambda^*)\Big[\frac{1}{4}u(w_A^H(0,1)) + \frac{1}{4}u(w_A^H(0,0))\Big] - \lambda^* p_1 u(w_A^H(0,1)) \\ &< \lambda^* u(w_A^H(1,1)) - \lambda^* p_1 u(w_A^H(11|01)) \\ &\quad + (1-p_1)(1-\lambda^*)\Big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\Big] \\ &= F(\lambda = \lambda^*),\end{aligned}$$

indicating that the payment to the agents in the collusive-supervision contract is no greater than the payment in the collusion-proof contract, $w_A^H(1,1) < \tilde{w}_{cp}^H$, when $\lambda = \lambda^*$.

Next, if $\lambda = 1$, $F(\lambda = 1)$ can be written as

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

The right-hand side of the inequality above contains the terms associated with high output in (IC_{cp}), indicating that the payment to the agents in the collusive-supervision contract is no less than the payment in the collusion-proof contract, $w_A^H(1,1) > \tilde{w}_{cp}^H$, when $\lambda = 1$. We further check the derivative of G with respect to λ; that is,

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

The derivative of F with respect to λ is also positive. Therefore, by the continuity of λ, the two functions G and F will cross only once, denoted by $\overline{\lambda} \in (\lambda^*, 1)$, such that if $\lambda = \overline{\lambda}$, $\tilde{w}_A^H = w^H(1,1)$.

Summarizing the analysis above shows that if $\underline{\lambda} < \lambda < \overline{\lambda}$, allowing a sub-coalition induces higher incentives for the agents to work. □

Proof of Proposition 5. From part (a) of Proposition 4, we have $w_A^H(11|00) = w_A^H(1,1)$. For the principal's cost-minimization problem (P_{cs}), we have the following Lagrangian.

$$\begin{aligned}\mathcal{L} &= 2w_A^H(1,1) + s^H(1,1)\\ &- \delta_1\Big\{\lambda u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(11|01)) + (1-p_1)(1-\lambda)\big[\frac{3}{4}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01))\big]\\ &- (1-p_1)\Big(\lambda u(w_A^L(11|01)) + (1-\lambda)\big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00))\big]\Big) - \varphi\Big\}.\end{aligned}$$

with the additional non-negativity constraints. It is clear that with the realization of output L, it is optimal for the principal not to reward both the supervisor and the agents, and thus, $w_i^L(r) = s^L(r) = 0, \ \ \forall\ r \in \Theta, \ \ i = A, B$. Therefore, our proof needs to focus only on the realization of output H.

Step I. We show that $w_A^H(1,1) = w_A^H(1,0) > 0$. The Kuhn-Tucker conditions for $w_A^H(1,1)$ are

$$\begin{aligned}\text{(C1)}: \quad \frac{\partial L}{\partial w_A^H(1,1)} &= 2 - \delta_1\Big\{\big[\lambda + \frac{3}{4}(1-\lambda)(1-p_1)\big]u'(w_A^H(1,1))\\ &+ \big[\frac{1}{4}(1-\lambda)(1-p_1) - \lambda p_1\big]u'(w_A^H(11|01))\frac{\partial w_A^H(11|01)}{\partial w_A^H(1,1)}\Big\} \geqslant 0,\\ w_A^H(1,1) \geqslant 0 \ \text{ and } \ & w_A^H(1,1)\frac{\partial L}{\partial w_A^H(1,1)} = 0;\end{aligned}$$

In (C1), it is impossible to have $\delta_1 = 0$ because $w_A^H(1,1) = 0$ will violate the IC constraint, yielding a contradiction. Thus, we should have $\delta_1 > 0$, indicating that $w_A^H(1,1) > 0$. By Lemma 2, we further have $w_A^H(1,0) = w_A^H(1,1) > 0$.

Step II. We show that it is optimal to set $w_A^H(0,1) = 0$ and $s^H(1,1) = 0$. $w_A^H(0,1)$ does not appear in the objective function and enters in the IC constraint only through $w_A^H(11|01)$ via the threatpoint payoff of agent A in the Nash bargaining problem. $w_A^H(11|01)$ and $s^H(11|01)$ are determined by Nash bargaining; clearly, in (2), $w_A^H(11|01)$ decreases when $w_A^H(0,1)$ decreases. Therefore, it is optimal for the principal to set $w_A^H(0,1) = 0$ in the IC constraint.

Because collusion proofness holds between signals of (1, 1) and (0, 0), we have $T^H(1,1) = T^H(0,0)$, which allows us to rewrite the Nash bargaining as

$$\text{Max}\big[u(T^H(0,0) - s^H(11|01)) - u(w_A^H(0,1))\big]^{\alpha}\big[s^H(11|01) - s^H(0,1)\big]^{1-\alpha}.$$

This fact tells us that $s^H(1,1)$ appears only in the objective function. The Kuhn-Tucker conditions for $s^H(1,1)$ are

$$\text{(C2)}: \qquad \frac{\partial L}{\partial s^H(1,1)} = 1 > 0,\ s^H(1,1) \geqslant 0 \ \text{ and } \ s^H(1,1)\frac{\partial L}{\partial s^H(1,1)} = 0;$$

Therefore, $s^H(1,1) = 0$. Given $s^H(1,1) = 0$ and $w_A^H(1,1) = w_A^H(0,0)$, the bargaining solution is $\alpha(w_A^H(1,1) - s^H(0,1))$ for agent A and $(1-\alpha)(w_A^H(1,1) - s^H(0,1))$ for the supervisor.

Step III. We show that $s^H(1,0) = 0$ in the contract. $s^H(1,0)$ does not appear in the objective function. Intuitively, we should have $w_A^H(1,0) + s^H(1,0) \leqslant w_A^H(1,1) + s^H(1,1)$; otherwise, the supervisor is induced to report (1, 0) with a signal of $(1,1)$. Given that $s^H(1,1) = 0$ in *Step* II and $w_A^H(1,0) = w_A^H(1,1)$ by Lemma 2, the inequality implies $s^H(1,0) \leqslant 0$. However, by definition, $s^H(1,0) \geqslant 0$. To satisfy both inequalities, we should therefore have $s^H(1,0) = 0$. Given that the agents are symmetric, $s^H(1,0) = s^H(0,1) = 0$.

Step IV. We show that it is optimal to set $s^H(0,0) = 0$. Although $s^H(0,0)$ does not appear in the objective function, intuitively, we should have $2w_A^H(0,0) + s^H(0,0) \leqslant 2w_A^H(1,1)$; otherwise, the supervisor will report (0, 0) with a signal of (1, 1). Because $w_A^H(0,0) = w_A^H(1,1)$, it yields $s^H(0,0) \leqslant 0$. However, by definition, $s^H(0,0) \geqslant 0$. To satisfy both inequalities, we should therefore have $s^H(0,0) = 0$.

Let us use $\breve{w}_{cs}^H$ to denote the equilibrium payment of $w_A^H(1,1)$. The payoff structure for agent A can be summarized as follows. Rewarding $\breve{w}_{cs}^H$ with signals of $(1,1)$ or $(1,0)$ or $(0,0)$ and rewarding 0 with signal of $(0,1)$, where $\breve{w}_{cs}^H$ is determined by ($\breve{IC}_{cs}$):

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

where $\alpha \in (0,1)$ is agent A's bargaining power in the Nash bargaining game in the sub-coalition. The payment to the supervisor is 0 across all signal states. By the symmetry of agents, we can easily construct the payment scheme for agent B as well. Given the payment feature, a sub-coalition arises with signals of $(1,0)$ and $(0,1)$, but the full-coalition is prevented. In the collusive-supervision contract, the principal incurs a total cost of $\check{C}_{cs} = 2\breve{w}_{cs}^H$. □

Proof of Proposition 7. Part (a): Differentiating $\overline{\lambda}$ with respect to α in (4) yields

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

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

Part (b): Plugging $\alpha = 0$ into (4) gives $\overline{\lambda} = 1$. With $\alpha = 0$, ($\breve{IC}_{cs}$) can be written as follows:

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

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

Plugging $\alpha = 1$ into (4) gives $\overline{\lambda} = \lambda^*$. With $\alpha = 1$, ($\breve{IC}_{cs}$) can be written as

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

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 $\breve{w}_{cs}^H \geqslant \tilde{w}_{cp}^H$.

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

References

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

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

Burlando, A. and A. Motta (2015): "Collusion and the organization of the Firm," *American Economic Journal: Microeconomics*, 7, 54–84. 4

Cadot, O. (1987): "Corruption as a gamble," *Journal of Public Economics*, 33, 223–244. 4

Celik, G. (2009): "Mechanism design with collusive supervision," *Journal of Economic Theory*, 144, 69–95. 2

Che, Y. K. (1995): "Revolving doors and the optimal tolerance for agency collusion," *RAND Journal of Economics*, 26, 378–397. 2

Dewatripont, M. and J. Tirole (2005): "Modes of communication," *Journal of Political Economy*, 113, pp. 1217–1238. 4

Grimaud, A. F., J. J. Laffont, and D. Martimort (2003): "Collusion, delegation and supervision with soft information," *Review of Economic Studies*, 70, 253–279. 2

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

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. 2

Khalil, F., D. Kim, and J. Lawarrée (2013): "Contracts offered by bureaucrats," *RAND Journal of Economics*, 44, 686–711. 4

Khalil, F., J. Lawarrée, and T. J. Scott (2015): "Private monitoring, collusion, and the timing of information," *RAND Journal of Economics*, 46, 872–890. 4

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, 2, 7, 7.1

Kofman, F. and J. Lawarrée (1993): "Collusion in hierarchical agency," *Econometrica*, 629–656. 4

——— (1996): "On the optimality of allowing collusion," *Journal of Public Economics*, 61, 383–407. 4, 2, 7

Laffont, J.-J. (1990): "Analysis of hidden gaming in a three-level hierarchy," *Journal of Law, Economics, and Organization*, 6, pp. 301–324. 2, 6.1

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

——— (1998): "Collusion and delegation," *RAND Journal of Economics*, 280–305. 4

——— (1999): "Separation of regulators against collusive behavior," *RAND journal of economics*, 232–262. 4

——— (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. 4

Lambert-Mogiliansky, A. (1998): "On optimality of illegal collusion in contracts," *Review of Economic Design*, 3, 303–328. 4

Liang, P. (2013): "Optimal delegation via a strategic intermediary," *Games and Economic Behavior*, 82, 15–30. 4

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. 4

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

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

Strausz, R. (1997a): "Collusion and renegotiation in a principal–supervisor–agent relationship," *The Scandinavian Journal of Economics*, 99, 497–518. 4

——— (1997b): "Delegation of monitoring in a principal-agent relationship," *Review of Economic Studies*, 64, pp. 337–357. 4

Tirole, J. (1986): "Hierarchies and bureaucracies: On the role of collusion in organizations," *Journal of Law, Economics, and Organization*, 2, 181. 1, 2, 7.2

——— (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, 2

Vafaï, K. (2010): "Opportunism in organizations," *Journal of Law, Economics, and Organization*, 26, 158–181. 4