Supervisory Efficiency and Collusion in a Multiple-Agent Hierarchy*

Xiaogang Che† Yangguang Huang‡ Le Zhang§

August 28, 2021

Abstract

We analyze a principal-supervisor-two-agent hierarchy with inefficient supervision. The supervisor may collect an incorrect signal on the agents' effort levels. When reporting to the principal, the supervisor may collude with one or both agents to manipulate the signal in exchange for a bribe. In the hierarchy, we identify a new trade-off between inefficient supervision and supervisor-agent collusion: Due to the incorrect supervisory signal, truthfully reporting the supervisory signal under collusion proofness may mistakenly punish the agents. As a result, allowing a certain type of collusion helps correct the incorrect signal and provides a higher incentive for the agents to work. We characterize the optimal no-supervision, collusion-proof, and collusive-supervision contracts, and show that the collusive-supervision contract dominates the others when supervisory efficiency is at an intermediate level.

Keywords: three-level hierarchy, collusion, supervisory efficiency, multiple agents, optimal contract.
JEL codes: D73, D82, D86.

*We are grateful to the editor, the advisory editor, and two anonymous referees for their insightful suggestions. We would like to thank Kunal Sengupta for the 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, and Leslie Reinhorn for their 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, Beihang University, SPMiD Conference, EARIE Conference, International Conference on Game Theory, and Tokyo Conference on Economics of Institutions and Organizations.

†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

When teamwork is used to complete a task, it is common for all team members to be rewarded based on both objective performance indicators and subjective evaluations made by their supervisor. For example, in medical research, a manager (the principal) 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, that is, some researchers may free ride on others. To provide proper incentives, an intermediate supervisor is sent to assess the agents' performance and obtain signals about the contribution of each team member. Compared to the principal, the supervisor has better expertise to evaluate team members. Although having the supervision can be valuable in preventing free riding and incentivizing the entire team, it may lead to two potential 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 worked hard," which, in turn, mitigates the effectiveness of the supervision. It is therefore natural to ask how the principal should design the optimal contract in such an environment.

We study the contracting problem in a multiple-agent hierarchy with both possibly inefficient supervision and collusion. Specifically, we consider a three-level hierarchy with a principal, a supervisor (she), and two productive agents (he/they). Supervisory technology can be either efficient or inefficient. Specifically, information on the agents' effort levels is accurate under efficient technology, but an incorrect signal may be collected under inefficient technology. The principal, who prefers both agents to work, initiates a contract with the supervisor and the two agents for a production task. Each agent can choose to either work or shirk. The possible output level depends on the joint effort of both agents. After the production output is realized, the supervisor 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 agents to forge a signal that favors them. Formally, we call this the supervisor-agent coalition and assume that the information is soft for the coalition. The supervisor can either collude with one agent and form a sub-coalition or collude with both agents and form a full-coalition. Contingent on the realized output level and the supervisory report, the principal pays transfers to the agents and the supervisor according to the contract.

In the analysis, we first characterize the no-supervision contract that depends on the output level only, ignoring the supervisory information. Detaching the signal with payments can therefore avoid the problems of possibly incorrect signals and collusion. We next establish another benchmark case where the supervisor is honest and always reports the observed signal truthfully. Although there is no collusion problem, inefficient supervisory technology may yield an incorrect

signal. Therefore, an agent may be mistakenly punished when he works but his signal is negative. To incentivize the agents, it is no longer true that the principal only rewards the agents after observing the positive evidence on their performance. A comparison between the two contracts shows that the principal only uses a supervisor when the supervisory technology is sufficiently accurate; otherwise, the principal prefers the no-supervision contract to the collusion-free contract.

We then examine corruptible supervision and derive the collusion-proof contract under which the supervisor and the agent(s) have no incentive to form any coalition (Tirole, 1986). In the collusion-proof contract, an agent is rewarded for a positive signal but not for a negative signal. Moreover, to prevent collusion and induce truthful reporting, the principal rewards the supervisor with a payment equivalent to the agent's wage when she reports a negative signal. We show that when supervisory efficiency is sufficiently high, the principal prefers the collusion-proof contract to the no-supervisor contract, which indicates that a corruptible and possibly inefficient supervisor is still useful to the principal.

These benchmark cases help us identify a novel trade-off between inefficient supervision and supervisor-agent collusion. In the hierarchy with multiple agents, the principal's goal is to induce both agents to work; thus, the payment structure must satisfy the incentive compatibility constraint of each agent, which guarantees that the agent does not want to unilaterally deviate from working given that the other agent chooses to work. As a result, when both agents are observed with the negative signal $(0, 0)$, it must be an inaccurate signal collected by the inefficient supervisor. Therefore, instead of punishing the agents by rewarding the supervisor, it is optimal to reward the agents under the negative signal $(0, 0)$ the same as under the positive signal $(1, 1)$. Doing so avoids the agents being mistakenly punished by an obviously incorrect signal, and this correction improves the agents' incentives to work and prevents the full-coalition among the supervisor and both agents.

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.[1] 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.[2] Tirole (1986) also argues that as supervisory information becomes less verifiable, the supervisor is less useful. Thus, when information is soft,

[1] 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); Kessler (2000); Khalil et al. (2013, 2015); Burlando and Motta (2015).

[2] 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).

i.e., entirely unverifiable, supervision becomes completely useless in the hierarchy.

Almost all prior studies that examine collusion in a three-tier hierarchy focus on a single productive agent. With soft supervisory information, Kofman and Lawarrée (1996) and Khalil et al. (2010) find that there can be benefit from allowing collusion in the hierarchy. In Kofman and Lawarrée (1996), the supervisor (auditor) can be either honest or dishonest; therefore, 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. Khalil et al. (2010) consider the case in which the supervisor has an ability to conceal a positive signal to extort the agent. They show that both bribery (supervisor-agent collusion) and extortion weaken the incentive scheme, but extortion is more severe; thus, the principal benefits by allowing collusion to attenuate the room for extortion.

Although our model reaches the same result that collusion becomes useful in lowering the principal's total cost in the contract, the underlying trade-off is different: imperfect supervisory technology drives the usefulness of collusion in correcting supervisory signals. Moreover, this improvement exists only in the multiple-agent environment. In Section 4.1, we show that such a trade-off disappears, and collusion proofness becomes optimal in a single-agent hierarchy.

The studies on contract design in multiple-agent organizations with supervisor-agent collusion are rather limited.[3] Laffont (1990) examines 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, then the optimal payment scheme should be purely personalized; if, however, information is soft, then it may be optimal to utilize some of the aggregate information to design the incentive scheme. In this paper, we study a similar setting with soft information only, and collusion formation may generate an externality to affect a non-colluding agent's payoff. Our analyses contribute to this strand of literature by eliciting the reasons why and characterizing the conditions under which collusion benefits the principal.

In the hierarchy with multiple agents, the possible output level is determined by the joint effort, but the agents' performance is evaluated individually. Thus, contract design becomes rather complicated and challenging in contrast to a single-agent setting. 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. That is, 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

[3]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.

(2012); Ryall and Sampson (2016); Biener et al. (2018).

We find that the no-supervision contract and the collusion-proof contract are along the lines of individual performance, whereby an agent's reward is based on his own signal but not the other agent's signal. Nevertheless, the collusive-supervision contract shows a special feature that combines both performance measures. If the agent's signal is negative, then his rewards depend on the other agent's signal; if, however, his signal is positive, then 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.

Our results offer several managerial implications. First, in many organizations such as companies and institutions, the managerial hierarchy involves several layers from top and intermediate managers to multiple productive units or teams. Team members' effort levels are typically unobservable; thus, subjective and objective assessments are generally combined to evaluate their performance (MacLeod, 2003). Our analysis above provides a fundamental setting in which to model such complicated managerial environments and the related contract design problems to provide instructive answers to the questions of not only whether and when collusion should be allowed but also to what extent. This information provides guidance to the manager on what types of collusion she may allow in practice. Second, one should interpret our results with caution because although allowing collusion is beneficial in certain circumstances, it is still more desirable to advance supervisory efficiency and implement collusion proofness to incentivize agents; as our main results show, when supervisor efficiency is sufficiently high, the collusion-proof contract becomes preferred. This finding indicates that in organizations, the principals should consider not only the innovations of production technologies but also the improvement of supervisory technology.[4] Finally, our analysis suggests that when designing a contract in a multiple-agent hierarchy, the principal should give attention to the potential externality caused by a supervisor-agent coalition, which would jeopardize non-colluding agents' incentives for production and fail the collusion-proof principle.

The remainder of the paper proceeds as follows. In Section 2, we provide the model setup and present three benchmark cases. Section 3 shows our novel results of the collusive-supervision contract and the benefit of allowing collusion. We discuss the robustness of our results when relaxing certain assumptions in Section 4. Section 5 concludes. Essential proofs are presented in Appendix I. Nonessential proofs and computation details are presented in Appendix II.

[4]When the hierarchy is in a repeated-contracting environment, the principal would asymptotically learn a supervisor's true type and then almost surely contract with efficient supervisors. Moreover, repeated interaction would also help a supervisor improve her monitoring skills. As a consequence, both effects entail a more accurate signal from supervision, which would weaken the benefits from allowing collusion.

2 Model Setup and Benchmark Cases

2.1 The 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, which are 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, and $H > L > 0$. The principal is risk-neutral and has *ex post* payoff, $\pi = y - w_A^y - w_B^y - s$, where w_A, w_B, and s are payments to agent A, agent B, and the supervisor, respectively, given output y. 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; thus, there is no separable output from an individual agent. If both agents work, then 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, which is 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 facilitates our focus on the uncertainty entailed by supervisory technology and clearly identifies the trade-off between inefficient supervision and supervisor-agent collusion.[5] With the symmetry of agents, we have $p_1 \equiv p(0, 1) = p(1, 0)$. Parameter p_1 measures how easy it is for an agent to free ride on the other agent.

The principal strictly prefers both agents to exert their efforts on production, i.e., $e = (1, 1)$. Given $p(1, 1) = 1$, maximizing the expected revenue is equivalent to minimizing the expected total payments for the principal. Therefore, in designing the incentive schemes, the principal aims to minimize the expected total payments from implementing $e = (1, 1)$. Agent i has a utility function $u(w_i^y) - \varphi e_i$, where w_i^y is the payment that agent i receives, and $\varphi > 0$ denotes the disutility level of working. $u(w_i^y)$ satisfies $u(0) = 0$, $u'(\cdot) > 0$ and $u''(\cdot) \leqslant 0$. 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 $\Theta \equiv \{(1, 1), (0, 1), (1, 0), (0, 0)\}$.[6] For each signal $\theta \equiv (\theta_A, \theta_B)$, θ_A and θ_B represent the signals of the effort levels of agent A and agent B, respectively. The agents can also observe the signal θ but cannot make their own reports to the principal. Supervisory technology is imperfect, which means that the supervisor can be either *efficient*

[5] In Section 4.4, we analyze the case of $p(1, 1) < 1$ and show that it is still beneficial to allow collusion in the presence of production uncertainty.

[6] We discuss different state-space settings with the possibility of uninformative signal in Section 4.3.

or *inefficient* with probabilities λ and $1-\lambda$, respectively. Parameter $\lambda \in [0,1]$ captures 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, then she observes a random signal, that is, each $\theta \in \Theta$ is randomly observed with an equal probability of $1/4$. The supervisor observes an incorrect signal when $\theta \neq e$. Thus, the overall probabilities that the supervisor collects an incorrect and a correct signal are given by $\frac{3}{4}(1-\lambda)$ and $\lambda + \frac{1}{4}(1-\lambda)$, respectively. After collecting the signal, the supervisor sends a report $r \equiv (r_A, r_B) \in \Theta$ to the principal about both agents' effort levels.

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. Let $T^y(r)$ denote the aggregate transfer made by the principal and $T^y(r) \equiv w_A^y(r) + w_B^y(r) + s^y(r)$. 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 collude and manipulate the signal, for example, by reporting 1 for 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.[7]

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 objective of the supervisor-agent coalition is to forge a report r that maximizes the total payment from the principal. We 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, members in the coalition divide the total payment by a Pareto efficient bargaining procedure in which each party in the coalition receives no less than what they would receive from choosing not to collude. For example, if the contract has the feature that $w_A^H(1,1) + s^H(1,1) > w_A^H(0,1) + s^H(0,1)$, then when observing $y = H$ and $\theta = (0,1)$,

[7]This assumption indicates that the supervisor cannot manipulate the signal to extort the agents by herself. For instance, in the development of a new drug, it is almost impossible for the supervisor to modify the pharmaceutical research data without the researchers' (agents') help.

the supervisor can cooperate with agent A, report $r = (1,1)$, and then split the total payment $w_A^H(1,1) + s^H(1,1)$; in this case, agent A obtains $w_A^H(11|01)$, and the supervisor obtains $s^H(11|01)$, where $w_A^H(11|01) + s^H(11|01) = w_A^H(1,1) + s^H(1,1)$.

With multiple agents, the supervisor can collude with one of the two agents or with both agents. A *sub-coalition* between the supervisor and agent i requires the following necessary conditions:

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

$$w_i^y(r|\theta) \geqslant w_i^y(\theta) \text{ and } s^y(r|\theta) \geqslant s^y(\theta) \text{ for } r \neq \theta, \tag{2}$$

where (1) indicates that given the observed signal θ, the total payment to agent i and the supervisor in the coalition is equal to the total payment from the principal with report r; (2) guarantees that each party in the coalition receives no less than what they would receive when choosing not to collude.

A *full-coalition* between the supervisor and both agents requires the following necessary conditions:

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

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

where (3) and (4) serve similar roles as (1) and (2), respectively.

The supervisor and the agents will not collude if they are indifferent between colluding and not colluding.[8] In the following, we say that a contract offered by the principal is non-collusion-proof if either a sub-coalition or a full-coalition is formed. If no coalition is formed, then the supervisor will report truthfully.

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

(1) The principal offers a contract that specifies 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, then the game ends, and all parties receive their respective reservation utilities.
(3) If the contract has been accepted, then the two agents simultaneously decide whether to work ($e_i = 1$) or to shirk ($e_i = 0$). After the agents making their decisions, output y is realized.

[8] In this case, the principal can break ties by increasing relevant payments by a penny to ensure "not colluding." It is easy to check that our characterization of the optimal collusive-supervision contract in Proposition 4 will not be affected by the tie-breaking rule.

(4) The supervisor is sent to assess both agents' performance, and output y is observed by all parties.[9]
(5) Signal θ is realized and observed by the supervisor and the two agents. The supervisor and the agent(s) choose whether to collude and make a side contract. If the side contract is rejected, then the supervisor will play noncooperatively (report truthfully).
(6) The supervisor makes the report r to the principal.
(7) Transfers are paid according to the contract (and the side contract if necessary).

Our analyses proceed as follows. We first establish the relevant contracts with no supervision, honest supervision, and collusion proofness. Establishing these three benchmark contracts helps us identify how inefficient supervision and collusion affect the principal's total costs and the trade-off between them. Finally, we examine the collusive-supervision contract, and show that it is beneficial to allow a certain level of collusion between the supervisor and the agents.

2.2 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. A simple way to avoid both problems is to completely ignore the supervisory information. In this case, the agents' payments are based solely on output y, i.e., $w_i^y(r) = w_i^y$ for all r, and the supervisor receives no payment. We refer to this contract as the no-supervision (no) contract.

The principal's objective is to implement effort choice $e = (1, 1)$ with the minimum total payments $C_{no} = w_A^H + w_B^H$. Taking agent A as the representative, given that agent B chooses to work ($e_B = 1$), the incentive compatibility (IC) constraint is

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

Note that the participation constraint is also satisfied when the IC constraint is satisfied. We thus omit the participation constraint here (and in the following analysis). The principal's cost-minimization problem can then be written as follows

$$(P_{no}) \qquad \begin{aligned} &\min C_{no} = w_A^H + w_B^H \\ &\text{s.t. } u(w_i^H) - \varphi \geqslant p_1 u(w_i^H) + (1 - p_1)u(w_i^L) \text{ for } i = A, B. \end{aligned}$$

[9] A typical example that satisfies such an environment is the medical research mentioned above, where the researchers (agents) are involved in only some components of the research task, and the supervisor will be used to assess the agents' performance only after the pharmaceutical research outcome is realized. In this environment, it is unlikely that the supervisor and the agents can collude before effort is exerted.

We focus on a symmetric equilibrium. The solution to (P_{no}) yields the *optimal no-supervision contract* as follows.

Proposition 1. *In the optimal no-supervision contract,*

(a) *for $y = L$, the agents do not obtain any rewards, i.e., $w_i^L = 0$ for $i = A, B$;*

(b) *for $y = H$, each agent obtains $w_i^H = \hat{w}_{no}^H$ 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 $y = H$.

2.3 Honest supervision and collusion-free contract

We now consider another benchmark case in which the supervisor is honest and always truthfully reports the observed signal, i.e., $r = \theta$. Note that the signal may not accurately reflect the agents' effort levels. In this collusion-free (cf) environment, we denote the principal's expected cost from implementing $e = (1, 1)$ by C_{cf}, which is given by

$$C_{cf} = \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(1,0) + T^H(0,1) + T^H(0,0)\Big],$$

where the first term is the payment when the supervisor is efficient (with probability λ); the second term is the payment when the supervisor is inefficient (with probability $1 - \lambda$).

Taking agent A as the representative, given that $e_B = 1$ and the supervisor reports truthfully, the IC constraint is

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

where the left-hand side of (IC_{cf}^A) is the expected payoff when agent A chooses to work, 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_{cf}^A) is the payoff when agent A chooses to shirk, 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. Given the symmetry of the two agents, a similar IC constraint (IC^B_{cf}) can be constructed for agent B.

The principal's cost minimization problem can then be written as follows:

$$(P_{cf}) \qquad \begin{aligned} &\min C_{cf} = \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(1,0) + T^H(0,1) + T^H(0,0)\Big] \\ &\text{s.t. } (IC^A_{cf}),\ (IC^B_{cf}). \end{aligned}$$

The cutoff value $\underline{\lambda} = \frac{1-p_1}{1+3p_1}$ is defined as the root of the equation

$$\frac{1}{4}(1-p_1)(1-\underline{\lambda}) - \underline{\lambda}p_1 = 0. \tag{5}$$

The solution to (P_{cf}) yields the *optimal collusion-free contract*, in which the contingent transfers to the supervisor and agents are characterized as follows.[10]

Proposition 2. *Given $\lambda > \underline{\lambda}$, the optimal collusion-free contract is given by the following:*

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

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

Report r	*Agent A*	*Agent B*	*Supervisor*
$(1,1)$	$\tilde{w}^H_{cf}$	$\tilde{w}^H_{cf}$	0
$(1,0)$	$\tilde{w}^H_{cf}$	0	0
$(0,1)$	0	$\tilde{w}^H_{cf}$	0
$(0,0)$	$\tilde{w}^H_{cf}$	$\tilde{w}^H_{cf}$	0

where $\tilde{w}^H_{cf}$ is determined by equation

$$(\widetilde{IC}_{cf}) \qquad \lambda u(\tilde{w}^H_{cf}) + \frac{3}{4}(1-p_1)(1-\lambda)u(\tilde{w}^H_{cf}) = \varphi.$$

The expected cost of the principal is $\tilde{C}_{cf} = (\frac{3}{2} + \frac{1}{2}\lambda)\tilde{w}^H_{cf}$.

The optimal collusion-free contract exhibits an interesting feature that given the possibility of acquiring an inaccurate supervisory signal, it is no longer true for the principal to only reward the agents after obtaining definitive evidence on their performance. Specifically, with the observation

[10] Other equilibria exist in which the agents' wages after observing signal $(0,0)$ do not need to be positive, particularly when the agents are risk-neutral. However, given that our study focuses on the trade-off between inefficient supervision and collusion, we restrict our attention to the stated wages in Proposition 2 for the ease of comparison across contracts.

of output H and signal $(0,0)$, the principal needs to reward the agents positively in the contract, i.e., $\tilde{w}_{cf}^{H}$. The reason is that by construction, the IC constraint reflects an agent's unilateral decision on whether to work or shirk, given that the other agent works. Thus, when signal $(0,0)$ is observed, it must be an incorrect signal from an inefficient supervisor. To correct the wrong signal and incentivize both agents, the principal should reward $w_i^H(0,0) = w_i^H(1,1)$.

Note that the supervisory signal will only be used if it is sufficiently accurate, i.e., $\lambda > \underline{\lambda}$. This is because when the signal is highly inaccurate ($\lambda \leqslant \underline{\lambda}$), the agents are exposed to a high possibility of mistaken punishment due to inefficient supervision; thus, they require more compensation. As a result, it is better for the principal not to make the transfers contingent on the highly inaccurate supervisory report, even though the supervisor always honestly reports what she observes. Therefore, when $\lambda \leqslant \underline{\lambda}$, the principal is better off adopting the no-supervisor contract.

2.4 Collusion-proof contract

Next, we examine collusive and possibly inaccurate supervision in the hierarchy and characterize the collusion-proof (cp) contract that leaves no incentive for the supervisor and the agents to collude. To prevent collusion, the principal must ensure that truthful reporting does not result in strictly less joint payments for all possible coalitions. These conditions are called coalition incentive compatibility (CIC) constraints (Tirole, 1992). Let us first consider the full-coalition deterrence. Given output y, to ensure truthful reporting, the CIC constraints are $T^y(\theta) \geqslant T^y(r)$ for all $\theta, r \in \Theta$. This implies that the aggregate payments to the two agents and the supervisor must be exactly the same across the four signal states:

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

Next, we consider the sub-coalition deterrence. Taking agent A as the representative, given agent B's signal, the CIC constraints are $w_A^y(\theta_A,1) + s^y(\theta_A,1) \geqslant w_A^y(r_A,1) + s^y(r_A,1)$ and $w_A^y(\theta_A,0) + s^y(\theta_A,0) \geqslant w_A^y(r_A,0) + s^y(r_A,0)$ for all $\theta_A, r_A \in \{0,1\}$. Satisfying the inequalities requires that the total payment to agent A and the supervisor are exactly the same across the two signal states. Similar inequalities can be constructed for agent B. These inequalities imply the following conditions:

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

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* $y = L, H$*.*

Lemma 1 indicates that to fully deter both types of coalitions, the incentive scheme for an agent should not depend on the other agent.

Under collusion proofness, the principal's cost-minimization problem can be written as follows:

$$(P_{cp}) \qquad \begin{aligned} &\min C_{cp} = \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(1,0) + T^H(0,1) + T^H(0,0)\Big] \\ &\text{s.t.}(IC_{cp}^A),\ (IC_{cp}^B),\ (CIC_f),\ (CIC_s). \end{aligned}$$

Because (CIC_f) and (CIC_s) guarantee truthful reporting, the expression of the objective function C_{cp} is the same as C_{cf}. The IC constraint (IC_{cp}^A) is the same as (IC_{cf}^A). The cutoff value $\lambda^* \equiv \frac{1-p_1}{1+p_1}$ is defined as the root of the equation

$$\frac{1}{2}(1-p_1)(1-\lambda^*) - \lambda^* p_1 = 0. \tag{6}$$

The solution to (P_{cp}) yields the *optimal collusion-proof contract* described below:

Proposition 3. *Given* $\lambda > \lambda^*$*, the optimal collusion-proof contract is given by*

(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*

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 $\tilde{C}_{cp} = 2\tilde{w}_{cp}^H$*.*

Lemma 1 and Proposition 3 explain why the collusion-proof contract induces a higher cost to the principal. Collusion proofness requires that an agent's wage transfer only depends on the signal about his own effort level. Thus, deterring supervision-agent collusion prevents the principal from effectively utilizing equilibrium information, i.e., that the other agent exerts high effort, when providing incentive to an agent. As a result, both agents are rewarded with zero wage after observing signal (0, 0) in the optimal collusion-proof contract, and this increases the cost of implementing high efforts.[11]

Proposition 3 also shows how supervisory efficiency, which is measured by parameter λ, affects the principal's contract choices. The optimal collusion-proof contract dominates the no-supervision contract only when supervisory information is sufficiently accurate, i.e., $\lambda > \lambda^*$; otherwise, the principal prefers the no-supervision contract in which the supervisory signal is ignored. Moreover, the corruptibility of the supervisor and the cost of collusion prevention lower the principal's incentive to hire a supervisor; therefore, $\lambda^* > \underline{\lambda}$.

3 Incentive Improvement by Allowing Collusion

The two benchmark cases of the optimal collusion-free contract and collusion-proof contract help us clearly identify the trade-off between inefficient supervision and supervisor-agent collusion. In the following, we explore the possibility of striking a balance in the trade-off. In particular, is it possible to correct an incorrect signal and lower the principal's total cost by allowing collusion? The answer is *yes*. We hereby characterize the optimal collusive-supervision (cs) contract and then demonstrate the benefit of allowing collusion.

3.1 Collusive-supervision contract

We examine the contract design problem in which the (CIC_f) and (CIC_s) constraints are removed from the principal's cost minimization problem. The principal needs to consider signal manipulation and the payoffs that result from supervisor-agent collusion. We first establish the following lemma.

Lemma 2. *For* $y = L, H$, (a) $T^y(1,1) \geqslant T^y(r)\ \forall\ r = \{(1,0),(0,1),(0,0)\}$; *and* (b) $s^y(1,1) = 0$.

Part (a) states that, to implement $e = (1,1)$, the aggregate transfers to the two agents and the supervisor under signal $(1,1)$ should be no less than those under other signals. By symmetry, we have $w_A^H(1,1) = w_B^H(1,1)$. Moreover, with $\theta = (1,1)$, both agents have no incentive to collude and

[11] Note that by comparing ($\widetilde{IC}_{cp}$) and ($\widetilde{IC}_{cf}$), we can easily show that $\tilde{w}_{cp}^H > \tilde{w}_{cf}^H$. Therefore, $\tilde{C}_{cp} = 2\tilde{w}_{cp}^H > (1+\lambda)\tilde{w}_{cf}^H = \tilde{C}_{cf}$, which indicates that the collusion-free contract (with an honest supervisor) always dominates the collusion-proof contract (with a corruptible supervisor).

the supervisor truthfully reports $r = (1,1)$. Thus, there is no need to reward the supervisor when she reports $r = (1,1)$. This gives Part (b).

With Lemma 2, we only need to consider three possible cases of the upward adjustment of the supervisory signal. Specifically, taking agent A as the representative, there are three relevant cases of collusion. First, given $\theta = (0,0)$, a sub-coalition is formed, and $r = (1,0)$ is reported. Second, given $\theta = (0,0)$, full-coalition is formed, and $r = (1,1)$ is reported. Third, given $\theta = (0,1)$, a sub-coalition is formed, and $r = (1,1)$ is reported.

Let us denote the principal's objective function in the cost minimization problem as

$$C_{cs} = \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(r|10) + T^H(r|01) + T^H(r|00)\Big],$$

where $T^H(r|\theta)$ denotes the aggregate transfer to the agents and the supervisor when signal θ is observed but the supervisor reports r. Furthermore, given $e_B = 1$, the IC constraint of the representative agent A can be written as

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

where the left-hand side of (IC_{cs}^A) is the expected payoff when agent A chooses to work; the right-hand side of (IC_{cs}^A) is the payoff when agent A shirks. The notations $w_A^y(r|10)$, $w_A^y(r|01)$, and $w_A^y(r|00)$ denote agent A's final payments under the possible signal manipulations. Given the symmetry of the two agents, a similar IC constraint (IC_{cs}^B) can be constructed for agent B. The principal's cost minimization problem is given by

$$\begin{aligned}
(P_{cs})\qquad &\min C_{cs} = \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(r|10) + T^H(r|01) + T^H(r|00)\Big] \\
&\text{s.t. } (IC_{cs}^A),\ (IC_{cs}^B).
\end{aligned}$$

Let $\alpha \in (0,1)$ denote the agent's share of the total payment received by the sub-coalition in the Pareto-efficient bargaining outcome.[12] The solution to (P_{cs}) yields the *optimal collusive-supervision contract* as follows.

[12] In a Nash bargaining game, α and $1-\alpha$ capture the bargaining powers of the colluding agent and the supervisor, respectively. See Section 3.3 for a further discussion.

Proposition 4. *Given $\lambda > \underline{\lambda}$, the optimal collusive-supervision contract is given by the following:*

(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*

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

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

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

From Proposition 4, the payment structure in the optimal collusive-supervision contract is the same as in the collusion-free contract (Section 2.3).[13] To correct the incorrect signal $(0,0)$, the principal rewards the agents positively with compensation. Therefore, we have the same payment across the signals of $(1,1)$ and $(0,0)$, and as a consequence, the full-coalition is prevented. However, under the payment structure, a sub-coalition will be formed. Specifically, with $\theta = (0,1)$ (or $(1,0)$), agent A (or agent B) and the supervisor will collude to manipulate the signal and report $r = (1,1)$; the colluding agent obtains $\alpha\check{w}_{cs}^H$, and the supervisor receives $(1-\alpha)\check{w}_{cs}^H$ as the bribe.

The collusive-supervision contract further shows that when collusion is allowed, it is still optimal for the principal to reward $w_A^H(1,1) = w_A^H(1,0)$ and $w_B^H(1,1) = w_B^H(0,1)$. In other words, the payment to the agent with $e_i = 1$ is regardless of the other agent's effort level. If $w_A^H(1,1) > w_A^H(1,0)$, then when signal $\theta = (1,0)$ is realized, agent A will obtain $w_A^H(11|10) > w_A^H(1,0)$ by colluding with the supervisor and agent B to report $r = (1,1)$. However, since $w_A^H(11|10) < w_A^H(1,1)$, rewarding $w_A^H(1,1) = w_A^H(1,0)$ instead prevents the full-coalition and provides a higher incentive to work. If $w_A^H(1,1) < w_A^H(1,0)$, then when signal $\theta = (0,0)$ is realized, since $w_A^H(1,0) + s^H(1,0) > w_A^H(0,0) + s^H(0,0)$, agent A will collude with the supervisor who manipulates $\theta = (0,0)$ to $r =$

[13] The comparison between the $(\widetilde{IC}_{cs})$ and $(\widetilde{IC}_{cf})$ constraints shows that with $\alpha \in (0,1)$, $\check{w}_{cs}^H > \tilde{w}_{cf}^H$ and $\check{C}_{cs} = 2\check{w}_{cs}^H > (\frac{3}{2} + \frac{1}{2}\lambda)\tilde{w}_{cf}^H = \tilde{C}_{cf}$, which indicates that the optimal collusion-free contract dominates the optimal collusive-supervision contract. Note that if the supervisor can take the full share of the total payment, i.e., $\alpha = 0$, then the two contracts are equivalent.

$(1,0)$, which jeopardizes the non-colluding agent B's incentive to work without changing his signal. Accordingly, to eliminate the negative externality, the principal should set $w_A^H(1,1) = w_A^H(1,0)$. Since both agents are symmetric, the same argument can be applied to agent B.

Given $w_i^H(0,0) = w_i^H(1,1)$, if the supervisor were rewarded positively with $r = (1,0)$, then this would lead to $w_A^H(1,0) + s^H(1,0) > w_A^H(0,0) + s^H(0,0)$. As a result, the supervisor would then collude with agent A who manipulates $\theta = (0,0)$ to $r = (1,0)$, which jeopardizes the non-colluding agent B's incentive to work without changing his signal ($w_B^H(1,0) < w_B^H(1,1)$). To eliminate this negative externality, the principal has to compromise by setting $s^H(1,0) = 0$. Similarly, one must set $s^H(0,1) = 0$ to prevent a sub-coalition with agent B and the supervisor when $\theta = (0,0)$. This externality caused by a sub-coalition is a special feature of the multiple-agent environment, which does not exist in a single-agent hierarchy. As a result, this payment scheme allows another type of sub-coalition that manipulates $\theta = (0,1)$ to $r = (1,1)$. Note that this type of sub-coalition does not generate the negative externality to agent A, as $w_B^H(0,1) = w_B^H(1,1)$. Therefore, to correct the obviously incorrect signal $\theta = (0,0)$ and avoid the negative externality caused by a sub-coalition, the principal must allow the sub-coalition that manipulates $= (0,1)$ or $(1,0)$ to $r = (1,1)$.

3.2 Comparisons across contracts

We are ready to compare the optimal no-supervision, collusion-proof, and collusive-supervision contracts, and identify the conditions under which the optimal collusive-supervision contract is better. The three contracts share a common feature that the aggregate payment by the principal 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 determined by ($\widehat{IC}_{no}$), ($\widetilde{IC}_{cp}$), and ($\breve{IC}_{cs}$), respectively. Thus, given the same payment $w_A^H(1,1)$, we can compare which equilibrium IC constraint provides a higher incentive to the agent. Write the equilibrium IC constraints in the form of function $Z \geqslant 0$:

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

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

$$
\begin{aligned}
Z_{cs} - Z_{no} &= \left[\lambda + \frac{3}{4}(1-p_1)(1-\lambda) - (1-p_1)\right]u(w_A^H(1,1)) + \left[\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1\right]u(w_A^H(11|01)) \\
&= \left[\lambda p_1 - \frac{1}{4}(1-p_1)(1-\lambda)\right]\Big[u(w_A^H(1,1)) - u(w_A^H(11|01))\Big].
\end{aligned} \tag{7}
$$

Clearly, $Z_{cs} - Z_{no}$ is increasing in λ. Since $u(w_A^H(1,1)) - u(w_A^H(11|01)) > 0$ for all $\alpha \in (0,1)$, we have $Z_{cs} - Z_{no} > 0$ when $\lambda > \underline{\lambda}$. In this case, collusive supervision provides greater incentives for the agents to work.

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

$$
\begin{aligned}
Z_{cs} - Z_{cp} &= \frac{1}{4}(1-p_1)(1-\lambda)u(w_A^H(1,1)) + \left[\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1\right]u(w_A^H(11|01)) \\
&= \frac{1}{4}(1-p_1)(1-\lambda)\Big[u(w_A^H(1,1)) - u(w_A^H(11|01))\Big] - \lambda p_1 u(w_A^H(11|01)).
\end{aligned} \tag{8}
$$

When $\lambda = 1$, $Z_{cs} - Z_{cp} < 0$, collusion proofness provides a higher incentive. However, when $\lambda = \lambda^*$, we have $Z_{cs} - Z_{cp} > 0$, which implies that the collusive-supervision contract is preferable. Furthermore, $d(Z_{cs} - Z_{cp})/d\lambda = -\frac{1}{4}(1-p_1) - p_1 u(w_A^H(11|01)) < 0$ indicates that $Z_{cs} - Z_{cp}$ is decreasing in $\lambda \in [\lambda^*, 1]$. Accordingly, a unique cutoff value must exist, which is denoted by $\overline{\lambda} \in (\lambda^*, 1)$, such that $Z_{cs} - Z_{cp} = 0$; equivalently,

$$
\overline{\lambda} \equiv \frac{(1-p_1)\big[u(w_A^H(1,1)) + u(w_A^H(11|01))\big]}{\big[(1-p_1)u(w_A^H(1,1)) + (1+3p_1)u(w_A^H(11|01))\big]}. \tag{9}
$$

The collusive-supervision contract provides greater incentives for the agents to work when $\lambda \in [\lambda^*, \overline{\lambda})$. Therefore, we can conclude the following.

Proposition 5. *In the multiple-agent hierarchy, the principal uses*

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

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

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

Figure 1 depicts the total costs of implementing $e = (1,1)$ under different contracts.[14] When λ is small, the supervisory technology is inaccurate. Allowing the payments to depend on the signal will expose the agents to excessive uncertainty that requires very high compensation. Hence, it is better to adopt the no-supervision contract. When λ is large, the supervisory technology is sufficiently reliable; thus, it is optimal to let the payments be contingent on truthfully reported signals obtained from collusion-proof implementation. When λ is in the intermediate range $[\underline{\lambda}, \overline{\lambda}]$, collusive supervision balances the gains and losses from using inefficient supervisory technology and becomes the optimal way to provide incentives.

[14] The computation details of Figures 1 and 2 are presented in Appendix II.

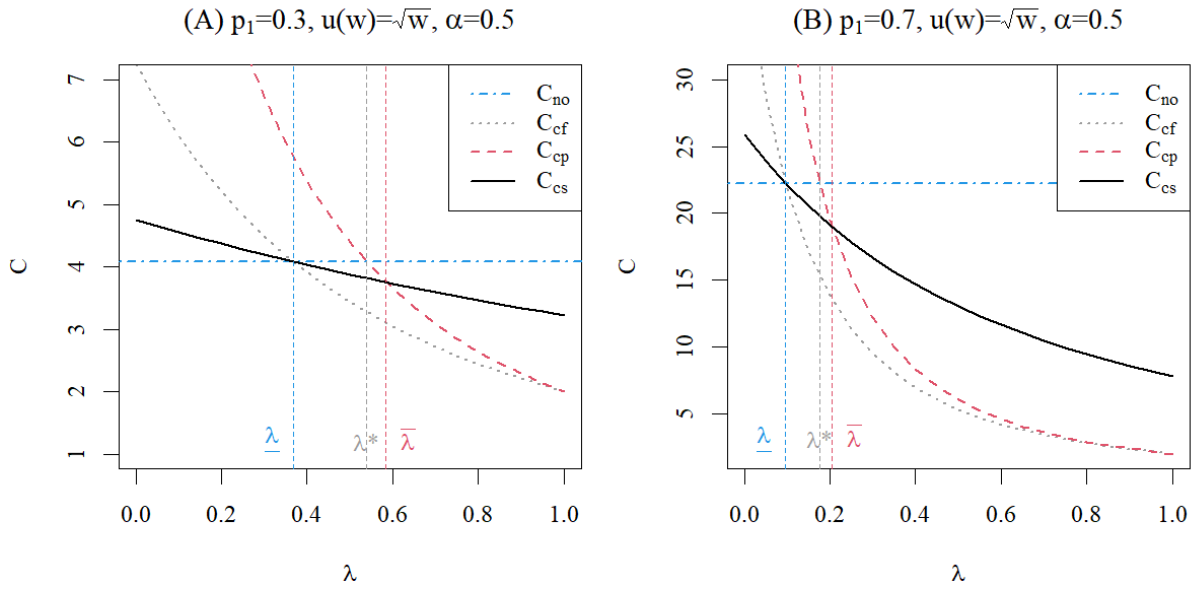

Figure 1: Comparison of the Total Costs under Different Contracts

Furthermore, both $\underline{\lambda}$ and $\overline{\lambda}$ decrease in p_1. As shown in Figure 1, Panel (B) with $p_1 = 0.7$ has lower $\underline{\lambda}$ and $\overline{\lambda}$ than those in Panel (A) with $p_1 = 0.3$. Recall that p_1 measures how severe the moral hazard problem is. Therefore, as shirking (and colluding with the supervisor) becomes more attractive to the agents, the principal would be more inclined to adopt the collusion-proof contract. The cost of implementing $e = (1, 1)$ is also considerably higher with a larger p_1.

3.3 Nash bargaining in the side contract

Assume that the side contract is conducted through a Nash bargaining problem (Nash, 1950). 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

$$\begin{aligned} &\max \left[w_A^H(11|01) - w_A^H(0,1)\right]^{\alpha} \left[s^H(11|01) - s^H(0,1)\right]^{1-\alpha}, \\ &\text{s.t.} \quad w_A^H(11|01) + s^H(11|01) = w_A^H(1,1) + s^H(1,1). \end{aligned}$$

From Proposition 4, the solution of the bargaining problem gives $w_A^H(11|01) = \alpha \check{w}_{cs}^H$ and $s^H(11|01) = (1-\alpha)\check{w}_{cs}^H$. Note that the principal's choice between the collusive-supervision and collusion-proof

contracts depends on $\overline{\lambda}$, which is given by

$$\overline{\lambda} = \frac{(1-p_1)\left[u(\check{w}_{cs}^H) + u(\alpha\check{w}_{cs}^H)\right]}{\left[(1-p_1)u(\check{w}_{cs}^H) + (1+3p_1)u(\alpha\check{w}_{cs}^H)\right]}.$$

Therefore, bargaining power α affects the principal's contract choice and total payment in equilibrium. We can then obtain the following proposition.

Proposition 6. *Regarding the agent's bargaining power α in the side contract, we find the following:*

(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, whereas if $\alpha \to 1$, then $\overline{\lambda} \to \lambda^*$, and the opposite dominance holds; and*

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

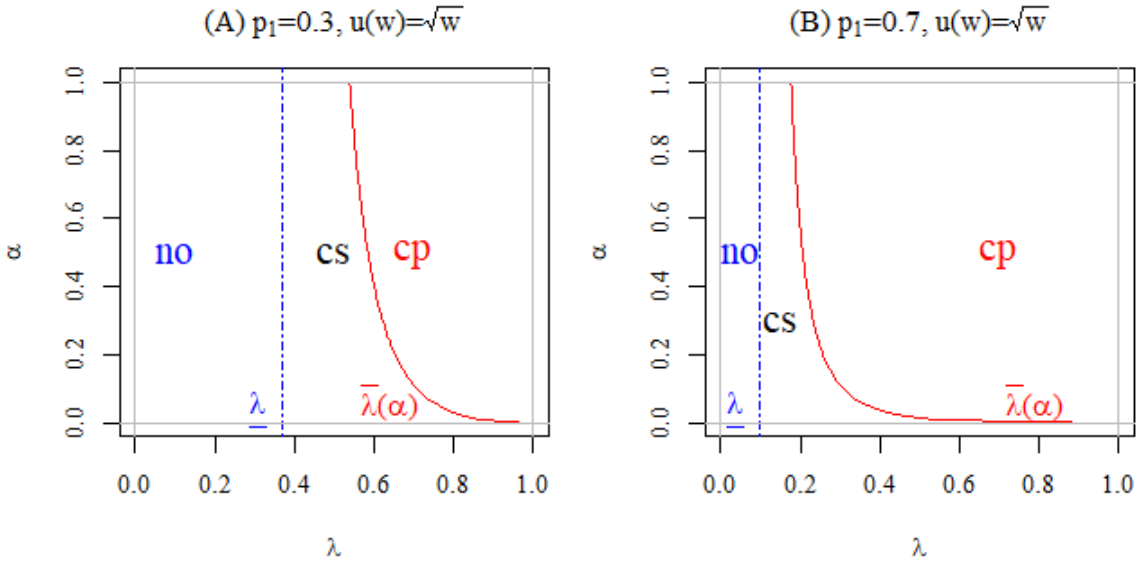

Figure 2: Illustration of the Cutoff Values of λ

Figure 2 further illustrates Propositions 5 and 6. In each diagram, the no-supervision contract is optimal in the left-hand area of the blue dashed line that depicts $\underline{\lambda}$, whereas the collusion-proof contract is optimal in the right-hand area of the red curve that depicts $\overline{\lambda}$. The dominance of the collusive-supervision contract occurs in the area between the blue dashed line and the red curve. Note that $\underline{\lambda}$ determines whether the principal should hire a possibly incorrect supervisor. As discussed above, once the supervisory technology is very likely to be inaccurate and the principal

decides not to use the supervisory signal, the payment scheme depends only on the output level. Given $\alpha \in (0,1)$, equation (7) indicates that the bargaining outcome in the side contract does not affect the comparison; therefore, $\underline{\lambda}$ is independent of α ($\underline{\lambda}$ is vertical in all diagrams).[15] In contrast, $\overline{\lambda}$ determines whether supervisor-agent collusion should be allowed conditional on the payment scheme being dependent on the supervisory signal. The bargaining outcome in the side contract affects the level of $\overline{\lambda}$; see equation (8). This reflects the trade-off between inefficient supervision and supervisor-agent collusion. When the agent acquires a stronger bargaining power in the side contract, he can pay less bribe ($(1-\alpha)\check{w}_{cs}$) to manipulate the signal and shirk. As a result, the principal must pay more to incentivize the agent; thus, $\check{C}_{cs}$ increases in α. Hence, as α increases, the principal tends to adopt the collusion-proof contract.

4 Discussion

4.1 Single-agent hierarchy

We examine the case of single-agent hierarchy, which helps clarify why the underlying trade-off exists in a multiple-agent setting only. The detailed analysis is given in Appendix II. We find that if $\lambda \geqslant \lambda^*$, then collusion proofness is optimal; otherwise, the principal should not use the supervisory signal at all. This is because unlike signal $\theta = (0,0)$ in the two-agent hierarchy, no signal in the single-agent hierarchy is entirely driven by the mistaken supervision. Moreover, when designing the payment schemes, the principal does not need to consider the negative externality caused by a sub-coalition. Therefore, the trade-off between inefficient supervision and collusion prevention disappears, and we return to the classic collusion-proof results in Tirole (1992). This further implies that if the two agents collude by coordinating their efforts and acting as a single agent in production, then the supervisor-agent coalition should be prevented in the hierarchy.

4.2 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 the payment structure below, where the *ex post* payments in Proposition 4 are offered directly by the principal. This payment structure creates the possibility of a sub-coalition: Because $\check{w}_{cs}^H + 0 < \check{w}_{cs}^H + (1-\alpha)\check{w}_{cs}^H$ for all $\alpha \in (0,1)$, when the supervisor observes a signal $(0,0)$, she has an incentive to collude with agent B and report $(0,1)$; this lowers the non-colluding agent A's payoff from $\check{w}_{cs}^H$

[15] Note that if the agent has the full bargaining power, i.e., $\alpha = 1$, then the optimal collusive-supervision contract is equivalent to the optimal no-supervision contract.

to $\alpha\check{w}_{cs}^H$, which jeopardizes his incentive to work. The collusion-proofness principle fails because of the negative externality of the sub-coalition.[16]

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

4.3 Uninformative signal

We here investigate the possibility that an uninformative signal about the agents' effort levels, which is denoted by $\varnothing$, can be observed in the hierarchy (Tirole, 1986). Specifically, we consider the following two cases.

Case 1. We examine the variant that the state space becomes $\Theta \equiv \{(1,1),(0,1),(1,0),(0,0),\varnothing\}$. If the supervisor is inefficient (with probably $1-\lambda$), then she observes a random signal with an equal probability of $1/5$. When signal $\varnothing$ is observed, a supervisor-agent coalition can report any of the other four signals. Our analysis in Appendix II shows that with the low output, it is still optimal to reward both agents zero regardless of the signal. With output $y = H$ and signal $\varnothing$, the payment structure is the same as under signal $(0,0)$; that is, $w(1,1) = w(0,0) = w(\varnothing)$. Thus, the possibility of full-coalition under signal $\varnothing$ is prevented. Furthermore, to prevent the negative externality, the principal should reward the supervisor zero payoff across all five signals. As a result, within a certain range of λ, allowing a sub-coalition that manipulates signal $(0,1)$ to report $(1,1)$ improves agent A's incentive to work.

Case 2. Suppose that the supervisor observes either the true signal with probability λ or the uninformative signal $\varnothing$ with probably $(1-\lambda)$. With this supervisory technology, signals $(0,1)$, $(1,0)$, and $(0,0)$ are off the equilibrium path, and the payments to the supervisor after these signals do not appear in the principal's objective function. In Appendix II, we show that the collusive-supervision contract is equivalent to the collusion-free contract in which the principal only pays the agents with signals $(1,1)$ and $\varnothing$. In this case, all collusion possibilities can be deterred without any cost after observing these signals, and thus, there is no need for the principal to consider the

[16] One might inquire whether a cross-checking mechanism (Baliga, 1999) can help eliminate the negative externality. In Appendix II, we explore such a possibility and show that the principal can achieve a lower cost than that of the optimal collusive-supervision contract. However, in many complicated tasks, such as medical research and new product development, each agent is usually responsible for a small component of the task and lacks sufficient knowledge and information to evaluate the performance of others. Thus, in practice, it is difficult to conduct a cross-checking mechanism in a multiple-agent environment.

negative externality from a sub-coalition. As a result, the collusive-supervision contract dominates the collusion-proof contract. This indicates that the novel trade-off that we examine in the paper is rooted in inefficient supervisory technology that possibly observes incorrect signals.

4.4 Production uncertainty

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. Here, we examine how the production uncertainty affects our results in the current multiple-agent hierarchy. When $0 < p(1,1) < 1$, a possibility exists that both agents have exerted effort but that the production yields a lower output. Propositions 7 and 8 in Appendix II show that it is still beneficial to allow supervisor-agent collusion when supervisory efficiency is at an intermediate level. This implies that our main result is robust to the setting of production uncertainty.

Next, let us consider an extreme case where $\lambda = 1$ 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. This case demonstrates that the failure of the collusion-proof principle in Tirole (1986) is rooted in inefficient supervision in the multiple-agent hierarchy. The trade-off between inefficient supervision and collusion prevention identified in this paper is novel.

5 Conclusion

We study a three-level hierarchy with multiple agents and a possibly inefficient supervisor. In the hierarchy, the supervisor and the agents can collude to forge the supervisory report to benefit themselves. In this study, we provide novel insights into the trade-off between inefficient supervision and supervisor-agent collusion. Allowing a sub-coalition that permits a revision of the incorrect supervisory signal rooted in inefficient supervisory technology provides higher incentives for agents to work. We further provide a full solution of the principal's contract design problem under different levels of supervisory efficiency, which shows that the collusive-supervision contract dominates both the non-supervision contract and the collusion-proof contract when supervisory efficiency is at an intermediate level.

Appendix I: Essential Proofs

Proof of Lemma 2. *Part (a).* Let us assume that there exists a signal τ such that $T^y(\tau) > T^y(1,1)$. This implies that when another signal rather than τ is observed, the supervisor and the two agents will collude to report τ. Note that we here allow a downward adjustment of an agent's signal if it is beneficial to do so. Taking agent A as the representative, the IC constraint to implement $e = (1,1)$ can be written as

$$\begin{aligned}
&\lambda u(w_A^H(\tau|11)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(\tau|11)) + u(w_A^H(\tau|10)) + u(w_A^H(\tau|01)) + u(w_A^H(\tau|00))\Big] - \varphi \\
&\geqslant p_1\left\{\lambda u(w_A^H(\tau|01)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^H(\tau|11)) + u(w_A^H(\tau|10)) + u(w_A^H(\tau|01)) + u(w_A^H(\tau|00))\Big]\right\} \\
&+ (1-p_1)\left\{\lambda u(w_A^L(\tau|01)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^L(\tau|11)) + u(w_A^L(\tau|10)) + u(w_A^L(\tau|01)) + u(w_A^L(\tau|00))\Big]\right\}.
\end{aligned} \tag{10}$$

We rewrite (10) in the form of $Z'_{cp} \geqslant 0$, where

$$\begin{aligned}
Z'_{cp} = &\Big(\lambda + \frac{1}{4}(1-\lambda)(1-p_1)\Big)u(w_A^H(\tau|11)) + (1-\lambda)(1-p_1)u(w_A^H(\tau|10)) \\
&+ (1-\lambda)(1-p_1)u(w_A^H(\tau|00)) - \Big(p_1\lambda - (1-\lambda)(1-p_1)\Big)u(w_A^H(\tau|01)) \\
&- (1-p_1)\left\{\lambda u(w_A^L(\tau|01)) + (1-\lambda)\frac{1}{4}\Big[u(w_A^L(\tau|11)) + u(w_A^L(\tau|10)) + u(w_A^L(\tau|01)) + u(w_A^L(\tau|00))\Big]\right\} \\
&- \varphi.
\end{aligned} \tag{11}$$

To incentivize the agent, when $y = L$, it is optimal for the principal to reward nothing to both agents and the supervisor. Then, (11) can be rewritten as

$$\begin{aligned}
Z'_{cp} = &\Big(\lambda + \frac{1}{4}(1-\lambda)(1-p_1)\Big)u(w_A^H(\tau|11)) + (1-\lambda)(1-p_1)u(w_A^H(\tau|10)) \\
&+ (1-\lambda)(1-p_1)u(w_A^H(\tau|00)) - \Big(p_1\lambda - (1-\lambda)(1-p_1)\Big)u(w_A^H(\tau|01)) - \varphi.
\end{aligned} \tag{12}$$

Given $\lambda > \underline{\lambda}$, this implies that $(\lambda p_1 - (1-\lambda)(1-p_1)) > 0$. Now, suppose that (12) is binding, that is, $Z'_{cp} = 0$. Since $w_A^H(11|01) < w_A^H(\tau|01)$, we have

$$\begin{aligned}
Z'_{cp} = 0 \;<\; &Z''_{cp} \\
= \;&\Big(\lambda + \frac{1}{4}(1-\lambda)(1-p_1)\Big)u(w_A^H(\tau|11)) + (1-\lambda)(1-p_1)u(w_A^H(\tau|10)) \\
+ \;&(1-\lambda)(1-p_1)u(w_A^H(\tau|00)) - \Big(p_1\lambda - (1-\lambda)(1-p_1)\Big)u(w_A^H(11|01)) - \varphi.
\end{aligned} \tag{13}$$

Clearly, we have $u(w_A^H(1,1)) < u(w_A^H(\tau|11))$, $u(w_A^H(11|10)) < u(w_A^H(\tau|10))$, and $u(w_A^H(11|00)) <$

$u(w_A^H(\tau|00))$. To make $Z''_{cp}=0$, we can lower the positive terms in Z''_{cp}, which gives

$$
\begin{aligned}
Z''_{cp} = &\Big(\lambda+\frac{1}{4}(1-\lambda)(1-p_1)\Big)u(w_A^H(1,1))+(1-\lambda)(1-p_1)u(w_A^H(11|10))\\
&+(1-\lambda)(1-p_1)u(w_A^H(11|00))-\Big(p_1\lambda-(1-\lambda)(1-p_1)\Big)u(w_A^H(11|01))-\varphi=0
\end{aligned}
\tag{14}
$$

In this way, the principal can satisfy the IC constraint by paying less. This indicates that when $\lambda>\underline{\lambda}$, the principal cannot do better than rewarding $T^y(1,1)\geqslant T^y(\tau)$ to implement $e=(1,1)$.

Part (b). With output L, from Part (a), we know that it is optimal for the principal not to reward the supervisor. With $y=H$ and $\theta=(1,1)$, both agents have no incentives to collude; therefore, there is no need to reward the supervisor when she reports $r=(1,1)$, that is, $s^H(1,1)=0$ when output H is observed. □

Proof of Proposition 4. Before setting up the Lagrangian for the optimization problem, we rewrite (IC_{cs}^A) in the form of $Z_{cs}\geqslant 0$, where

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

We then examine the maximization of Z_{cs} 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.

Step 1. Since the sign of Term 5 is negative, the principal should minimize Term 5 by setting all payments associated with output L to zero.

Result 1. It is optimal for the principal not to reward the agents for $y=L$, i.e., $w_i^L(r)=0,\ \ \forall\, r\in\Theta$, $i=A,B$.

Step 2. 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 $\breve{w}_{cs}^H>0$ as the payment to agent A that implements $e=(1,1)$, which is determined by the principal's cost minimization problem or, specifically, the binding equilibrium IC constraint.

Result 2. $w_A^H(1,1)=w_B^H(1,1)=\breve{w}_{cs}^H>0$.

Step 3. Consider Term 2 associated with signal $\theta = (1,0)$. Apparently, since $\frac{1}{4}(1-p_1)(1-\lambda) \geqslant 0$, 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.

Result 3. It is optimal to reward an agent if his own signal is positive, regardless of the other agent's signal, i.e., $w_A^H(1,1) = w_A^H(1,0) = \check{w}_{cs}^H > 0$ and $w_B^H(1,1) = w_B^H(0,1) = \check{w}_{cs}^H > 0$.

This indicates that agent A will not have the incentive to collude given $\theta = (1,0)$. However, note that according to Lemma 2, $T(1,0) \leqslant T(1,1)$, 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 $r = (1,1)$.

Step 4. We now turn to Term 3. 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, namely, $r = (1,1)$, $(1,0)$, and $(0,0)$, which correspond 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 $T(1,1) > T(0,0)$, then the report is $r = (1,1)$, and the total payment is $T^H(1,1) = w_A^H(1,1) + w_B^H(1,1) = 2\check{w}_{cs}^H$. The 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; therefore, $w_A^H(11|00) < \check{w}_{cs}^H$. 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, the payment structure is the same across signals $(1,1)$ and $(0,0)$; this corrects the wrong signal from the inefficient supervisor, and the full-coalition among the agents and the supervisor is prevented.

Second, consider the sub-coalition case. If a sub-coalition with agent B is allowed, then $w_B^H(0,1) + s(0,1) > w_B^H(0,0) + s(0,0) = \check{w}_{cs}^H$ must hold. From Lemma 2, $T^H(0,1) \leqslant T^H(0,0) = 2\check{w}_{cs}^H$. These two inequalities yield the following:

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

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

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

Result 4. (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, $s^H(1,0) = s^H(0,1) = 0$.

Step 5. Finally, we consider Term 4, where there are two possible reports, specifically, $r = (1,1)$ and $(0,1)$, corresponding to a sub-coalition and truthful reporting, respectively. The coefficient in front of $w_A^H(r|01)$ can be positive or negative. Given $\underline{\lambda} \equiv \frac{1-p_1}{1+3p_1}$, apparently, if $\lambda \leqslant \underline{\lambda}$, since $\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1 \geqslant 0$, then a higher $w_A^H(r|01)$ provides a greater incentive for agent A to work. If, however, $\lambda > \underline{\lambda}$, then 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 Result 3, $w_B^H(0,1) = \check{w}_{cs}^H$. From Result 2, 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 $w_A^H(01|01) = \check{w}_{cs}^H$ implies that $w_A^H(0,1) = \check{w}_{cs}^H$ provides a greater incentive for agent A to work. Furthermore, by combining the results in Results 1-4, we can simplify function Z_{cs}:

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

Given that $w_A^H(1,1) = w_A^H(1,0) = w_A^H(0,1) = w_A^H(0,0) = \check{w}_{cs}^H$, we have $Z_{cs} = (1-p_1)u(\check{w}_{cs}^H) - \varphi$. The equilibrium IC constraint, $(1-p_1)u(\check{w}_{cs}^H) - \varphi = 0$, is the same as ($\widehat{IC}_{no}$) for the no-supervision contract, which implies that $\check{w}_{cs}^H = \hat{w}_{no}^H$. 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 $w_A^H(r|01)$ increases agent A's incentive to work. Under the sub-coalition, $w_A^H(11|01) + s^H(11|01) = w_A^H(1,1) + s^H(1,1) = \check{w}_{cs}^H$, where $s^H(1,1) = 0$ according to Lemma 2. To form the coalition, the supervisor must receive $s^H(11|01) > 0$; thus, agent A receives $w_A^H(11|01) < \check{w}_{cs}^H$. As collusion can lower $w_A^H(r|01)$, 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 $\check{w}_{cs}^H$. The discussion above yields the following result.

Result 5. (a) If $\lambda \leqslant \underline{\lambda}$, it is optimal to reward the agent regardless of the signal, i.e., $w_A^H(1,1) = w_A^H(1,0) = w_A^H(0,1) = w_A^H(0,0) = \check{w}_{cs}^H > 0$ for agent A. The contract is equivalent to the no-supervision contract; (b) if $\lambda > \underline{\lambda}$, then for agent A, $w_A^H(1,1) = w_A^H(1,0) = w_A^H(0,0) > 0$ and $w_A^H(0,1) = 0$, but for agent B, $w_B^H(1,1) = w_B^H(0,1) = w_B^H(0,0) > 0$ and $w_B^H(1,0) = 0$. In the

payment structure, a sub-coalition to manipulate the signal from $(0,1)$ or $(1,0)$ to $(1,1)$ is allowed.

From the analysis above, given $\alpha \in (0,1)$ and $\lambda > \underline{\lambda}$, let us set up the Lagrangian for (P_{cs}). Since the principal does not reward the two agents when the output is L, it is optimal to set $s^L(r) = 0$ for all r. We can rewrite the objective function C_{cs} as follows:

$$\begin{aligned}
C_{cs} &= \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(r|10) + T^H(r|01) + T^H(r|00)\Big] \\
&= \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(11|10) + T^H(11|01) + T^H(0,0)\Big] \\
&= \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(1,1) + T^H(1,1) + T^H(0,0)\Big] \\
&= 2\lambda w^H(1,1) + (1-\lambda)\frac{1}{4}\Big[2w^H(1,1) + 2w^H(1,1) + 2w^H(1,1) + 2w^H(0,0)\Big] \\
&= 2w^H(1,1).
\end{aligned}$$

Then, the Lagrangian for (P_{cs}) is given by

$$\begin{aligned}
\mathcal{L} = 2w_A^H(1,1) - \delta\Bigg\{ &\lambda\Big[u(w_A^H(1,1)) - p_1 u(\alpha w_A^H(1,1))\Big] \\
&+ (1-\lambda)(1-p_1)\Big[\frac{3}{4}u(w_A^H(1,1)) + \frac{1}{4}u(\alpha w_A^H(1,1))\Big] - \varphi\Bigg\}.
\end{aligned}$$

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

$$\begin{aligned}
\text{(A1)}: \quad \frac{\partial \mathcal{L}}{\partial w_A^H(1,1)} = 2 - \delta\Bigg\{ &\lambda\Big[u'(w_A^H(1,1)) - p_1\alpha u'(\alpha w_A^H(1,1))\Big] \\
&+ (1-\lambda)(1-p_1)\Big[\frac{3}{4}u'(w_A^H(1,1)) + \frac{1}{4}\alpha u'(\alpha w_A^H(1,1))\Big]\Bigg\} \geqslant 0, \\
w_A^H(1,1) \geqslant 0, \ \text{ and } \ & w_A^H(1,1)\frac{\partial \mathcal{L}}{\partial w_A^H(1,1)} = 0;
\end{aligned}$$

plus the complementary slackness conditions for the constraints.

Step 1. It is impossible to have $\delta = 0$ because this implies that $w_A^H(1,1) = 0$ in (A1), which violates (IC_{cs}^A) and yields a contradiction. We therefore have $\delta > 0$ and $w_A^H(1,1) > 0$.

Step 2. When (IC_{cs}^A) is binding, which is denoted by $(\widetilde{IC}_{cs})$, we have the value of $\check{w}_{cs}^H$ as follows:

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

Because of the symmetry of the two agents, the total payment of the principal is $\check{C}_{cs} = 2\check{w}_{cs}^H$. $\square$

Proof of Proposition 6. *Part (a)*: Differentiating $\overline{\lambda}$ with respect to α yields

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

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

Part (b): When $\alpha = 0$, $\overline{\lambda} = 1$ and ($\widetilde{IC}_{cs}$) 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 $\lambda \in [\lambda^*, 1]$. When $\alpha = 1$, $\overline{\lambda} = \lambda^*$ and ($\widetilde{IC}_{cs}$) can be written as

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

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 ($\widetilde{IC}_{cs}$) 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 $\kappa < 0$; therefore, an increase in bargaining power α leads to a higher $\check{w}_{cs}^H$ to satisfy ($\widetilde{IC}_{cs}$). □

References

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

Nash, J. F. (1950): "The Bargaining problem," *Econometrica: Journal of the Econometric Society*, 155–162.

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

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

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

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

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

——— (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*.

Appendix II: Nonessential Proofs and Supplemental Materials (For Online Publication)

Proof of Proposition 1. We focus on the symmetric equilibrium and therefore write $C_{no} = w_A^H + w_B^H = 2w_A^H$. Before setting up the Lagrangian for the optimization problem, we rewrite (IC_{no}^A) in the form of $Z_{no} \geqslant 0$, where

$$Z_{no} \equiv (1-p_1)u(w_A^H) - (1-p_1)u(w_A^L) - \varphi.$$

We examine the maximization of Z_{no} term-by-term by choosing the payments. This helps us determine the payment structure that the principal should offer to incentivize the agents.

a. Because the payment associated with $y = L$ has a negative sign, it is obvious that the principal should set all of the payments to zero to incentivize the agent. Therefore, we have $w_A^L = 0$.

b. Since $(1-p_1) \geqslant 0$, this indicates that, to incentivize agent A, the principal should reward positively. Therefore, it is optimal to set $w_A^H > 0$ to maximize Z_{cf}.

Then, the Lagrangian for (P_{no}) is given by

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

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

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

plus the complementary slackness conditions for the constraints.

Step 1. It is impossible to have $\delta = 0$ because this implies that $w_A^H = 0$ in (A1), which violates (IC_{no}^A) and yields 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. When (IC_{no}^A) is binding, which is 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 $\hat{C}_{no} = 2\hat{w}_{no}^H$. □

Proof of Proposition 2. Before setting up the Lagrangian for (P_{cf}), we rewrite (IC^A_{cf}) in the form of $Z_{cf} \geqslant 0$, where

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

We first examine the maximization of Z_{cf} term-by-term by choosing the payments. This helps us determine the payment structure that the principal should offer to incentivize the agents.

a. Because all the payments associated with $y = L$ have negative signs, it is obvious that the principal should set all of them to zero to incentivize the agent. Therefore, we have $w_i^L(r) = 0$ for all $r \in \Theta$, $i = A, B$.

b. Since $[\lambda + \frac{1}{4}(1-p_1)(1-\lambda)] \geqslant 0$ and $[\frac{1}{4}(1-p_1)(1-\lambda)] \geqslant 0$, this imply that to incentivize agent A, the principal should reward positively with signals of $(1,1)$, $(1,0)$, and $(0,0)$. Therefore, it is optimal to set $w_A^H(1,1) = w_A^H(1,0) = w_A^H(0,0) > 0$ to maximize Z_{cf}.

c. There is a unique cutoff, $\underline{\lambda} = \frac{1-p_1}{1+3p_1}$, that satisfies the equation

$$\frac{1}{4}(1-p_1)(1-\underline{\lambda}) - \underline{\lambda}p_1 = 0.$$

This is equation (5) in the main text that critically determines whether the signal is sufficiently accurate to be considered or not. When $\lambda \leqslant \underline{\lambda}$, we have $[\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1] \geqslant 0$, and it is optimal to reward agent A with signal of $(0,1)$. This implies that $w_A^H(1,1) = w_A^H(1,0) = w_A^H(0,0) = w_A^H(0,1) > 0$. This is equivalent to the no-supervision contract. When $\lambda > \underline{\lambda}$, this gives $[\frac{1}{4}(1-p_1)(1-\lambda) - \lambda p_1] < 0$, and therefore, it is optimal for the principal to reward zero with a signal of $(0,1)$, i.e., $w_A^H(0,1) = 0$.

Given the analysis above, let us consider $\lambda > \underline{\lambda}$ and set up the Lagrangian for (P_{cf}). Since the supervisor does not need to be incentivized to tell the truth, it is optimal for the principal to reward no payment $s^y(r) = 0$ for all y and r. We here focus on symmetric equilibrium; therefore

the objective function C_{cf} can be written as follows:

$$\begin{aligned} C_{cf} &= \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(1,0) + T^H(0,1) + T^H(0,0)\Big] \\ &= 2\lambda w_A^H(1,1) + (1-\lambda)\frac{1}{4}\Big[2w_A^H(1,1) + w_A^H(1,0) + w_A^H(1,0) + 2w_A^H(0,0)\Big] \\ &= 2\lambda w_A^H(1,1) + \frac{3}{2}(1-\lambda)w_A^H(1,1) \\ &= (\frac{3}{2} + \frac{1}{2}\lambda)w_A^H(1,1). \end{aligned}$$

Then, the Lagrangian for (P_{cf}) is given by

$$\mathcal{L} = (\frac{3}{2} + \frac{1}{2}\lambda)w_A^H(1,1) - \delta\left[\big[\lambda + \frac{3}{4}(1-\lambda)(1-p_1)\big]u(w_A^H(1,1)) - \varphi\right],$$

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

$$\begin{aligned} \text{(A1)}: \quad & \frac{\partial \mathcal{L}}{\partial w_A^H(1,1)} = (\frac{3}{2} + \frac{1}{2}\lambda) - \delta\big[\lambda + \frac{3}{4}(1-\lambda)(1-p_1)\big]u'(w_A^H(1,1)) \geqslant 0, \\ & w_A^H(1,1) \geqslant 0, \ \text{ and } \ w_A^H(1,1)\frac{\partial \mathcal{L}}{\partial w_A^H(1,1)} = 0; \end{aligned}$$

plus the complementary slackness conditions for the constraints.

Step 1. In (A1), if $\delta = 0$, it implies that $\partial\mathcal{L}/\partial w_A^H(1,1) = \lambda + \frac{1}{4}(1-\lambda) > 0$ and $w_A^H(1,1) = 0$, which violates (IC_{cf}^A) and yields a contradiction. Therefore, we have $\partial\mathcal{L}/\partial w_A^H(1,1) = 0$, which indicates that $\delta > 0$ and $w_A^H(1,1) > 0$.

Step 2. When (IC_{cf}^A) is binding, which is denoted by $(\widetilde{IC}_{cf})$, we have the value of $\tilde{w}_{cf}^H$ as follows:

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

The expected cost of the principal is $\tilde{C}_{cf} = (\frac{3}{2} + \frac{1}{2}\lambda)\tilde{w}_{cf}^H$.

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

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

Furthermore, (CIC_s) indicates that $w_B^y(1,1) + s^y(1,1) = w_B^y(1,0) + s^y(1,0)$. Therefore, $w_A^y(1,0) = w_A^y(1,1)$. Furthermore, (CIC_f) requires $T^y(0,1) = T^y(0,0)$; 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 3. Before setting up the Lagrangian for (C_{cp}), given Lemma 1, we rewrite IC_{cp} in the form of $Z_{cp} \geqslant 0$, where

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

We first examine the maximization of Z_{cp} term-by-term by choosing the payments. This helps us determine the payment structure that the principal should offer to incentivize the agents.

a. Because all the payments associated with $y = L$ have negative signs, it is obvious that the principal should set all of them to zero to incentivize the agent. Therefore, we have $w_i^L(r) = 0$ for all $r \in \Theta$, $i = A, B$.

b. There is a unique cutoff, $\lambda^* = \frac{1-p_1}{1+p_1}$, that satisfies the equation:

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

When $\lambda \leqslant \lambda^*$, $\frac{1}{2}(1-\lambda)(1-p_1) - \lambda p_1 \geqslant 0$, and it is optimal to reward agent A with a signal of $(0,0)$ (or a signal of $(0,1)$). This implies that $w_A^H(1,1) = w_A^H(1,0) = w_A^H(0,0) = w_A^H(0,1) > 0$. This is equivalent to the no-supervision contract. When $\lambda > \lambda^*$, this gives $\frac{1}{2}(1-\lambda)(1-p_1) - \lambda p_1 < 0$, and therefore, it is optimal to reward zero with a signal of $(0,0)$ (or a signal of $(0,1)$), i.e. $w_A^H(0,0) = 0$.

Assuming that $\lambda > \lambda^*$, let us set up the Lagrangian for (C_{cp}). Given that we focus on symmetric equilibrium, the objective function C_{cp} can be written as follows:

$$\begin{aligned} C_{cf} &= \lambda T^H(1,1) + (1-\lambda)\frac{1}{4}\Big[T^H(1,1) + T^H(1,0) + T^H(0,1) + T^H(0,0)\Big], \\ &= \lambda\Big[2w_A^H(1,1) + s^H(1,1)\Big] + (1-\lambda)\Big[2w_A^H(1,1) + s^H(1,1)\Big] \\ &= 2w_A^H(1,1) + s^H(1,1). \end{aligned}$$

The Lagrangian is given by

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

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

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

$$\text{(A2)}: \quad \frac{\partial \mathcal{L}}{\partial s^H(1,1)} = 1 \geqslant 0,\ s^H(1,1) \geqslant 0 \text{ and } s^H(1,1)\frac{\partial \mathcal{L}}{\partial s^H(1,1)} = 0;$$

$$\text{(A3)}: \quad \frac{\partial \mathcal{L}}{\partial w_A^H(0,0)} = -\delta\big[\frac{1}{2}(1-\lambda)(1-p_1) - \lambda p_1\big]u'(w_A^H(0,0)) \geqslant 0,$$
$$w_A^H(0,0) \geqslant 0 \text{ and } w_A^H(0,0)\frac{\partial \mathcal{L}}{\partial w_A^H(0,0)} = 0;$$

plus the complementary slackness conditions for the constraints.

Step 1. It is impossible to have $\delta = 0$ because this implies that $\partial\mathcal{L}/\partial w_A^H(1,1) = 2 > 0$ and $w_A^H(1,1) = 0$ in (A1), which violates (IC_{cp}) and yields a contradiction. Therefore, we should have $\delta > 0$ and $w_A^H(1,1) > 0$, which implies that $\partial\mathcal{L}/\partial w_A^H(1,1) = 0$ and $\delta = 2/\big[\lambda + \frac{1}{2}(1-\lambda)(1-p_1))u'(w_A^H(1,1)\big]$. According to Lemma 1, we further have $w_A^H(1,1) = w_A^H(1,0) > 0$.

Step 2. From (A2), we clearly have $s^H(1,1) = 0$.

Step 3. From (A3), when $\lambda > \lambda^*$, $\partial\mathcal{L}/\partial w_A^H(0,0)$ is strictly positive, which implies that $w_A^H(0,0) = 0$. According to Lemma 1, we also have $w_A^H(0,1) = 0$. Furthermore, satisfying (CIC_f) and (CIC_s) implies that $s^H(0,0) = 2w_A^H(1,1)$ and $s^H(1,0) = s^H(0,1) = w_A^H(1,1)$.

Step 4. Finally, we denote $w_A^H(1,1)$ in equilibrium by $\tilde{w}_{cp}^H$, which is uniquely determined by (IC_{cp}):

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

Because of the symmetry of the two agents, the principal incurs a total cost of $\tilde{C}_{cp} = 2\tilde{w}_{cp}^H$. □

S1: Single-agent Hierarchy.

We modify the model in the following way. There is only one productive agent in the hierarchy. If the agent works, then the probability of producing output $y = H$ is 1. However, if the agent shirks, then the probability of $y = H$ is $p \in (0, 1)$. 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 payment to the agent when the signal is 0 but the report is 1, and let $w^y(0|0)$ denote the agent's payoff under truthful reporting.

We here examine whether allowing collusion improves the agents' incentives to work. The IC constraint can then be written as follows:

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

which can be rewritten as

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

First, to provide incentives for the agent to work, it is optimal for the principal not to reward the agent when $y = L$. Second, if $\lambda \leqslant \lambda^*$, then we have $\frac{1}{2}(1-\lambda)(1-p) - \lambda p \geqslant 0$, which implies that truthful reporting (where $w^H(0|0) = w^H(1)$ to the agent and $s^H(0|0) = 0$ to the supervisor) generates a higher incentive to work than collusion (where $w^H(1|0) < w^H(1)$ to the agent and $s^H(0|0) > 0$ to the supervisor). Plugging $w^H(0|0) = w^H(1)$ into the IC constraint implies that the collusive-supervision contract is equivalent to the no-supervision contract. If $\lambda > \lambda^*$, then we have $\frac{1}{2}(1-\lambda)(1-p) - \lambda p < 0$, which indicates that truthful reporting (where $w^H(0|0) = 0$ to the agent and $s^H(0|0) = w^H(1)$ to the supervisor) generates a higher incentive to work than collusion (where $w^H(1|0) > 0$ to the agent and $s^H(0|0) > 0$ to the supervisor). Plugging $w^H(0|0) = 0$ into the IC constraint implies that the collusive-supervision contract is equivalent to the collusion-proof contract. Thus, we can conclude that it is not beneficial to allow supervisor-agent collusion in a single-agent setting. □

S2: Cross-checking Mechanism.

Ideally, if there were no negative externality, then the principal would reward the supervisor positively to prevent the sub-coalition from manipulating signal $(0,1)$ (or $(1,0)$) to $(1,1)$, which would provide higher incentives for the agents to work. We call such a contract the *superior collusion-proof (scp) contract*. Let us denote the equilibrium payment by $\check{w}^H_{scp}$ in the contract, and the payment structure is given as follows:

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

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

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

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

We here explore the implementation of the superior collusion-proof contract by allowing both the supervisor and the agents to submit their reports. This is called the cross-checking mechanism introduced by Baliga (1999). In reality, this setting resembles the condition in which the principal has a direct communication channel with the agents. After observing θ, agent i and the supervisor make their own reports r_i and r_s, respectively, where $i = A,\ B$. The cross-checking mechanism is implemented as follows:

(a) If $r_A = r_B = r_s = (0,0)$ or $(1,1)$, then both agents are rewarded $\check{w}^H_{scp}$, and the supervisor is not rewarded.

(b) If $r_A = r_B = r_s = (0,1)$ or $(1,0)$, then the agent with a signal of 1 and the supervisor obtain rewards $\check{w}^H_{scp}$, and the agent with a signal of 0 receives nothing.

(c) If the reports are not the same, then all the parties receive no reward.

The cross-checking mechanism characterized above helps eliminate the possible sub-coalition that manipulates the signal from $(0,0)$ to $(0,1)$ (or $(1,0)$) and its associated negative externality. This is because if the supervisor colludes with one of the agents, then this will induce the colluding members and the non-colluding member to submit different reports, which leads to zero payment to all the parties. By comparing $(\widetilde{IC}_{cs})$ and $(\widetilde{IC}_{scp})$, it is easy to see that the superior collusion-proof contract dominates the collusive-supervision contract when $\underline{\lambda} < \lambda < \overline{\lambda}$. □

S3: Uninformative Signal.

Case 1. Given the possibility of observing an uninformative signal, we can then write the IC constraint of the representative agent A as follows:

(IC_{un})

$$\lambda u(w_A^H(1,1)) + (1-\lambda)\frac{1}{5}\Big[u(w_A^H(1,1)) + u(w_A^H(r|10)) + u(w_A^H(r|01)) + u(w_A^H(r|00)) + u(w_A^H(r|\varnothing))\Big] - \varphi$$
$$\geqslant p_1\Bigg\{\lambda u(w_A^H(r|01)) + (1-\lambda)\frac{1}{5}\Big[u(w_A^H(1,1)) + u(w_A^H(r|10)) + u(w_A^H(r|01)) + u(w_A^H(r|00)) + u(w_A^H(r|\varnothing))\Big]\Bigg\}$$
$$+ (1-p_1)\Bigg\{\lambda u(w_A^L(r|01)) + (1-\lambda)\frac{1}{5}\Big[u(w_A^L(1,1)) + u(w_A^L(r|10)) + u(w_A^L(r|01)) + u(w_A^L(r|00)) + u(w_A^L(r|\varnothing))\Big]\Bigg\}.$$

By rearranging the constraint (IC_{un}), we have

$$\Big[\lambda + \frac{1}{5}(1-\lambda)(1-p_1)\Big]u(w_A^H(1,1)) + \Big[\frac{1}{5}(1-\lambda)(1-p_1) - p_1\lambda\Big]u(w_A^H(r|01))$$
$$+ \frac{1}{5}(1-\lambda)(1-p_1)\Big[u(w_A^H(r|10)) + u(w_A^H(r|00)) + u(w_A^H(r|\varnothing))\Big] - (1-p_1)\Bigg\{\lambda u(w_A^L(r|01))$$
$$+ \frac{1}{5}(1-\lambda)\Big[u(w_A^L(1,1)) + u(w_A^L(r|10)) + u(w_A^L(r|01)) + u(w_A^L(r|00)) + u(w_A^L(r|\varnothing))\Big]\Bigg\} - \varphi \geqslant 0.$$

When $y = L$, it is optimal to reward zero to both agents and the supervisor across all the signals. With $y = H$, given $\frac{1}{5}(1-\lambda)(1-p_1) > 0$, truthful reporting (where $u(w_A^H(10|10)) = u(w_A^H(00|00)) = u(w_A^H(\varnothing|\varnothing)) = u(w_A^H(1,1)) > 0$) gives a higher incentive for agent A to work than allowing collusion. Furthermore, to prevent the negative externality from a sub-coalition, the supervisor is rewarded with zero payoff across the five signals, $s^H(r) = 0$. Finally, let $\underline{\lambda} \equiv \frac{1-p_1}{1+4p_1}$. When signal $(0,1)$ is observed, if $\lambda > \underline{\lambda}$, then $\frac{1}{5}(1-\lambda)(1-p_1) - p_1\lambda < 0$, and allowing a sub-coalition in which $u(w_A^H(11|01)) < u(w_A^H(1,1))$ improves agent A's incentive to work. If, however, $\lambda \geqslant \underline{\lambda}$, then $\frac{1}{5}(1-\lambda)(1-p_1) - p_1\lambda \geqslant 0$, and rewarding $u(w_A^H(01|01)) = u(w_A^H(1,1))$ is optimal (that is equivalent to the no-supervision contract).

Case 2. The inefficient supervisor always observes $\varnothing$. Let us first examine the collusion-free contract and the collusion-proof contract. Given the supervisory technology, the IC constraint is the same for the two contracts given by

$$\lambda u(w_A^H(1,1)) + (1-\lambda)u(w_A^H(\varnothing)) - \varphi$$
$$\geqslant p_1\Big[\lambda u(w_A^H(0,1)) + (1-\lambda)u(w_A^H(\varnothing))\Big] + (1-p_1)\Big[\lambda u(w_A^L(0,1)) + (1-\lambda)u(w_A^L(\varnothing))\Big].$$

Rearranging the constraint, we have

$$\begin{aligned}(15)\qquad &\lambda u(w_A^H(1,1)) - p_1\lambda u(w_A^H(0,1)) + (1-\lambda)(1-p_1)u(w_A^H(\varnothing)) \\ &- (1-p_1)\Big[\lambda u(w_A^L(0,1)) + (1-\lambda)u(w_A^L(\varnothing))\Big] - \varphi \geqslant 0.\end{aligned}$$

With the honest supervision, when $y = L$, it is optimal to reward zero to both agents and the supervisor across all the signals. When $y = H$, given $\lambda > 0$ and $(1-\lambda)(1-p_1) > 0$, to incentivize the agent, the payment structure should be that $w_A^H(1,1) = w_A^H(\varnothing) > 0$ and $w_A^H(0,1) = 0$. Therefore, in equilibrium, the principal pays the agent with signals $(1,1)$ and $(\varnothing)$, and reward zero with other signals. The equilibrium IC is written as $\lambda u(w_h^H) + (1-\lambda)(1-p_1)u(w_h^H) - \varphi = 0$. Since the supervisor does not obtain any payments from the principal, the aggregate payment is given by $C_h = \lambda 2 w_A^H(1,1) + (1-\lambda)2w_A^H(\varnothing) = \lambda 2 w_h^H + (1-\lambda)2w_h^H = 2w_h^H$. See the payment structure below.

Report r	Agent A	Agent B	Supervisor S
$(1,1)$	w_h^H	w_h^H	0
$(1,0)$	0	0	0
$(0,1)$	0	0	0
$(0,0)$	0	0	0
$(\varnothing)$	w_h^H	w_h^H	0

We next consider the collusion-proof contract. Both agents are paid, and the supervisor obtains zero when signal $(1,1)$ is observed. To report truthfully (to satisfy the CIC constraint), the supervisor is paid $s^H(0,0) = s^H(\varnothing) = w_A^H(1,1) + w_B^H(1,1)$ when signals $(1,1)$ and $(\varnothing)$ are observed, and $s^H(1,0) = w_B^H(1,1)$ and $s^H(0,1) = w_A^H(1,1)$ when signals $(1,0)$ and $(0,1)$ are observed. The equilibrium IC is written as $\lambda u(w_c^H) - \varphi = 0$. Since the supervisor does not obtain any payments from the principal, the aggregate payment is given by $C_c = \lambda 2 w_A^H(1,1) + (1-\lambda)2w_A^H(\varnothing) = \lambda 2 w_c^H + (1-\lambda)2w_c^H = 2w_c^H$. See the payment structure below.

Report r	Agent A	Agent B	Supervisor S
$(1,1)$	w_c^H	w_c^H	0
$(1,0)$	w_c^H	0	w_c^H
$(0,1)$	0	w_c^H	w_c^H
$(0,0)$	0	0	$2w_c^H$
$(\varnothing)$	0	0	$2w_c^H$

Comparing the IC constraints of the two contracts shows that $w_c^H > w_h^H$ and therefore $C_c > C_h$.

Now let us consider the case where collusion is allowed. (15) can be rewritten as follows:

$$\begin{aligned}(16)\qquad &\lambda u(w_A^H(1,1)) - p_1\lambda u(w_A^H(r|01)) + (1-\lambda)(1-p_1)u(w_A^H(r|\varnothing)) \\ &- (1-p_1)\Big[\lambda u(w_A^L(r|01)) + (1-\lambda)u(w_A^L(r|\varnothing))\Big] - \varphi \geqslant 0.\end{aligned}$$

When $y = L$, it is optimal to reward zero to both agents and the supervisor across all the signals. With $y = H$, since $-p_1\lambda < 0$, it is optimal to set $u(w_A^H(01|01)) = 0$. Furthermore, given $(1-\lambda)(1-p_1) > 0$, truthful reporting (where $u(w_A^H(\varnothing|\varnothing)) = u(w_A^H(1,1)) > 0$) gives a higher incentive for agent A to work than allowing collusion. Then, a full-coalition is prevented in the contract. As the signals $(1,0)$, $(0,1)$, and $(0,0)$ are off the equilibrium path, collusion possibilities can be deterred without any cost after observing these signals, and thus, there is no need for the principal to consider the negative externality from a sub-coalition. As a result, when collusion is allowed, there will be no scope for collusion and the contract is equivalent to the collusion-free contract characterized above. That is, the equilibrium IC is written as $\lambda u(w_h^H) + (1-\lambda)(1-p_1)u(w_h^H) - \varphi = 0$ and $C_h = \lambda 2 w_A^H(1,1) + (1-\lambda)2w_A^H(\varnothing) = 2w_h^H$. □

S4: Production Uncertainty.

In this section, 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 the following, we provide the IC constraints for the no-supervision, collusion-proof, and collusive-supervision contracts, and we show that allowing collusion can improve the agents' incentives to work. For the no-supervision contract, (IC_{no}^A) is replaced by the following equation:

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

For the collusion-proof contract, the IC constraint can be rewritten as follows:

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

For the collusive supervision, (IC^A_{cs}) is given by

$$
\begin{aligned}
& p_2\Big\{\lambda u(w_A^H(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^H(1,1) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\Big]\Big\} \\
& + (1-p_2)\Big\{\lambda u(w_A^L(1,1)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00)))\Big]\Big\} - \varphi \\
(IC'_{cs}) \qquad & \geqslant p_1\Big\{\lambda u(w_A^H(11|01) + (1-\lambda)\Big[\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\Big]\Big\} \\
& + (1-p_1)\Big\{\lambda u(w_A^L(11|01)) + (1-\lambda)\Big[\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00))\Big]\Big\}.
\end{aligned}
$$

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 7. *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

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

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

$$
\begin{aligned}
X = {} & \lambda(1-p_2)u(w_A^L(1,1)) - \lambda(1-p_1)u(w_A^L(0,1)) \\
& - (p_2-p_1)(1-\lambda)\Big(\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(0,1)) + \frac{1}{4}u(w_A^L(0,0))\Big).
\end{aligned}
$$

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:

$$
\begin{aligned}
X \geqslant {} & \lambda(1-p_2)u(w_A^L(1,1)) - \lambda(1-p_1)u(w_A^L(11|01)) \\
& - (p_2-p_1)(1-\lambda)\Big(\frac{1}{2}u(w_A^L(1,1)) + \frac{1}{4}u(w_A^L(11|01)) + \frac{1}{4}u(w_A^L(11|00)\Big).
\end{aligned}
$$

The right-hand side of the inequality above comprises 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 8. *Given $p_2 \in (0,1)$, with $y = H$, unique cutoffs $\underline{\lambda} \in (0, \lambda^*)$ exist and $\overline{\lambda} \in (\lambda^*, 1)$, where $\lambda^* \equiv \frac{p_2 - p_1}{p_2 + p_1}$, such that if $\underline{\lambda} < \lambda < \overline{\lambda}$, then allowing collusion induces a higher incentive for the agents to work than the no-supervision and the collusion-proof supervision.*

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:

$$
\begin{aligned}
F(\lambda) &= \lambda p_2 u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(11|01)) \\
&\quad + (p_2 - p_1)(1-\lambda)\left(\frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) \\
&= \lambda p_2 u(w_A^H(1,1)) - \lambda p_1 u(w_A^H(11|01)) + (p_2 - p_1)(1-\lambda)\Bigg(u(w_A^H(1,1)) \\
&\quad - \frac{1}{2}u(w_A^H(1,1)) + \frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\Bigg) \\
&> (p_2 - p_1)u(w_A^H(1,1)) + (p_2 - p_1)(1-\lambda)\left(\frac{1}{4}u(w_A^H(11|01)) + \frac{1}{4}u(w_A^H(11|00))\right) \\
&\quad - \lambda p_1 u(w_A^H(11|01)).
\end{aligned} \tag{17}
$$

Given that $w_A^y(11|00) \geqslant w_A^y(11|01)$ and $u(.)$ is increasing and concave, if $\lambda = \lambda^*$, then 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 comprises the term associated with high output in the IC constraint (equation (IC'_{no})) of the no-supervision contract, which indicates 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. Therefore, 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, because of the continuity of $F(\lambda)$, a unique cutoff $\underline{\lambda} \in (0, \lambda^*)$ must exist such that if $\lambda = \underline{\lambda}$, then $w_A^H(1,1) = \hat{w}^H$.

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 $\lambda = \lambda^*$, then 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

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

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 contains 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 deriva-

tive 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. We 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 $\lambda = \overline{\lambda}$.

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

□

Numerical Example. Let $u(w) = \sqrt{w}$. By ($\widehat{IC}_{no}$), $(1-p_1)\sqrt{\hat{w}_{no}^H} = \varphi$,

$$\hat{w}_{no}^H = \left(\frac{\varphi}{1-p_1}\right)^2,\ \hat{C}_{no} = 2\left(\frac{\varphi}{1-p_1}\right)^2.$$

By ($\widetilde{IC}_{cf}$), $\lambda\sqrt{\tilde{w}_{cf}^H} + \frac{3}{4}(1-p_1)(1-\lambda)\sqrt{\tilde{w}_{cf}^H} = \varphi$,

$$\tilde{w}_{cf}^H = \left(\frac{\varphi}{\lambda + \frac{3}{4}(1-p_1)(1-\lambda)}\right)^2,\ \hat{C}_{cf} = (\frac{3}{2} + \frac{1}{2}\lambda)\left(\frac{\varphi}{\lambda + \frac{3}{4}(1-p_1)(1-\lambda)}\right)^2.$$

By ($\widetilde{IC}_{cp}$), $\lambda\sqrt{\tilde{w}_{cp}^H} + \frac{1}{2}(1-p_1)(1-\lambda)\sqrt{\tilde{w}_{cp}^H} = \varphi$,

$$\tilde{w}_{cp}^H = \left(\frac{\varphi}{\lambda + \frac{1}{2}(1-p_1)(1-\lambda)}\right)^2,\ \tilde{C}_{cp} = 2\left(\frac{\varphi}{\lambda + \frac{1}{2}(1-p_1)(1-\lambda)}\right)^2.$$

By ($\widetilde{IC}_{cs}$), $\lambda\big[\sqrt{\check{w}_{cs}^H} - p_1\sqrt{\alpha\check{w}_{cs}^H}\big] + (1-\lambda)(1-p_1)\big[\frac{3}{4}\sqrt{\check{w}_{cs}^H} + \frac{1}{4}\sqrt{\alpha\check{w}_{cs}^H}\big] = \varphi$,

$$\check{w}_{cs}^H = \left(\frac{\varphi}{\lambda(1-p_1\sqrt{\alpha}) + (1-\lambda)(1-p_1)(\frac{3}{4} + \frac{1}{4}\sqrt{\alpha})}\right)^2,$$

$$\check{C}_{cs} = 2\left(\frac{\varphi}{\lambda(1-p_1\sqrt{\alpha}) + (1-\lambda)(1-p_1)(\frac{3}{4} + \frac{1}{4}\sqrt{\alpha})}\right)^2.$$

The cutoff values are $\underline{\lambda} = \frac{1-p_1}{1+3p_1}$, $\lambda^* \equiv \frac{1-p_1}{1+p_1}$, and $\overline{\lambda} = \frac{(1-p_1)(1+\sqrt{\alpha})}{\big[(1-p_1)+\sqrt{\alpha}(1+3p_1)\big]}$. By fixing $\alpha = 0.5$, we obtain Figure 1 by plugging in $p_1 = 0.3$ and $p_1 = 0.7$ into the above formula. By varying $\alpha \in (0,1)$ and plugging in $p_1 = 0.3$ and $p_1 = 0.7$ into $\underline{\lambda}$ and $\overline{\lambda}$, respectively, we obtain Figure 2.