ANNALS OF ECONOMICS AND FINANCE **22-2**, 317–365 (2021)

Collusion, Shading, and Optimal Organization Design in a Three-tier Agency Model with a Continuum of Types*

Yutaka Suzuki†

We apply the First-Order Approach and Monotone Comparative Statics to the continuous-type, three-tier agency model with hidden information and collusion à la Tirole (1986,1992), characterize the nature of equilibrium contract implemented under the possibility of collusion between supervisor and agent, and obtain a general comparison result on the two-tier vs. three-tier organization structures. We then introduce a behavioral idea, "shading" (Hart and Moore (2008)). By combining the two ideas, collusion and shading, we obtain a new result on the choice of Collusion-proof vs. Equilibrium Collusion regimes, and give a micro foundation to ex-post haggling costs, addressed by Transaction Cost Economics.

Key Words: Collusion; Supervision; Mechanism Design; First Order (Mirrlees) Approach; Behavioral Economics; Shading.

JEL Classification Numbers: D82, D86.

1. INTRODUCTION

In this paper, we apply the "First-Order Approach" and the "Monotone Comparative Statics" to the continuous-type, three-tier agency model with hidden information and collusion à la Tirole (1986,1992), thereby providing a framework that can address the issues treated in the existing literature

*This research was supported by Grant-in-Aid for Scientific Research by Japan Society for the Promotion of Science (C) 20530162 and 23530383.

† Faculty of Economics, Hosei University, 4342 Aihara, Machida-City, Tokyo 194-0298 Japan. E-mail: yutaka@hosei.ac.jp. I would like to thank Oliver Hart, Hideshi Itoh, Michihiro Kandori, Hitoshi Matsushima, Kathryn Spier, Kenichi Amaya, Charles Angelucci, audiences at Hosei University, Association for Public Economic Theory (PET 2008), UECE Lisbon Meetings: Game Theory and Applications (2009), Japanese Economic Association (2010), Game Theory Conference at Stony Brook (2010), Contract Theory Workshop East (CTWE), Harvard/MIT Contracts and Organization Lunch (2012), and Micro Workshop at University of Tokyo (2012) for their valuable comments. I also would like to thank an anonymous referee of this Journal for the detailed comments and suggestions, and appreciate the Co-editor's helpful advice for the revision of the paper.

1529-7373/2021

in a much simpler fashion. Then, we characterize the nature of equilibrium contract that can be implemented under the possibility of collusion between the supervisor and the agent, and obtain a general comparison result on the two-tier vs. three-tier organization structures. Next, we introduce a behavioral contract theory idea, "shading" (Hart and Moore (2008)) into the model. By combining the two ideas, i.e., collusion and shading, we can not only enrich the existing collusion model, including a new result on the choice of Collusion-proof vs. Equilibrium Collusion regimes, but also give a micro foundation to ex-post haggling costs, addressed by Transaction Cost Economics (e.g. Coase (1937) and Williamson (1975)). This will contribute to a deep understanding of resource allocation and decision process in hierarchical organization.[1]

The research, which deals with a three-tier agency model with hidden information and collusion, has so far been developed by Tirole (1986, 1992), Laffont and Tirole (1991, 1993), and Laffont and Martimort (1997) etc. In addition, Kofman and Lawarree (1993) applied a three-tier agency model — consisting of the two-type (productivity) agent, the internal and external auditors (supervisors), and the principal — to the issue of auditing and collusion.[2] However, such literature has a rather complicated model whose structure involves a Kuhn-Tucker problem with many IC (Incentive Compatibility) and IR (Individual Rationality) constraints, and is not a simple mathematical model. This mathematical complexity is a disadvantage.

In contrast, we construct a three-tier agency model with a continuum of types in this paper, where we exploit the "Monotone Comparative Statics" à la Topkis (1978) and Edlin and Shannon (1998), and the "First-Order Approach" à la Mirrlees (1971), which is a widely-used way to reduce the number of incentive constraints by replacing them with the corresponding First-Order Conditions. We thereby provide a framework that can address the issues treated in the literature in a much simpler fashion.

The basic tradeoff in our model is the benefit from the reduction in information rent by adding the auditor (supervisor) versus the resource cost of adding him into the hierarchy, and this bottom line is basically preserved through the model. The optimal collusion-proof contract in the Principal-Supervisor-Agent three-tier regime has the property whereby (1) Efficiency at the top (the highest type) and (2) Downward distortion for all other types, and the downward distortion is aggravated at the optimum, in

[1]The examples of hierarchical organizations would include, though not be restricted to, the corporate hierarchy, consisting of Manager, Foreman, and Worker, the governmental procurement, consisting of Department of Defense, Contracting Company, and Subcontractor, and the regulatory institutions, consisting of Congress, Government Agency, and Regulated Firm, à la Laffont and Tirole (1991).

[2]Bolton and Dewatripont (2005)'s textbook presents a simple version of the collusion models (Tirole (1986), Kofman and Lawarree (1993)).

comparison with the Principal-Agent two-tier regime. The optimal solution allows simple comparative statics, which shows that downward distortions from the first best output levels increase when the accuracy of supervision and the efficiency of collusion increase. This would be a specific contribution to the literature.

We compare the payoffs between two regimes, that is, the 'Three-tier' Collusion-proof regime (TH) and the 'No-Supervisor' (standard 'Two-tier') regime (TW). We find that under the assumption that the cost of introducing the supervisor (a transaction cost) is zero, the principal prefers the 'Three-tier' Collusion-proof regime (TH) in terms of his expected payoff. Intuitively, since the principal does not commit himself not to adjust the output (quantity) rule as well as the price rule in the 'Three-tier' Collusion-proof regime (TH), he optimally adjusts both of them and tries to design a "more state-contingent" contract through more efficient use of supervisor's report, which is more efficient than the pooling output (quantity) rule, where the principal insists on implementing an output function which varies only with the type of the agent.[3]

Then, we incorporate behavioral elements à la Fahr and Schmidt (1999) into the model, and examine their effects on the optimal solution in the principal-supervisor-agent hidden information model with collusion. We find that these behavioral elements can change the monetary reward for inducing the true information, and so the virtual surplus for each type is also altered through the change in the information rent (an incentive cost for inducing a truthful information revelation). Thus, the optimal solution with behavioral elements can be different from the one with no behavioral elements. More concretely, we introduce the recent behavioral contract theory idea, "shading" (Hart and Moore (2008)) into our collusion model. Hart and Moore (2008) introduced a behavioral idea that a contract provides a reference point for parties' feelings of entitlement. A party who felt aggrieved in terms of his entitlement shades (punishes) the party who aggrieved him to the point where his payoff falls by a constant multiplied by the aggrievement level, that is, the former shades (punishes) the latter by a constant times the aggrievement level. In their model, contracting parties possess behavioral preferences: they prefer to impose losses on their contracting partner if they perceive that their partner has chosen an action within the range permitted formally that falls short of "consummate" performance. In sum, each party interprets the contract in a way that is most favorable to him, which generates a conflict of entitlements. When he does not obtain the most favored outcome within the contract, he engages

[3]Suzuki (2018) applied this analysis to the three-tier hierarchical structure of Global Pollution Control, which consists of the Supra-National Regulator, the Government, and the Polluting Firm.

in shading. This will lead to ex post controversy and mutual punishment, which may bring about a great deal of ex post inefficiency.

We introduce this behavioral idea, "shading" as a component of ex-post haggling (addressed by Coase (1937) and Williamson (1975)) into our model, thereby constructing a new model of hierarchical organization. By combining the two ideas, i.e., collusion and shading, we can enrich the existing model and obtain a new result on Collusion-proof vs. Equilibrium Collusion in that the increase in shading pressure (behavioral element) strengthens the incentive for collusion, thereby makes it difficult to implement the collusion-proof (Supervisor's truth telling) incentive schemes, which leads to the Equilibrium Collusion. That is, the collusion-proof principle does not hold any more in the presence of strong shading pressures (behavioral elements) and weak accuracy of supervision.[4]

Further, by considering shading as a component of ex-post haggling (addressed by Coase (1937) and Williamson (1975), more generally, Transaction Cost Economics (TCE)), we can give a micro foundation (an explicit modeling) to ex-post adaptation costs, where we view rent-seeking associated with collusive behavior and ex-post haggling generated from aggrievement and shading as the two sources of the costs.[5] By using this model, we can analyze the optimal organizational design problem as an optimal response to the trade-off between gross total surplus and ex-post haggling cost. We believe that our model can help a deep understanding of resource allocation and decision process in hierarchical organizations.

Our paper is constructed as follows. In section 2, we present our continuous-type, three-tier agency model consisting of three risk neutral parties, and explain the timing and the full information benchmark (the First Best solution). Then, we consider the optimal collusion-free contract, assuming that side contracts are infeasible (coalitions do not form), and at the same time explain how to apply the Mirrlees First-Order approach and the Monotone Comparative Statics. In section 3, we substantially analyze a three-tier hierarchy, where the principal communicates not only with the agent, but also with the supervisor. After introducing the possibility of collusion between the supervisor and the agent, but still with no behav-

[4]Suzuki (2019) shows how the "shading" mechanism (Hart and Moore, 2008) can mitigate the ratchet effect and renegotiation problem in the dynamic adverse selection setting.

[5]Theoretically, our model deals with a situation where bilateral collusive contracts are feasible while the grand contract is not feasible (i.e., an incomplete grand contract situation), which corresponds to a case where the Coase Theorem will not hold since externalities cannot be fully internalized (like in the Coase's 1937 paper). It would be novel and interesting to model the situation where the third party who suffers from the negative externality brought by such bilateral, collusive contracts shades ex post the colluding party (especially, the supervisor) by a constant times the aggrievement (the negative externality he suffers from), in the three-tier agency framework.

ioral element, we characterize the optimal collusion-proof contracts in our unified framework. Then, we examine the payoff comparison between the two (two-tier vs. three-tier) regimes, and provide some comparative statics in the accuracy of monitoring, the possibility of collusion. In section 4, we introduce the behavioral contract theory idea, "shading" into our collusion model. By combining the two ideas, i.e., collusion and shading, we enrich our collusion model and show how the analytical results are changed by the introduction of behavioral elements, including a new result on the choice of Collusion-proof vs. Equilibrium Collusion regimes. Section 5 concludes the paper.

2. MODEL

2.1. The Parties

The framework of our analysis is a simple three-tier hierarchy. The top of the hierarchy is the residual claimant of profits generated by the whole structure: the principal (P). The bottom layer is the agent (A), the only level that actually produces any output. The intermediate layer is a supervisor (S), who is capable of collecting information on the agent's unobservable characteristics.

The agent is the productive unit of the structure; he controls a technology that generates the productive outputs. The agent is endowed with a productive parameter θ, which has a continuous type space $\Theta = [\underline{\theta}, \overline{\theta}]$, with the cumulative distribution function $F(\cdot)$ and a strictly positive density $f(\theta) = F'(\theta)$, and is private information for the agent. $C(X, \theta)$ is the effort cost for the agent of type θ to attain the output X, and for each θ satisfies $C(X, \theta) > 0, \partial C(X, \theta)/\partial X > 0, \partial^2 C(X, \theta)/\partial X^2 > 0, \forall X \in \mathbb{R}_+$. In addition, we assume that the marginal cost of the output, $\partial C(X, \theta)/\partial X$, is strictly decreasing in type θ, i.e., higher types always have gentler cost functions. $W(X)$ is the wage scheme which the agent of type θ is faced with, and then his utility function is described as $U(X, \theta) = W(X) - C(X, \theta)$. We assume that $U(X, \theta)$ has the Single Crossing Property (SCP), in the sense that the derivative $U_X(X, \theta)$ exists and is strictly increasing in $\theta \in \Theta$ for all X.[6] Under the assumption that the wage scheme $W(X)$ is differentiable, the SCP of $U(X, \theta)$ is satisfied, because the marginal cost of output $C_X(X, \theta)$ is strictly decreasing in type θ. We normalize the agent's reservation utility as 0 for all types.

The supervisor has a monitoring role in the structure. The principal has access, at a cost Z, to the supervisor who is an internal auditor and

[6] Edlin and Shannon (1998) introduced this SCP under the name of "increasing marginal returns". This condition, which can be also referred to as "Strict Increasing Difference" (Amir (2005)), is a key property to ensure our monotone comparative statics.

can, for each θ, provide proof of the fact (θ) with probability p, and with $1-p$, is unable to obtain any information. We assume that proofs of θ cannot be falsified, and thus the agent is protected against false claims that his type θ is higher/lower than it really is, and that this is hard information- in the way Tirole (1986) defines this term. In other words, the supervisor has to document every report he makes to the principal on the agent's productivity, and he has no way to produce enough supporting documentation for a false report. Therefore, the principal can verify the truth of the supervisor's report. Payoff of the supervisor is described by the wage payment W_S and S's reservation utility is 0.

The principal is risk neutral: he observes both the productive output X and the report of the supervisor r which are both verifiable to third parties.

2.2. Timing

We now describe the information structure and the extensive form of our model. The information structure is such that before contracting the agent knows his unobservable productivity θ while the other parties share a common prior $f(\theta)$. Negotiation takes place among the principal, the supervisor, and the agent. The principal is assumed to have all the bargaining power: he proposes a take-it-or-leave-it offer C (contract) to both the agent and the supervisor, which specifies a schedule of compensations for both supervisor and agent as a function of the output X and the supervisor's report r. That is, the contract C consists of $W(X,r)$ for the agent and $W_S(X,r)$ for the supervisor. The agent and the supervisor observe each other's contracts and take the decision to accept or reject C, simultaneously and independently.

If the contract is accepted, then the supervisor learns the signal on the productivity of the agent, and the collusion between the agent and the supervisor may take place. We assume, for simplicity, that in the collusion game the agent has all the bargaining power and makes a take-it-or-leave-it collusive offer to the supervisor. The supervisor can only accept or reject the offer.

The supervisor then produces a report for the principal. This report is public information. The agent chooses effort, output is realized, and the three parties exchange transfers according to the latest contractual agreements (main and side contracts).

2.3. The Full Information Benchmark (First Best)

As a benchmark, we consider the case in which the principal observes the agent's type θ. Given θ, he offers the bundle (X, W_S, W) to solve:

$$\begin{aligned} \max_{(X,W_S,W)} \quad & X - W(X) - W_S(X) \\ \text{s.t.} \quad & W(X) - C(X,\theta) \geq 0 \ \ \text{(IR of the agent)} \\ & W_S(X) \geq 0 \ \ \text{(IR of the supervisor).} \end{aligned}$$

The supervisor's and the agent's Individual Rationality constraints bind at an optimal solution. Then, the principal eventually solves $\max_X X - C(X,\theta)$, which is exactly the total surplus maximization. Let $X^{FB}(\theta)$ denote a solution to this maximization problem, called the First Best (FB) solution.

Now, we assume that the First Best output levels $X^{FB}(\theta)$ exist and unique for each type θ. Uniqueness of $X^{FB}(\theta)$ is ensured by assuming that total surplus $TS = X - C(X,\theta)$ is strictly concave in X, which is satisfied because $\frac{\partial^2 TS}{\partial X^2} = -\frac{\partial^2 C(X,\theta)}{\partial X^2} < 0$. Then, according to Edlin and Shannon (1998), we check whether our assumptions ensure that the First Best output $X^{FB}(\theta)$ is strictly increasing in type θ. If $-C(X,\theta)$ satisfies SCP, then total surplus $X - C(X,\theta)$ satisfies SCP, and if $X^{FB}(\theta)$ is in the interior for each θ, we see that $X^{FB}(\theta)$ is strictly increasing in θ.

2.4. Optimal Contract when Side Contracts are infeasible (Collusion-free Solution: CF)

We first consider the optimal contract, assuming that side contracts are infeasible (coalitions do not form). As the proposition 1 of Tirole (1986) says, in the absence of coalitions, the optimal contract is equivalent to the optimal contract between the principal and the agent when the principal has the supervisor's information structure. The supervisor's wage is constant in all states of nature, and he obtains his reservation utility 0. Since the supervisor has no incentive to lie (conceal the evidence), the principal can obtain the supervisor's information at "minimal cost." In this case, the three-tier structure substantially boils down to the two-tier principal/agent one, and the supervisor plays a completely passive role, just like a machine. The principal has full information when the supervisor observes the true state $s = \theta$ and can implement the first best contracts, which occurs with probability p, and when the agent observes θ but the supervisor observes nothing $s = \phi$ with probability $1 - p$, the incentive problem is restricted to the agent's truth-telling problem, the issue wherein is to induce the agent to reveal the true information θ.

2.4.1. The Revelation Principle

Now we consider a different contract from the contract $W : X \to \mathbb{R}$ which we have considered so far, where the agent is asked to announce his type $\hat{\theta}$, and receives payment $W(\hat{\theta})$ in exchange for an output $X(\hat{\theta})$ on the basis of his announcement $\hat{\theta}$. This contract $(W(\hat{\theta}), X(\hat{\theta}))$ is called a Direct Revelation Contract. According to the Revelation Principle, any contract $W : X \to \mathbb{R}$ can be replaced with a Direct Revelation Contract that has an equilibrium in which all types receive the same bundles as in the original contract.[7]

2.4.2. Incentive Constraints for the Agents with a Continuum of Types under Asymmetric Information

The principal's problem for designing the optimal collusion-free contract can be written as:

$$\begin{aligned}
\max_{\langle X(\cdot),W(\cdot)\rangle} \quad & p\int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)]f(\theta)d\theta \\
& +(1-p)\int_{\underline{\theta}}^{\overline{\theta}} [X(\theta) - W_S(\theta,\phi) - W(\theta)]f(\theta)d\theta \\
\text{s.t.} \quad & W(\theta) - C(X(\theta),\theta) \geq W(\hat{\theta}) - C(X(\hat{\theta}),\theta) (IC_{\theta\hat{\theta}}) \quad \forall\theta,\hat{\theta} \in \Theta \\
& W(\theta) - C(X(\theta),\theta) \geq 0 \quad (IR_\theta \text{ of the agent}) \quad \forall\theta \in \Theta \\
& W_S(\theta,s) = 0 \qquad (IR \text{ of the supervisor is binding}) \quad \forall\theta \in \Theta, s \in \{\theta,\phi\}
\end{aligned}$$

The first term of the principal's objective function comes from the fact that the principal has full information when the supervisor observes the true state $s = \theta$, with probability p, and then can implement the first best solutions $X^{FB}(\theta)$ for each θ. Just as in the two-type case, only the lowest type agent's IR binds out of all the participation constraints.

LEMMA 1. *At a solution $(X(\cdot), W(\cdot))$, all IR_θ with $\theta > \underline{\theta}$ are not binding, and only $IR_{\underline{\theta}}$ is binding.*

As for the analysis of ICs with a continuum of types, Mirrlees (1971) introduced a widely used way to reduce the number of incentive constraints by replacing them with the corresponding First-Order Conditions.[8] The "trick" is as follows.

[7] As for the Revelation Principle, see, Fudenberg and Tirole (1991) and Bolton and Dewatripont (2005).

[8] Fudenberg and Tirole (1991) pp257-268 reviews the First Order Approach.

(IC) can be written as $\theta \in \arg\max_{\hat{\theta}\in\Theta} U(\hat{\theta},\theta)$, where $U(\hat{\theta},\theta) = W(\hat{\theta}) - C(X(\hat{\theta}),\theta)$ is the utility that the agent of type θ receives by announcing that his type is $\hat{\theta}$. If $\theta \in (\underline{\theta},\overline{\theta})$ and $U(\hat{\theta},\theta)$ is differentiable in $\hat{\theta}$, then the first order condition $\frac{\partial U(\hat{\theta},\theta)}{\partial\hat{\theta}}|_{\hat{\theta}=\theta} = 0$ is necessary for the above optimality. We define the Agent's equilibrium utility (the value):

$$U(\theta) \equiv U(\theta,\theta) = W(\theta) - C(X(\theta),\theta). \tag{1}$$

Note that this utility depends on θ in two ways — through the agent's true type and through his announcement. Differentiating with respect to θ, we have $U'(\theta) = U_{\hat{\theta}}(\theta,\theta) + U_{\theta}(\theta,\theta)$, where the first derivative of U is with respect to the agent's announcement (the first argument) and the second derivative is with respect to the agent's true type (the second argument). Since the first derivative equals zero by $\frac{\partial U(\hat{\theta},\theta)}{\partial\hat{\theta}}|_{\hat{\theta}=\theta} = 0$, we have the Envelope condition

$$U'(\theta) = U_{\theta}(\theta,\theta) = -\frac{\partial C(X(\theta),\theta)}{\partial\theta}. \tag{2}$$

By integrating it, we have the important formula:

$$U(\theta) = U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau),\tau)}{\partial\tau} d\tau \tag{ICFOC}$$

(ICFOC) demonstrates that with a continuum of types, incentive compatibility constraints pin down up to a constant plus all types' utilities for a given output rule $X(\cdot)$.[9] This is a remarkable result that holds only for the continuous-type case.

(ICFOC) is consistent with the truthful announcement $\hat{\theta} = \theta$ being a local maximum, but may not be a global maximum. It is even consistent with truthful announcement being a local minimum. To rule out these situations Topkis (1978) and Edlin and Shannon (1998) establish that the agent's output choices in any incentive compatible contract are nondecreasing in type. Thus, any piecewise differentiable IC contract must satisfy that $X(\cdot)$ is nondecreasing (M). Under SCP, ICFOC in conjunction with (M) do ensure that truth telling is a global maximum, i.e., all ICs are satisfied:

LEMMA 2. *$(X(\cdot), W(\cdot))$ is Incentive Compatible if and only if both (ICFOC) and (M) hold, where $U(\cdot)$ is given by (1).*

[9]Our methodology is related to the "Envelope Approach" in auction theory, e.g., the analysis of first price auction by the envelope approach. As for it, e.g., see Milgrom (2004).

Proof. See, Appendix A.1. ▮

Given (ICFOC), we can express transfers: $\underbrace{W(\theta)}_{\text{Wage Payment}} = \underbrace{C(X(\theta),\theta)}_{\text{Effort Cost}} + \underbrace{U(\theta)}_{\substack{\text{Information Rent} \\ \text{given for type } \theta}}$

The collusion-free optimal solution will be derived as a special case of the unified analysis in section 3.

3. OPTIMAL CONTRACT WHEN SIDE CONTRACTS ARE FEASIBLE (COALITIONS CAN FORM)

3.1. The Collusion-proof Problem

In this section, we consider the three-tier hierarchy, where the principal can have access to the supervisor at a cost Z.[10] In that case, the supervisor can, for each θ, provide a proof of this fact with probability p, and with $1-p$, is unable to obtain any information. We assume that proofs of θ cannot be falsified. That is, θ is hard information.[11] On the other hand, the agent can potentially benefit from a failure by the supervisor to truthfully report that his type is θ when the supervisor observed the signal θ. The supervisor will collude with the agent if he benefits from such behavior. We assume the following collusion technology: if the agent offers the supervisor a transfer (side payment) t, he benefits up to kt, where $k \in [0,1]$. That is, only a fraction, $k \in [0,1]$, of the agent's side payment ends up in the supervisor's hands. The idea is that transfers of this sort may be subject to transaction costs.[12] We assume that side-contracts of this sort are enforceable (See, Tirole 1992).[13]

Now, suppose that the type of the agent is θ. When the supervisor cannot obtain any information for θ (which occurs with probability $1-p$), the only thing the supervisor can do is reporting $r = \phi$. Then, the principal implements the screening contract (Direct Revelation Contract) $\{X(\hat{\theta}), W(\hat{\theta})\}$ under asymmetric information in the Principal-Agent, two-

[10] Z is the cost for the principal to communicate with the supervisor, which includes a cost for verification of the supervisor's report with proof (evidence).

[11] We assume that the agent correctly knows whether the supervisor is informed of his type information θ or not. This is the same assumption as the early literature, e.g., Tirole (1986).

[12] The case $k = 0$ corresponds to a full dead weight loss in the side transaction and yields the collusion-free (no-collusion) case.

[13] Of course, enforceability of side contracts should have some more (theoretical or behavioral) foundation. In section 4, we introduce a new idea where the behavioral element ("Shading") becomes a strong driver that implements Equilibrium Collusion (side contract) between the supervisor and the agent.

tiered hierarchy. Then, the type θ agent can obtain the information rent $U(\theta)$.

When the supervisor has obtained the hard evidence for θ (which occurs with probability p), the supervisor can choose a report $r \in \{\phi, \theta\}$, where ϕ means that he did not obtain any information. If the principal receives the report (with hard evidence) from the supervisor that the type information is θ, the principal can eliminate the downward distortion and implement the first best contract $\{X^{FB}(\theta), W^{FB}(\theta)\}$ and then exploit the information rent $U(\theta)$ from the type θ agent. (This arrangement is committed in the initial contracts.) If the type θ agent anticipates this outcome, since the agent can benefit from a failure by the supervisor to report his type θ truthfully, he will offer the supervisor the transfer (side payment) $t = U(\theta)$, the amount equivalent to his information rent, of which the supervisor benefits up to kt, where $k \in [0, 1]$. Hence, the principal must pay $W_S(\theta) = kU(\theta)$ to the supervisor in opposition to the collusive offer by the agent, in order to elicit true information.

In other words, in order to deter collusion between the supervisor and the agent, the principal will have to offer the supervisor a reward $W_S(\theta)$ for reporting $r = \theta$, such that the coalition incentive compatibility constraint $W_S(\theta) \geq kU(\theta)$ is satisfied, from which the optimal transfer $W_S(\theta) = kU(\theta)$ is derived. (This reward scheme is also committed in the initial contract to the supervisor.)

To summarize, when the supervisor obtains the proof of θ with probability p, the principal can implement the first best payoff $X^{FB}(\theta) - C(X^{FB}(\theta), \theta)$, but must pay the incentive reward $kU(\theta)$ for the supervisor to tell the truth $r = \theta$. This is the essential difference from the collusion-free regime with no supervisory reward.

On the other hand, when the supervisor cannot obtain any information for θ with probability $1-p$, the principal implements the screening contract (Direct Revelation Contract) $\{X(\hat{\theta}), W(\hat{\theta})\}$ under asymmetric information, and in equilibrium gives the information rent $U(\theta)$ to the type θ agent, in exchange for attaining total surplus $X(\theta) - C(X(\theta), \theta)$.

Substituting $X(\theta) = X^{FB}(\theta)$, $W_S(\theta) = kU(\theta)$ and $W(\theta) = C(X^{FB}(\theta), \theta)$ with probability p, and $X(\theta) = X(\theta)$, $W_S(\theta) = 0$ and $W(\theta) = C(X(\theta), \theta) + U(\theta)$ with probability $1 - p$, into the Principal's objective function $X(\theta) - W_S(\theta) - W(\theta)$, the expected total surplus minus the expected information

rent for type θ is written as

$$\begin{aligned}
&\underbrace{p}_{\substack{\theta \text{is} \\ \text{observed}}} \times \left[\underbrace{X^{FB}(\theta) - C(X^{FB}(\theta), \theta)}_{\text{(Ex post) First Best Allocative Efficiency}} - kU(\theta) \right] \\
+ \;& (1-p)[X(\theta) - C(X(\theta), \theta) - U(\theta)] \\
= \;& (1-p)[X(\theta) - C(X(\theta), \theta)] + p[X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] \\
- \;& [(1-p) + pk]U(\theta).
\end{aligned}$$

Hence, the principal's optimization problem is as follows.

$$\begin{aligned}
\max_{X(\cdot), U(\cdot)} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \left[(1-p)[X(\theta) - C(X(\theta), \theta)] + p[X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] - [(1-p) + pk]U(\theta) \right] f(\theta) d\theta - Z \\
\text{s.t.} \quad & dX(\theta)/d\theta \geq 0 : X(\theta) \text{is nondecreasing} \qquad \text{(M)} \\
& U(\theta) = U(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} \frac{\partial C(X(\tau), \tau)}{\partial \tau} d\tau \qquad \text{(ICFOC)} \\
& U(\underline{\theta}) = W(\underline{\theta}) - C(X(\underline{\theta}), \underline{\theta}) \geq 0 \qquad (IR_{\underline{\theta}}).
\end{aligned}$$

3.2. Solving the Relaxed Program

Thus, the principal's optimization problem can be rewritten as

$$\begin{aligned}
\max_{X(\cdot)} \quad & \int_{\underline{\theta}}^{\overline{\theta}} [(1-p)[X(\theta) - C(X(\theta), \theta)] - [(1-p) + pk]U(\theta)] f(\theta) d\theta \\
& + p \int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] f(\theta) d\theta - Z \\
\text{s.t.} \quad & dX(\theta)/d\theta \geq 0 \quad \text{(M)} \; \forall \theta
\end{aligned}$$

where $\int_{\underline{\theta}}^{\overline{\theta}} U(\theta) f(\theta) d\theta$ can be called the expected information rent.

LEMMA 3. *Expected Information Rent is transformed as follows.*

$$\int_{\underline{\theta}}^{\overline{\theta}} U(\theta) f(\theta) d\theta = U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta), \theta)}{\partial \theta} \frac{1 - F(\theta)}{f(\theta)} f(\theta) d\theta. \qquad (3)$$

Proof. See, Appendix A.2. ∎

Substituting these expected information rents into the principal's program, and ignoring the constant terms, the program for designing the collusion-proof contract becomes

$$\begin{aligned} \max_{X(\cdot)} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \left[(1-p)[X(\theta) - C(X(\theta),\theta)] + [(1-p)+pk] \frac{\partial C(X(\theta),\theta)}{\partial \theta} \frac{1-F(\theta)}{f(\theta)} \right] f(\theta) d\theta \\ \text{s.t.} \quad & dX(\theta)/d\theta \geq 0 \quad \text{(M)} \quad \forall \theta. \end{aligned} \tag{4}$$

We ignore the Monotonicity Constraint (M) and solve the resulting relaxed program.[14] Thus, the principal maximizes the expected value of the expression within the square brackets, the virtual surplus, and denoted by $J(X,\theta)$. This expected value is maximized by simultaneously maximizing virtual surplus for (almost) every type θ, i.e.,

$$\begin{aligned} X^S(\theta) \quad \in \quad & \arg\max_{X(\cdot)} (1-p)[X(\theta) - C(X(\theta),\theta)] + [(1-p)+pk] \left[\frac{1-F(\theta)}{f(\theta)} \right] \frac{\partial C(X(\theta),\theta)}{\partial \theta} \\ \iff \quad & (1-p) \left\{ X(\theta) - C(X(\theta),\theta) + \left[\frac{1-F(\theta)}{f(\theta)} \right] \frac{\partial C(X(\theta),\theta)}{\partial \theta} \right\} \\ + \quad & pk \left[\frac{1-F(\theta)}{f(\theta)} \right] \frac{\partial C(X(\theta),\theta)}{\partial \theta}. \end{aligned} \tag{5}$$

This defines the optimal output rule $X^S(\cdot)$ for the relaxed program. The principal's choice of $X^S(\theta)$ can be understood as a trade-off between maximizing the expected total surplus for type θ when the supervisor cannot obtain any information for θ with probability $1-p$ and reducing the sum of information rents both when the supervisor cannot obtain any information for θ with probability $1-p$ and when the supervisor can obtain the proof of true information, with probability p.

In particular, for the highest type $\overline{\theta}$, there are no higher types, i.e., $F(\overline{\theta}) = 1$ and the principal just maximizes total surplus, choosing $X^S(\overline{\theta}) = X^{FB}(\overline{\theta})$. In words, we have efficiency at the top. For all other types, the principal will distort output to reduce information rents. To see the direction of distortion, consider the parameterized maximization program

$$\max_{X \in X} \Psi(X,\xi) = \underbrace{X(\theta) - C(X(\theta),\theta)}_{\text{Total Surplus}} + \xi \left[\frac{1-F(\theta)}{f(\theta)} \right] \frac{\partial C(X(\theta),\theta)}{\partial \theta}.$$

Here $\xi = 0$ corresponds to total surplus-maximization (first-best), $\xi = 1$ ($p = 0$ or $k = 0$) corresponds to the principal's second best problem,[15]

[14] Since the Monotonicity Constraint (M) is the necessary condition for implementability, we present a sufficiency condition for the condition (M) to be satisfied, in the proposition2.

[15] $p = 0$ corresponds to the standard two-tier asymmetric information regime (TW), and $k = 0$ ($p > 0$) corresponds to the collusion-free (no-collusion) regime (CF) in section 2.4. When $p = 1$, the principal can implement the first-best solution $X^{FB}(\theta), \forall \theta$, where the optimal solution is $X^S(\theta) = 0, \forall \theta$, thereby the information rent is also $U(\theta) = 0, \forall \theta$.

and $\xi = 1 + \frac{pk}{1-p}$ corresponds to the collusion-proof three-tier problem (the above relaxed problem).[16]

Note that $\frac{\partial \Psi(X,\xi)}{\partial X \partial \xi} = \left[\frac{1-F(\theta)}{f(\theta)}\right] \frac{\partial^2 C(X(\theta),\theta)}{\partial X \partial \theta} < 0$ for $\theta < \overline{\theta}$ since the agent's payoff has the single crossing property (SCP), that is $\partial^2 U(X,\theta)/\partial X \partial \theta = -\partial^2 C(X,\theta)/\partial X \partial \theta > 0$. Thus, $\Psi(X,\xi)$ has SCP in $(X,-\xi)$. Based on Edlin and Shannon (1998), we have $X^*(\xi \geq 1) \leq X^*(\xi = 1) \leq X^*(\xi = 0)$, that is, $X^S(\theta) \leq X^{SB}(\theta) \leq X^{FB}(\theta)$ for all $\theta < \overline{\theta}$.

Thus, in the collusion-proof three-tier regime, the principal induces less marginal incentives than the second best regime, in order to reduce the information rents paid to the supervisor and the agent θ (and the information rents of all types above θ), in other words, in order to reduce the implementation costs for any $X < X^S(\overline{\theta}) = X^{FB}(\overline{\theta})$. Thus, we obtain the following proposition.

PROPOSITION 1. *In the Principal-Supervisor-Agent three-tier regime with a continuum of types, the optimal collusion-proof contract has the property that*

(1) Efficiency at the top (the highest type $\overline{\theta}$) $X^S(\overline{\theta}) = X^{FB}(\overline{\theta})$. (2) Downward distortion for all other types $\theta \in [\underline{\theta}, \overline{\theta})$ is aggravated, that is,

$$X^S(\theta) \underbrace{\leq}_{\substack{\text{Equality holds} \\ \text{at } p=0 \text{ or } k=0}} X^{SB}(\theta) \underbrace{\leq}_{\substack{\text{Equality holds} \\ \text{at } \theta=\overline{\theta}}} X^{FB}(\theta).$$

Now, remember that we ignored the monotonicity constraint (M) and solved the relaxed program. So, we need to check that the solution $X^S(\theta)$ indeed satisfies the monotonicity constraint (M), that is, the output rule $X^S(\theta)$ is nondecreasing. We define $h(\theta) \equiv f(\theta)/[1 - F(\theta)] > 0$, which is called the hazard rate of type θ[17]. Then, the principal's program can be rewritten as

$$\max_{X \in X} J(X,\theta) = X - C(X,\theta) + \left[1 + \frac{pk}{1-p}\right] \frac{1}{h(\theta)} \frac{\partial C(X,\theta)}{\partial \theta}. \qquad (5')$$

Based on Topkis (1978) and Edlin and Shannon (1998), assuming that $C(X,\theta)$ is sufficiently smooth, a sufficient condition for $X^S(\theta)$ to be non-

[16] $\xi = 1 + \frac{pk}{1-p}$ is the relative weight of the last term (virtual cost), i.e. $\frac{(1-p)+pk}{(1-p)}$.

[17] Fudenberg and Tirole (1991) examines when it is legitimate to focus on the relaxed program, by using the differentiability approach (i.e., analyzing total differentiation of the first order condition to the relaxed program), not using the monotone comparative statics method. They derive the monotone hazard rate condition, that is, the condition2 of Proposition 2 as the assumption sufficient to satisfy the monotonicity constraint (M).

decreasing in θ is for the following derivative to be strictly increasing in:

$$\frac{\partial J(X,\theta)}{\partial X} = 1 - \frac{\partial C(X,\theta)}{\partial X} + \left[1 + \frac{pk}{1-p}\right] \frac{1}{h(\theta)} \frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}. \tag{6}$$

Since $-C(X,\theta)$ satisfies SCP, the second term is strictly increasing in θ, and the first term does not depend on θ. The only problematic term, therefore, is the third term. Our result is ensured when the third term is nondecreasing in θ. Since $1/h(\theta)$ is positive and $\partial^2 C(X,\theta)/\partial X \partial \theta$ is negative, this is ensured when $\partial^2 C(X,\theta)/\partial X \partial \theta$ is nondecreasing. That is, we have

PROPOSITION 2. *A sufficiency condition for the optimal collusion-proof solution $X^S(\theta)$ to satisfy the monotonicity constraint (M) is that the following conditions hold.*

1. *$\partial^2 C(X,\theta)/\partial X \partial \theta$ is nondecreasing in θ.*
2. *The hazard rate $h(\theta)$ is nondecreasing.*

Example: The first assumption is satisfied e.g., in the following cost function forms:

$$C(X,\theta) = (X-\theta)^\alpha \quad \text{and} \quad C(X,\theta) = (X/\theta)^\alpha, \ \alpha \geq 2.$$

The second condition is called the "Monotone Hazard Rate Condition" and satisfied by many familiar probability distributions.[18] Now, we can present the following proposition on the comparative statics.

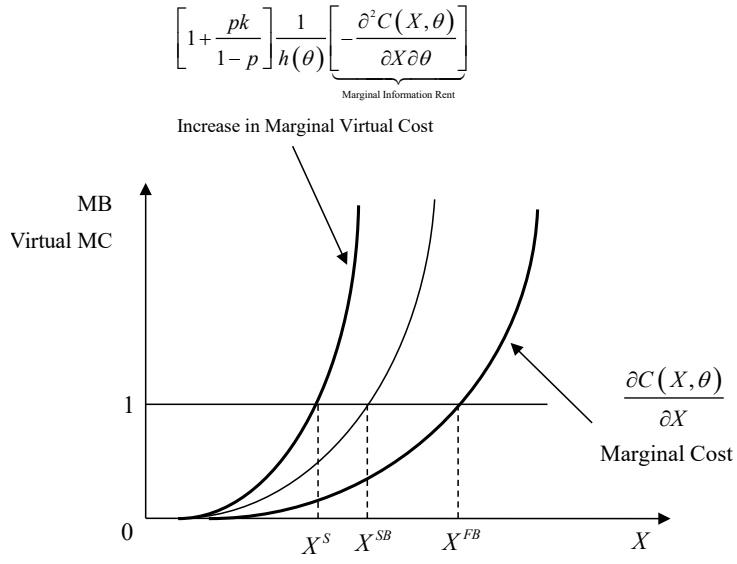
Graphical Explanation

Proposition 1 can be understood by using the Figure 1, which shows that the optimal solution $X^S(\theta)$ is determined such that the marginal benefit 1 equals the marginal virtual cost (marginal cost $\frac{\partial C(X,\theta)}{\partial X}$ plus marginal virtual information rent $-\left[1 + \frac{pk}{1-p}\right] \frac{1}{h(\theta)} \frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}$). The result of $X^S(\theta) \leq X^{SB}(\theta) \leq X^{FB}(\theta)$ basically comes from the increase in the virtual marginal cost due to $1 + pk/(1-p) \geq 1$, compared with the second-best case.

The condition 1 of Proposition 2 means that the marginal information rent $-\frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}$ is decreasing in θ, that is, shifts downwards as θ increases. Since the marginal cost $\frac{\partial C(X,\theta)}{\partial X}$ is also decreasing in θ, the proposition 2 as a whole refers to a sufficient condition for the virtual marginal cost to decrease in θ, that is, for $X^S(\theta)$ to increase in θ.

[18] For example, uniform, normal, logistic, and exponential distributions. See Fudenberg and Tirole (1991).

FIG. 1. Equilibrium Output in the Three-tier, Collusion-proof Regime X^S w.p $1-p$ and X^{FB} w.p p for each θ



PROPOSITION 3. *Suppose that the sufficiency condition in proposition 2 holds. Then, the optimal collusion-proof solution $X^S(\theta)$ is nonincreasing in the parameter p and the parameter k.*

Proof. From the equation (6), the derivative $J_X(X,\theta)$ is nonincreasing in the parameter p, because the derivative of $J_X(X,\theta)$ in the parameter p is $k/(1-p)^2 \geq 0$ for $k \in [0,1]$, multiplied by $\frac{1}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta} < 0$. Hence, due to the monotone comparative statics, the optimal solution $X^S(\theta)$ is nonincreasing in the parameter p. The latter part can also be proved in the same way: The derivative $J_X(X,\theta)$ is nonincreasing in the parameter k for $p \in [0,1]$, and so the optimal solution $X^S(\theta)$ is nonincreasing in the parameter k. That is, the distortion is nondecreasing in both p and k. ∎

3.3. Payoff Comparison between Two Regimes: Three-tier vs. Two-tier Structures

We compare the payoffs between two regimes, that is, 'Three-tier' regime (TH) and 'No-Supervisor' (two-tier) regime (TW).

The expected payoff for the principal in the 'Three-tier' regime (TH) is

$$(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X^S(\theta)-C(X^S(\theta),\theta)]f(\theta)d\theta+p\times\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta$$
$$+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^S(\theta),\theta)}{\partial\theta}f(\theta)d\theta-Z.$$

The expected payoff for the principal in the No-Supervisor (Two-tier) regime (TW), which is the standard second best regime and also corresponds to $p=0$ in the Three-tier regime, is

$$\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{TW}(\theta)-C(X^{TW}(\theta),\theta)+\frac{1}{h(\theta)}\frac{\partial C(X^{TW}(\theta),\theta)}{\partial\theta}\right]f(\theta)d\theta.$$

We first consider the comparison when $Z=0$ (The cost for the principal to communicate with the supervisor is zero). Then, we have the following proposition.

PROPOSITION 4. *Suppose $Z=0$. The principal prefers the 'Three-tier' regime with supervision (TH) to the 'Two-tier' regime with no supervision (TW) in terms of his expected payoff. That is,*

$$\begin{aligned}&(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X^S(\theta)-C(X^S(\theta),\theta)]f(\theta)d\theta+p\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta\\ +\ &[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^S(\theta),\theta)}{\partial\theta}f(\theta)d\theta\\ \geq\ &\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{TW}(\theta)-C(X^{TW}(\theta),\theta)+\frac{1}{h(\theta)}\frac{\partial C(X^{TW}(\theta),\theta)}{\partial\theta}\right]f(\theta)d\theta.\end{aligned}$$

Proof. See, Appendix A.3 ∎

Rationale

First, the principal compares the "Three-tier" regime (TH) with the "pooling" regime (PL) where the principal commits himself to the pooling output rule $X^P(\theta)$. In the "Three-tier" regime (TH), the principal designs a "more state-contingent" contract for more efficient use of supervisor's report $r\in\{\theta,\phi\}$, that is, he sets $X^{FB}(\theta)$ for the states $\{\theta,s=\theta\}$ where the agent type is θ and the supervisor's signal is $s=\theta$, and sets $X^S(\theta)$ for the states $\{\theta,s=\theta\}$ where the agent type is θ and the supervisor's signal

is $s = \phi$. On the other hand, in the Pooling regime, the principal does not use the supervisor's report $r \in \{\theta, \phi\}$ in a state-dependent way, but unanimously imposes the pooling output $X^P(\theta)$ for both states $\{\theta, s = \theta\}$ and $\{\theta, s = \phi\}$, which would not be efficient. If we use the terminology in Weitzman's paper (1974) "Prices vs. Quantities", the "Pooling" regime (PL) is the regime where the principal adjusts only the price rule $W(\theta)$ under the commitment to the pooling output (quantity) rule $X^P(\theta)$, in the form that he does not pay the information rent $U(\theta)$ to the agent of type θ when the supervisor's report is $r = \theta$.[19] On the other hand, in the "Three-tier" regime (TH), the principal optimally adjusts both of the output (quantity) rule $X(\theta)$ and the price rule $W(\theta)$, contingent on the supervisor's report $r \in \{\theta, \phi\}$. When the true type information θ is revealed from the supervisor with probability p, the principal implements the first-best outcome $\{X^{FB}(\theta), W^{FB}(\theta)\}$ based on its hard evidence. Otherwise, the downward distorted outcome $\{X^S(\theta), W^S(\theta)\}$ is implemented. These arrangements are optimally created and committed as the collusion-proof contract by the principal.

Next, when the principal compares the pooling regime $X^P(\theta)$ with the 'No Supervisor' two-tier regimes $X^{TW}(\theta)$, the virtual surplus for type θ is more increased in the former regime through the effective reduction of information rent due to $(1-p) + pk \leq 1$ that is,

$$\begin{aligned} & X(\theta) - C(X(\theta), \theta) + \underbrace{[(1-p) + pk]}_{\leq 1} \underbrace{\frac{\partial C(X(\theta), \theta)}{\partial \theta}}_{-} \frac{1}{h(\theta)} \\ \geq \; & X(\theta) - C(X(\theta), \theta) + \underbrace{\frac{\partial C(X(\theta), \theta)}{\partial \theta}}_{-} \frac{1}{h(\theta)}. \end{aligned}$$

Combining these two comparison results, we find that the principal always prefers the "Three-tier" regime (TH) to the "No Supervisor" two-tier regime (TW) when $Z = 0$.

The Choice of Organization Structure

[19] For the analysis of this "Pooling" regime, see Suzuki (2008). In contrast, in this paper, it is just a hypothetical regime used for payoff comparison between two regimes (Three-tier vs. Two-tier regimes).

Now define $Z^*(p,k)$ be the payoff difference between the 'Three-tier' regime (TH) and the 'Two-tier' regime (TW) when $Z=0$. That is,

$$\begin{aligned} Z^*(p,k) \;:=\; & \Bigg\{(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X^S(\theta)-C(X^S(\theta),\theta)]f(\theta)d\theta \\ & +p\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta \\ & +[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^S(\theta),\theta)}{\partial\theta}f(\theta)d\theta\Bigg\} \qquad (7) \\ & -\int_{\underline{\theta}}^{\overline{\theta}}\left[X^{TW}(\theta-C(X^{TW}(\theta),\theta))+\frac{\partial C(X^{TW}(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}\right]f(\theta)d\theta. \end{aligned}$$

$Z^*(p,k)$ is the relative importance of the three-tier structure with supervision, and could be rephrased as the "comparative (relative) advantage" à la Weitzman (1974).

Then, we have the following corollary for $Z>0$.

Corollary 1. *The optimal regulation structure R^* is determined based on the following rule:*

$$R^*(p,k,Z)=\begin{cases} TH: \textit{Three-tier structure} & \textit{if } Z\le Z^*(p,k) \\ TW: \textit{Two-tier structure} & \textit{if } Z>Z^*(p,k) \end{cases}$$

That is to say, the principal's optimal strategy is to choose the three-tier structure with supervision (TH) if $Z\le Z^*(p,k)$, and to choose the two-tier structure with no supervision (TW) if $Z>Z^*(p,k)$, for $0\le p,k\le 1$.

From the simple comparative statics, we have

$$\begin{aligned} \frac{\partial Z^*(p,k)}{\partial p} \;=\; & \int_{\underline{\theta}}^{\overline{\theta}}\underbrace{\{[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]-[X^S(\theta)-C(X^S(\theta),\theta)]\}}_{\ge 0}f(\theta)d\theta \\ & +\underbrace{(k-1)}_{\le 0}\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\underbrace{\frac{\partial C(X^S(\theta),\theta)}{\partial\theta}}_{<0}f(\theta)d\theta\ge 0 \\ \frac{\partial Z^*(p,k)}{\partial k} \;=\; & p\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^S(\theta),\theta)}{\partial\theta}f(\theta)d\theta\le 0 \quad \forall(p,k)\in[0,1]^2. \end{aligned}$$

As p (the accuracy of supervision/monitoring) increases, the relative importance $Z^*(p,k)$ of the three-tier structure increases. On the other hand,

as k (the efficiency of collusion, the easiness of collusion) increases, it decreases, since the increase in k increases the expected information rent the principal needs to pay to the supervisor[20].

4. OPTIMAL ORGANIZATION DESIGN UNDER COLLUSION AND SHADING

4.1. Introduction of Behavioral Element: Shading

We incorporate a behavioral element into the model, based on the "shading" model[21] by Hart and Moore (2008), which introduced a new idea that a contract provides a reference point for parties' feelings of entitlement. A party who felt aggrieved in terms of his entitlement shades (punishes) the party who aggrieved him to the point where his payoff falls by a constant multiplied by the aggrievement level, that is, the former shades (punishes) the latter by a constant times the aggrievement level.[22] Contracting parties possess behavioral preferences: they prefer to impose losses on their contracting partner if they perceive that their partner has chosen an action within the range permitted formally that falls short of "consummate" performance. In summary, each party interprets the contract in a way that is most favorable to him, which generates a conflict of entitlements. When he does not obtain the most favored outcome within contract, he engages in shading. This will lead to ex post controversy and mutual punishment. In our three-tier hierarchical structure, at the final stage, the agent and the principal may well shade (punish) the supervisor, who made a crucial report for payoff distribution, depending on their entitlements and aggrievements.

By introducing such a shading behavior as ex-post haggling into our collusion model and integrating two ideas, "collusion" and "shading", we try to give a micro foundation to ex-post adaptation costs, and understand the ex-post optimal adaptation as an optimal balance resulting from the trade-

[20]Conversely, as k decreases, the "comparative (relative) advantage" of the three-tier structure increases. The size of k will be related to the integrity/honesty of the supervisor. The lower k implies the possibility of less collusion between the supervisor and the agent, due to the higher integrity/honesty of the supervisor. It decreases the expected information rent the principal must pay to the supervisor.

[21]It is related to negative reciprocity in the behavioral economics literature, that is, "I am better off when someone who has tried to hurt me is hurt". Also see e.g. Falk and Fischbacher (2006) for the behavioral game theory literature on the formalization of reciprocity.

[22]Indeed, Fehr et al (2011) examine the realism of the shading concept in their experiment paper, and obtain a supportive result, which can be consistent with contractual opportunism and its punishment.

off between gross total surplus and ex-post adaptation costs associated with the output decisions.

4.2. Shading Model with Observable Collusion

In our model, the agent of type θ feels entitled to the information rent (indirect utility) $U(\theta)$ indicated by the initial contract. Nevertheless, the supervisor reported $r = \theta$ and aggrieved (disappointed) the agent by exploiting the information rent $U(\theta)$. Then, the agent shades (punishes) the supervisor by $\beta U(\theta)$. So, the net payoff of the supervisor when he reports the truth $r = \theta$ is $\underbrace{W_S(\theta)}_{\text{Wage Payment}} - \underbrace{\beta U(\theta)}_{\text{shading loss}}$.

As for the principal's shading, there exists a subtle informational point. Our model is basically a hidden information model and the supervisor's signal θ is not observed by the principal. Otherwise (if the principal directly observed θ), he would not need the supervisor. We have already assumed that the supervisor, with probability p, obtains a proof (evidence) that the agent type is θ. Now suppose that the principal can know that the above state (of probability p) has happened, i.e., the supervisor has observed some signal θ. But suppose that he cannot know the exact value of θ, and also cannot verify that the supervisor has observed some signal θ. Then, if the supervisor provides no proof (evidence), the principal knows that the collusion has occurred (a side contract has been signed) between the agent of some type and the supervisor, though it is not verifiable. Only when the principal implements the initial scheme $\{X(\theta), W(\theta)\}$ and enforces $X(\theta)$ for the agent's report $\hat{\theta} = \theta$, he can know the exact value of θ, and understand how much he has been aggrieved by the supervisor. Then, he can shade the supervisor. In summary, this information structure means that collusion (side contracting) between the supervisor and the type θ agent is observable but unverifiable.

Then, formally, the principal would feel that she had been entitled to $X^{FB}(\theta) - C(X^{FB}(\theta), \theta)$, since the type information was θ. Nonetheless, he could only attain the payoff under asymmetric information regime between the principal and the agent θ, $X(\theta) - C(X(\theta), \theta) - U(\theta)$, since the supervisor colluded with the agent and hid the information θ. In summary, he was aggrieved by

$$\{[X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] - [X(\theta) - C(X(\theta), \theta) - U(\theta)]\} \tag{8}$$

and so he shades (punishes) the supervisor by a constanttimes the aggrievement level

$$\gamma\{[X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] - [X(\theta) - C(X(\theta),\theta) - U(\theta)]\}. \tag{9}$$

Thus, we obtain the supervisor's incentive constraint with behavioral assumptions

$$\begin{aligned} & \underbrace{W_S(\theta)}_{\text{wage payment}} - \underbrace{\beta U(\theta)}_{shadingloss} \geq \underbrace{kU(\theta)}_{sidepayment} \\ & \underbrace{-\gamma\{[X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] - [X(\theta) - C(X(\theta),\theta) - U(\theta)]\}}_{\text{shading loss}} \qquad (10) \\ \Longleftrightarrow \quad & \underbrace{W_S(\theta)}_{\text{wage payment}} \geq \underbrace{kU(\theta)}_{\text{side payment}} + \underbrace{(\beta-\gamma)U(\theta)}_{\text{shading loss by agent}} \\ & \underbrace{-\gamma\{[X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] - [X(\theta) - C(X(\theta),\theta)]\}}_{\text{shading loss by principal}} \qquad (10') \end{aligned}$$

Substituting $W(\theta) = C(X(\theta),\theta) + U(\theta)$ and

$$W_S(\theta) = kU(\theta)+(\beta-\gamma)U(\theta)-\gamma\{[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]-[X(\theta)-C(X(\theta),\theta)]\}$$

into the principal's objective function, we have the formulation of virtual surplus for type θ

$$\begin{aligned} & p(X^{FB}(\theta) - C(X^{FB}(\theta),\theta) - W_S(\theta)) + (1-p)(X(\theta) - C(X(\theta),\theta) - U(\theta)) \\ = \; & p(X^{FB}(\theta) - C(X^{FB}(\theta),\theta) - kU(\theta)) + (1-p)(X(\theta) - C(X(\theta),\theta)) \\ - \; & p\{\beta U(\theta) - \gamma[(X^{FB}(\theta) - C(X^{FB}(\theta),\theta)) - (X(\theta) - C(X(\theta),\theta) - U(\theta))]\}. \end{aligned}$$

Hence, the program of designing the optimal collusion-proof contract with behavioral elements can be rewritten as

$$\begin{aligned} \max_{X(\cdot)} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \left[\begin{array}{l} p(X^{FB}(\theta) - C(X^{FB}(\theta),\theta) - kU(\theta)) + (1-p)(X(\theta) - C(X(\theta),\theta) - U(\theta)) \\ -p\{\beta U(\theta) - \gamma[(X^{FB}(\theta) - C(X^{FB}(\theta),\theta)) - (X(\theta) - C(X(\theta),\theta) - U(\theta))]\} \end{array} \right] f(\theta)d\theta - Z \\ \text{s.t.} \quad & dX(\theta)/d\theta \geq 0 \quad (M) \quad \forall\theta. \end{aligned} \tag{11}$$

From the lemma 3, which shows the transformation of the expected information rent, the program becomes

$$\max_{X(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} \left[\begin{array}{c} (1-p)(X(\theta) - C(X(\theta),\theta)) + [(1-p)+pk]\frac{\partial C(X(\theta),\theta)}{\partial \theta}\frac{1-F(\theta)}{f(\theta)} \\ -p\left[\gamma(X(\theta) - C(X(\theta),\theta)) - (\beta-\gamma)\frac{\partial C(X(\theta),\theta)}{\partial \theta}\frac{1-F(\theta)}{f(\theta)}\right] \end{array} \right] f(\theta)d\theta$$
$$+ (1+\gamma)p\int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)]f(\theta)d\theta - Z$$
$$\text{s.t.} dX(\theta)/d\theta \geq 0 \quad (M) \quad \forall\theta.$$

We ignore the Monotonicity Constraint (M) and solve the relaxed program. The principal maximizes the expected value of the modified virtual surplus, denoted by $J^B(X,\theta)$. This expected value is maximized by simultaneously maximizing the modified virtual surplus for (almost) every type θ, i.e.

$$\begin{aligned} & X^B(\theta) \in \arg\max_{X(\cdot)} J^B(X,\theta) \\ = & \underbrace{(1-p)(X(\theta) - C(X(\theta),\theta)) + \frac{[(1-p)+pk]}{h(\theta)}\frac{\partial C(X(\theta),\theta)}{\partial \theta}}_{\text{Standard Virtual Surpluss}J(X,\theta)} \\ & \underbrace{-p\left[\gamma(X(\theta) - C(X(\theta),\theta)) - \frac{(\beta-\gamma)}{h(\theta)}\frac{\partial C(X(\theta),\theta)}{\partial \theta}\right]}_{\text{Virtual Loss through Behavioral Elements}} \end{aligned} \tag{12}$$

where $h(\theta) = f(\theta)/(1-F(\theta))$ is the hazard rate. This defines the optimal output rule $X^B(\cdot)$ for the program.[23] We take the derivative:

$$\begin{aligned} \frac{\partial J^B(X,\theta)}{\partial X} = & \underbrace{[1-p]\left[1 - \frac{\partial C(X,\theta)}{\partial X}\right] + \frac{[(1-p)+pk]}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}}_{\text{Standard Marginal Virtual Surpluss}J(X,\theta)} \\ & \underbrace{-p\left[\gamma\left[1 - \frac{\partial C(X,\theta)}{\partial X}\right] - \frac{(\beta-\gamma)}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}\right]}_{\text{Marginal Virtual Loss through Behavioral Elements}}. \end{aligned} \tag{13}$$

[23] The principal can design the optimal output rule $X^B(\cdot)$ to modify shading behaviors by controlling the potential for aggrievement, e.g. information rent $U(\theta)$. In that sense, our framework of shading model is similar to the idea of efficient organization design which counters "influence activities" by Milgrom (1988). The difference is that influence activities are made before an important decision making, while shading behaviors are made after an important and aggrieving decision making.

PROPOSITION 5. *The optimal solution $X^B(\theta)$ with behavioral elements is smaller than the solution $X^S(\theta)$ with no behavioral elements, that is, $X^B(\theta) \leq X^S(\theta)$.*

Proof. See, Appendix A.4 ▌

Theoretical Intuition

The supervisor's (collusion-proof) reward is, from (10'),

$$\begin{aligned} W_S(\theta) &= kU(\theta) + (\beta - \gamma)U(\theta) \\ &- \gamma\{[X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] - [X(\theta) - C(X(\theta), \theta)]\}. \quad (14) \end{aligned}$$

First, when the output $X(\theta)$ decreases marginally, the information rent $U(\theta)$ goes down.

Next, since $X(\theta) - C(X(\theta), \theta)$ goes down for $X(\theta) \leq X^{FB}(\theta)$, the potential aggrievement $[X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] - [X(\theta) - C(X(\theta), \theta)]$ increases, and the shading threat goes up $\gamma\{[X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] - [X(\theta) - C(X(\theta), \theta)]\} \Uparrow$. These two effects will decrease the supervisor's wage $W_S(\theta)$ discretely, which generates **a first-order gain**. Though the decrease in $X(\theta)$ generates **a second-order loss** through the change of optimal solution, the principal's profit will go up totally (due to the **first-order gain vs. second-order loss**). Thus, the optimal solution with behavioral elements (shading) $X^B(\theta)$ will fall below the optimal solution with no behavioral elements $X^S(\theta)$.

Now, we can perform a comparative statics on the optimal solution $X^B(\theta)$.

COROLLARY 2. *The optimal solution with behavioral elements $X^B(\theta)$ is nonincreasing in both parameter β (the degree of shading strength by the agent) and γ (the degree of shading strength by the principal).*

Proof. See Appendix A.5. ▌

Theoretical Intuition

The intuition is very close to the former argument. The supervisor's (collusion-proof) reward is, again from (10'),

$$\begin{aligned} W_S(\theta) &= kU(\theta) + (\beta - \gamma)U(\theta) \\ &- \gamma\{[X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] - [X(\theta) - C(X(\theta), \theta)]\}. \quad (15) \end{aligned}$$

Now, suppose that γ increases. If the optimal solution $X(\theta)$ decreases marginally, the information rent $U(\theta)$ goes down, and $X(\theta) - C(X(\theta), \theta)$ also goes down for $X(\theta) \leq X^{FB}(\theta)$. Therefore, the shading threat $\gamma\{[X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] - [X(\theta) - C(X(\theta), \theta)]\}$ goes up. These two effects will decrease the supervisor's wage $W_S(\theta)$ discretely, which generates **a first-order gain**. Though the decrease in $X(\theta)$ generates **a second-order loss** through the change of optimal solution, the principal's profit will go up totally (due to the **first-order gain vs. second-order loss**). Thus, the optimal solution with behavioral element (shading) $X^B(\theta)$ will decrease as the shading parameters β, γ increase.

PROPOSITION 6.

1. The principal's equilibrium payoff can increase more likely in the three-tier regime (B) with behavioral elements, in comparison with the three-tier regime (TH) with no behavioral elements, when the shading strength γ by the principal is greater than the shading strength β by the agent, i.e. $\gamma \geq \beta$.

2. The principal's equilibrium payoff tends to decrease in the three-tier regime (B) with behavioral elements, in comparison with the three-tier regime (TH) without behavioral elements, when the shading strength β by the agent is greater than the shading strength γ by the principal, i.e., $\beta \geq \gamma$. This is particularly so when p, γ are smaller.

Proof. See Appendix A.6 for Proposition 6.1. and See Appendix A.7 for Proposition 6.2. ∎

Rationale

Proposition 7.1 implies that under the information structure where collusion (side contracting) between supervisor and agent is observable ex post for the principal but unverifiable, the fear of being "shaded" by the principal can relax the supervisor's incentive constraint (coalition incentive constraint) discretely, thereby can increase the principal's equilibrium profit.[24] That is, the principal can reduce the reward to the supervisor discretely through his shading threat (γ times aggrievement), thereby increasing his profit.

[24] As an analogy for the moral hazard model with risk averse agent, we can say that the principal can decrease the risk cost (risk compensation) discretely, where the risk cost (risk compensation) corresponds to the shading cost in our paper. The point is that the principal ultimately bears the shading cost for the supervisor in order to satisfy his IR constraint.

The main point for Proposition 6.2 is that when $\beta \geq \gamma$, the net positive shading cost by the agent must be compensated for the supervisor by the principal.

4.3. Shading Model with Unobservable Collusion

Now, suppose that the supervisor's signalis $s \in \{\theta, \phi\}$ not observed at all by the principal ex post, that is, the principal cannot know at all ex post whether the supervisor obtained the informative signal (evidence, proof on θ) or not (ϕ), as well as which state θ has occurred. Then, the principal cannot distinguish whether she was aggrieved or whether the supervisor just obtained no informative signal (ϕ). Hence, the principal cannot shade the supervisor. This information structure means that collusion (side contracting) between supervisor and agent is unobservable, and thus the shading loss by the principal would be zero due to $\gamma = 0$.

Then, the supervisor's incentive constraint (coalition incentive constraint) is reduced to

$$\underbrace{W_S(\theta)}_{\text{wage payment}} - \underbrace{\beta U(\theta)}_{\text{shading loss}} \geq \underbrace{kU(\theta)}_{\text{side payment}} \iff \underbrace{W_S(\theta)}_{\text{wage payment}} \geq \underbrace{kU(\theta)}_{\text{side payment}} + \underbrace{\beta U(\theta)}_{\text{shading loss}}. \tag{16}$$

Hence, shading only by the agent $\beta > 0$ tightens the supervisor's incentive constraint (coalition incentive constraint), and makes it more likely that the supervisor will collude with the agent.

PROPOSITION 7. *Suppose that collusion (side contracting) between supervisor and agent is unobservable ex post for the principal. Then, only agent can shade the supervisor, which corresponds to $\beta > 0, \gamma = 0$. Then, the principal's equilibrium payoff is reduced in the regime with behavioral elements, in comparison with that without behavioral elements $\beta = \gamma = 0$. That is, "shading" becomes detrimental to organization design.*

Proof. The principal's virtual surplus is written as follows.

$$J^B(X,\theta) = \underbrace{(1-p)(X(\theta) - C(X(\theta),\theta)) + \frac{[(1-p)+pk]}{h(\theta)} \frac{\partial C(X(\theta),\theta)}{\partial \theta}}_{\text{(Standard) Virtual Surplus}} + \underbrace{\frac{p\beta}{h(\theta)}}_{\geq 0} \underbrace{\frac{\partial C(X(\theta),\theta)}{\partial \theta}}_{-}. \tag{17}$$

where $\underbrace{\frac{p\beta}{h(\theta)}}_{\geq 0}\underbrace{\frac{\partial C(X(\theta),\theta)}{\partial\theta}}_{-}$ is the increase in dead weight loss (information rent) through shading by the type θ agent, which decreases the principal's virtual surplus. This completes the proof. ∎

4.3.1. Collusion-proof Regime vs. Equilibrium Collusion Regime

Now, the principal has two options, one of which is the Collusion-proof Regime, where the principal deters the collusion between the agent θ and the supervisor through the collusion-proof constraint and induces the supervisor's truth telling $r = \theta$, and the other of which is the Equilibrium Collusion Regime, where the principal allows the collusion between them in equilibrium and induces the truthful information from the agent by himself, while the supervisor reports $r = \phi$. Which regime the principal chooses between the Collusion-proof regime and the Equilibrium Collusion regime depends on the condition, which will be analyzed below.

Collusion-proof Regime (CP)

In order to satisfy the collusion-proof constraint, the principal must set the reward for the supervisor

$$\underbrace{W_S(\theta)}_{\text{wage payment}} = \underbrace{kU(\theta)}_{\text{side payment}} + \underbrace{\beta U(\theta)}_{\text{shading loss by agent}} = (k+\beta)U(\theta). \tag{18}$$

Then, the virtual surplus for type θ is

$$(1-p)[X(\theta)-C(X(\theta),\theta)-U(\theta)]+p[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)-(k+\beta)U(\theta)].$$

Hence, the expected virtual surplus for the principal is, due to Lemma 3 and $U(\underline{\theta}) = 0$,

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}} \{(1-p)[X(\theta)-C(X(\theta),\theta)-U(\theta)]+p[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)-(k+\beta)U(\theta)]\}f(\theta)d\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \left\{(1-p)[X(\theta)-C(X(\theta),\theta)]+[(1-p)+p(k+\beta)]\frac{\partial C(X(\theta),\theta)}{\partial\theta}\frac{1-F(\theta)}{f(\theta)}\right\}f(\theta)d(\theta) \\
&+ \int_{\underline{\theta}}^{\overline{\theta}} p[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta. \qquad (19)
\end{aligned}$$

The principal simultaneously maximizes the modified virtual surplus for (almost) every type θ, i.e.

$$\underbrace{(1-p)(X(\theta)-C(X(\theta),\theta))+[(1-p)+p(k+\beta)]\frac{\partial C(X(\theta),\theta)}{\partial\theta}\frac{1-F(\theta)}{f(\theta)}}_{J^B_{CP}(X,\theta)} \tag{20}$$

Or

$$\underbrace{(X(\theta)-C(X(\theta),\theta))}_{\text{Total Surplus}}+\underbrace{\left[1+\frac{pk}{1-p}\right]\frac{\partial C(X(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}}_{\text{Information Rent}} + \underbrace{\frac{p\beta}{1-p}\frac{\partial C(X(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}}_{\text{Increase in Information Rent through Shading}} . \tag{20'}$$

First order condition for the optimality is

$$\frac{\partial J^B_{CP}(X,\theta)}{\partial X}=\underbrace{[1-p]\left[1-\frac{\partial C(X,\theta)}{\partial X}\right]+\frac{[(1-p)+pk]}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}}_{\text{Marginal Virtual Surplus}} + \underbrace{\frac{p\beta}{h(\theta)}\underbrace{\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}}_{-}}_{\text{Marginal Shading Cost}}=0 \tag{21}$$

$$\iff \underbrace{\left[1-\frac{\partial C(X,\theta)}{\partial X}\right]+\frac{\left[1+\frac{pk}{1-p}\right]}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}}_{\text{Marginal Virtual Surplus}}+\underbrace{\frac{p\beta}{1-p}\frac{1}{h(\theta)}\underbrace{\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}}_{-}}_{\text{Marginal Shading Cost}}=0. \tag{21'}$$

PROPOSITION 8. *The optimal solution $X^B_{CP}(\theta)$ with behavioral elements under the collusion-proof regime is smaller than the optimal solution $X^S(\theta)$ with no behavioral elements, that is, $X^B_{CP}(\theta)\leq X^S(\theta)$.*

Proof. See the proof of Proposition 6. This is the case where $\beta>0,\gamma=0$ ∎

Equilibrium Collusion Regime (EC)

In this regime, when the supervisor obtains the proof on θ with probability p, the principal allows the collusion between the agent θ and the supervisor in equilibrium, which means that the supervisor reports $r = \phi$ and the agent θ self-selects $\{X(\theta), W(\theta)\}$ and obtains the information rent $U(\theta)$. Then, the principal pays the information rent $U(\theta)$ to the agent θ at the unit transfer price 1.

Now, the virtual surplus for type θ is

$$\begin{aligned}&(1-p)[X(\theta) - C(X(\theta),\theta) - U(\theta)] + p[X(\theta) - C(X(\theta),\theta) - U(\theta)]\\ =&X(\theta) - C(X(\theta),\theta) - U(\theta).\end{aligned}$$

Hence, the expected virtual surplus for the principal is, due to Lemma 3 and $U(\underline{\theta}) = 0$,

$$\begin{aligned}&\int_{\underline{\theta}}^{\overline{\theta}} \{X(\theta) - C(X(\theta),\theta) - U(\theta)\} f(\theta) d\theta\\ =&\int_{\underline{\theta}}^{\overline{\theta}} \left\{ X(\theta) - C(X(\theta),\theta) + \frac{\partial C(X(\theta),\theta)}{\partial \theta} \frac{1-F(\theta)}{f(\theta)} \right\} f(\theta) d\theta.\end{aligned}$$

Then, the principal simultaneously maximizes the modified virtual surplus for (almost) every type θ, $\underbrace{X(\theta) - C(X(\theta),\theta) + \frac{\partial C(X(\theta),\theta)}{\partial \theta} \frac{1}{h(\theta)}}_{J^B_{EC}(X,\theta)}$. First order condition for the optimality is

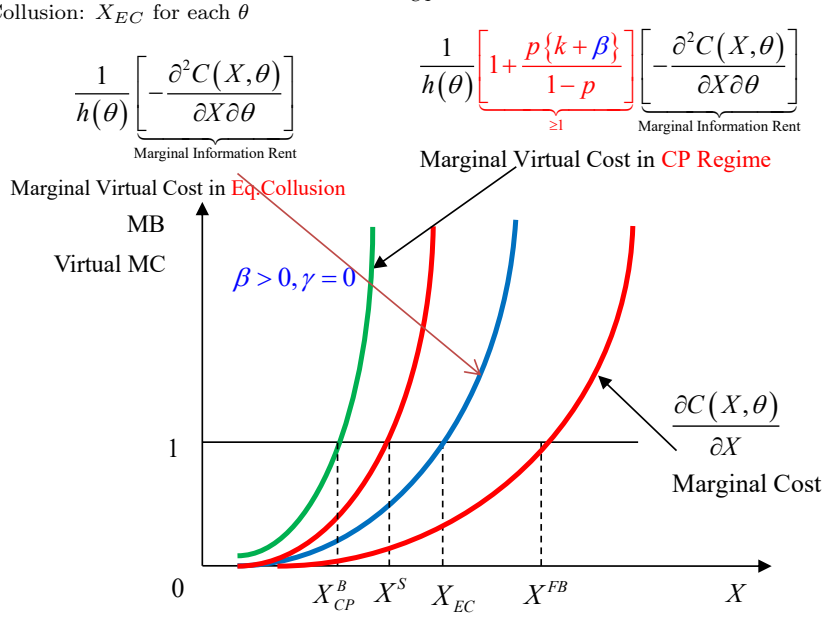
$$1 - \frac{\partial C(X(\theta),\theta)}{\partial X(\theta)} + \frac{\partial^2 C(X(\theta),\theta)}{\partial X \partial \theta} \frac{1}{h(\theta)} = 0. \tag{22}$$

Comparing First Order Conditions on marginal incentives in the two regimes (CP and EC), (21') and (22), we find that the coefficient of the marginal virtual cost $1 + \frac{p(k+\beta)}{1-p}$ in the collusion-proof regime (CP) is greater than that 1 in the equilibrium collusion regime (EC), that is, $1 + \frac{p(k\beta}{1-p} \geq 1$ for $\forall p, k, \beta \geq 0$. Hence, we have $X^B_{CP}(\theta) \leq X_{EC}(\theta) = X^{SB}(\theta)$. The below figure shows the determination of equilibrium incentives for type θ.

4.3.2. Payoff Comparison between Collusion-proof and Equilibrium Collusion Regimes

We analyze which regime the principal chooses between the Collusion-proof regime (CP) and the Equilibrium Collusion Regime (EC). We compare the payoffs between Collusion-proof vs. Equilibrium Collusion Regimes.

FIG. 2. Comparison of Equilibrium Outputs between Collusion-proof and Equilibrium Collusion Regimes Collusion-proof: X^B_{CP} w.p $1-p$ and X^{FB} w.p p vs. Equilibrium Collusion: X_{EC} for each θ



The expected payoff for the principal in the 'Collusion-proof' regime (CP) is

$$
\begin{aligned}
&(1-p)\int_{\underline{\theta}}^{\overline{\theta}} [X^B_{CP}(\theta) - C(X^B_{CP}(\theta),\theta)] f(\theta) d\theta \\
&+p \times \int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] f(\theta) d\theta \qquad (23) \\
&+[(1-p)+p(k+\beta)] \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)} \frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial \theta} f(\theta) d\theta.
\end{aligned}
$$

The expected payoff for the principal in the Equilibrium Collusion Regime (EC) is

$$
\int_{\underline{\theta}}^{\overline{\theta}} \left[X(\theta) - C(X(\theta),\theta) + \frac{1}{h(\theta)} \frac{\partial C(X(\theta),\theta)}{\partial \theta} \right] f(\theta) d\theta. \qquad (24)
$$

We consider the comparison between the two regimes under $Z = 0$, that is, the cost of introducing the supervisor (a transaction cost) is zero. Then, we have the following proposition.

PROPOSITION 9. *Collusion-proof vs. Equilibrium Collusion*

1. The principal prefers the Collusion-proof regime (CP) to the Equilibrium Collusion regime (EC) in terms of his expected payoff when the shading strength $\beta \leq 1 - k$.

2. The principal prefers the Collusion-proof regime (CP) to the Equilibrium Collusion regime (EC) for all $\beta \geq 0$ *when the accuracy of supervision* $p \geq p^*$. *Especially, as* $\beta \to +\infty$, *the optimal Collusion-proof contract has a property of "Shut Down" in all states of supervisory no information* (θ, ϕ).

3. The principal prefers the Equilibrium Collusion regime (EC) to the Collusion-proof regime (CP) when the shading strength $\beta > \beta^*$ *and the accuracy of supervision* $p < p^*$.

Proof. See the Appendix A.8. ▌

Rationale

As the degree of shading β ("threat" by the agent) increases, the incentive for collusion between the agent of type θ and the supervisor increases. Thereby, it becomes more costly for the principal to impose collusion-proof schemes and deter collusion, and to induce the truth telling from the supervisor. Theoretically, this implies that as the set of collusion-proof, Incentive compatible schemes becomes smaller, the attainable efficiency becomes lower.

Then, it may be better for the principal to allow collusion between the agent of type θ and the supervisor, and then attain the higher efficiency through discretely reducing the ex-post aggrievement and shading by the agent of type θ.[25]

This is **a new idea** in the Collusion literature a la Tirole (1986, 1992)[26] in that the increase in shading pressure (behavioral element) strengthens the incentive for collusion, thereby makes it difficult to implement the collusion-proof (Supervisor's truth telling) incentive schemes, which leads to the Equilibrium Collusion. The principal allows collusion between the agent and the supervisor in equilibrium, and the supervisor reports $r = \phi$ ("I did not observe any information") and the agent of type θ reveals his

[25] A clear understanding is that as β approaches to ∞ and p approaches to 0, the two player contract (Equilibrium Collusion) dominates, since the supervisor becomes very expensive to maintain relative to the probability that he will bring informative evidence. (Proposition 9.3).

[26] Tirole (1992) explains the intuitive ideas of the cases where Equilibrium Collusion can be optimal in the three-tier hierarchical structures with no behavioral elements. For the existing literature (with no behavioral elements) on the optimality of allowing collusion, see e.g., Itoh (1993) , Kofman and Lawarree (1996), and Suzuki (1997).

type information θ by self-selecting $\{X(\theta), W(\theta)\}$ and obtains the information rent $U(\theta)$.

Interpretation of the Result

We can interpret the results from the viewpoint of Transaction Cost Economics a la Coase (1937) and Williamson (1975). Let us assume that "Haggling Cost" in Transaction Cost Economics has two sources: Cost of Rent-seeking or Influence activity which accompanies Ex-ante Collusion before the supervisor's decision making (report), and Cost of Ex-post Shading which results from Ex-post aggrievement and shading behavior after the supervisor's decision making (report), as the below figure suggests.

Two Sources of Haggling Costs
(CP) Collusion—Proof but Ex-post Shading
(EC) Equilibrium Collusion but Ex-post No Shading

In the Collusion-proof regime, the principal deters collusion through collusion-proof schemes, and thus no ex-ante collusion occurs. But, ex-post shading by the agent of type θ occurs, since the agent of type θ expected to obtain the best reward for him, that is, the information rent $U(\theta)$, but was aggrieved to have lost it due to the supervisory report $r = \theta$. Therefore, the agent of type θ shades the supervisor by the shading parameter β times the aggrievement level $U(\theta)$. In this case, we have ex-ante no collusion costs but ex-post shading costs.

On the other hand, **in the Equilibrium Collusion Regime**, the principal allows ex-ante collusion between the agent of type θ and the supervisor, which may be costly by itself but does not generate any aggrievement for the agent of type θ, since he can indeed obtain the information rent $U(\theta)$ (as his "entitlement"). Hence, he does not shade the supervisor ex-post. In this case, we have ex-ante collusion costs but ex-post no shading costs.[27]

As the degree of shading β increases, the incentive for collusion between the agent of type θ and the supervisor increases. Thereby, it becomes more costly for the principal to impose collusion-proof schemes and deter

[27] Note that the trade-off here is different from the trade-off between gross total surplus and ex-post haggling cost at the optimal organization design problem. The latter trade-off is represented by mathematical formulas, e.g. (12) and (16).

collusion, and to induce the truth telling from the supervisor. Then, it can be better for the principal to let them collude in equilibrium, and attain the higher efficiency through reducing discretely the ex-post aggrievement and shading by the agent of type θ.

We believe that this is not only a new idea in the Collusion literature a la Tirole (1986, 1992) in that the increase in shading pressure (behavioral element) strengthens the incentive for collusion, thereby makes it difficult to implement the collusion-proof (Supervisor's truth telling) incentive schemes, which leads to the Equilibrium Collusion, but also gives a micro-foundation (an explicit modeling) for the "Ex-post Haggling Cost" in Transaction Cost Economics a la Williamson (1975).[28]

5. CONCLUSION

We applied the First Order (Mirrlees) Approach and the Monotone Comparative Statics method to the continuous-type, three-tier agency model with hidden information and collusion à la Tirole (1986), thereby providing a framework that can address the issues treated in the existing literature in a much simpler fashion. We characterized the nature of equilibrium contract that can be implemented under the possibility of collusion, and obtained a general comparison result on the organization structures. Then, we introduced the recent behavioral contract theory idea, "shading" (Hart and Moore (2008)) into the model. By integrating the two ideas, i.e., collusion and shading, we could not only enrich the existing collusion model, including a new result on the choice of Collusion-proof vs. Equilibrium Collusion regimes, but also gave a micro foundation to ex-post haggling costs, addressed by Transaction Cost Economics (TCE). By using this model, we examined the optimal organizational design problem as an optimal response to the trade-off between gross total surplus and ex-post haggling costs. This is also close to the idea of efficient organization design which counters influence activities by Milgrom (1988). We believe that our model can help a deep understanding of resource allocation and decision process in hierarchical organizations.

[28]Strictly speaking, our paper may only exogenously have introduced a type of haggling cost, shading, by applying the idea of Hart and Moore (2008). Indeed, no micro-model is constructed to explain how the haggling cost arises. Nonetheless, we endogenously examined the optimal organizational design problem as an optimal response to the trade-off between gross total surplus and ex-post haggling costs. This could be evaluated to have taken a step further the idea of efficient organization design which counters influence activities (Milgrom (1988)), by incorporating ex-post haggling (shading) costs into the model.

APPENDIX

A.1. PROOF OF LEMMA 2

Proof. The "⇒" part was established above. It remains to show that (ICFOC) and monotonicity (M) imply that $U(\hat{\theta},\theta) \leq U(\theta)$ for all $\hat{\theta},\theta$. For $\hat{\theta} > \theta$, we can write

$$\begin{aligned}
U(\hat{\theta},\theta) - U(\theta) =& W(\hat{\theta}) - C(X(\hat{\theta}),\theta) - U(\theta) \\
=& U(\hat{\theta}) + C(X(\hat{\theta}),\hat{\theta}) - C(X(\hat{\theta}),\theta) - U(\theta) \\
=& [C(X(\hat{\theta}),\hat{\theta}) - C(X(\hat{\theta}),\theta)] + [U(\hat{\theta}) - U(\theta)] \\
=& \int_{\theta}^{\hat{\theta}} \frac{\partial C(X(\hat{\theta}).\tau)}{\partial \tau} d\tau + \int_{\theta}^{\hat{\theta}} \left[-\frac{\partial C(X(\tau),\tau)}{\partial \tau} \right] d\tau \qquad \text{(A.1)} \\
=& \int_{\theta}^{\hat{\theta}} \left[\frac{\partial C(X(\hat{\theta}),\tau)}{\partial \tau} - \frac{\partial C(X(\tau),\tau)}{\partial \tau} \right] d\tau \leq 0. \qquad \text{(A.2)}
\end{aligned}$$

In (A.1), we used the following fact by (ICFOC) and Envelope theorem

$$U(\hat{\theta}) - U(\theta) = \int_{\theta}^{\hat{\theta}} \frac{dU}{d\tau}(\tau) d\tau = \int_{\theta}^{\hat{\theta}} -\frac{\partial C(X(\tau),\tau)}{\partial \tau} d\tau.$$

In (A.2), the last inequality is obtained by SCP and the fact that $X(\hat{\theta}) \geq X(\theta)$ by (M). As explained just below the Definition 1, SCP implies that the marginal cost of output $\frac{\partial C(X,\theta)}{\partial X}$ is decreasing in type θ in our model. That is, $\frac{\partial^2 C(X,\theta)}{\partial X \partial \theta} < 0$. This condition implies that $\frac{\partial C(X(\hat{\theta}),\theta)}{\partial \theta} - \frac{\partial C(X(\theta),\theta)}{\partial \theta} \leq 0$ for $X(\hat{\theta}) \geq X(\theta)$ due to (M). So, we obtain the last inequality. The proof for $\theta > \hat{\theta}$ is similar. ∎

A.2. PROOF OF LEMMA 3

Proof. We transform the expected information rents by exploiting "Integration by Parts". Because

$$[U(\theta)F(\theta)]' = U(\theta)f(\theta) + \underbrace{\frac{dU(\theta)}{d\theta}}_{\text{(Due to Envelope Theorem)}} F(\theta) = U(\theta)f(\theta) - \underbrace{\frac{\partial C(X(\theta),\theta)}{\partial \theta}}_{\text{(Due to Envelope Theorem)}} F(\theta),$$

and so $U(\theta)f(\theta) = [U(\theta)F(\theta)]' + \frac{\partial C(X(\theta),\theta)}{\partial \theta}F(\theta)$, we have

$$\begin{aligned}
\int_{\underline{\theta}}^{\overline{\theta}} U(\theta)f(\theta)d\theta =& [U(\theta)F(\theta)]_{\underline{\theta}}^{\overline{\theta}} + \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta}F(\theta)d\theta \\
=& U(\overline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta}F(\theta)d\theta \\
=& U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta}d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta}F(\theta)d\theta \\
& \left(\because U(\overline{\theta}) = U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta}d\theta\right) \\
=& U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta}(1 - F(\theta))d\theta \\
=& U(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial C(X(\theta),\theta)}{\partial \theta}\frac{1 - F(\theta)}{f(\theta)}f(\theta)d\theta.
\end{aligned}$$

∎

A.3. PROOF OF PROPOSITION 4

Proof. First, by definition, $X^S(\theta)$ is the optimal decision over the problem

$$\begin{aligned}
&\max_{X(\cdot)}(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X(\theta) - C(X(\theta),\theta)]f(\theta)d\theta \\
&+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X(\theta),\theta)}{\partial \theta}f(\theta)d\theta.
\end{aligned}$$

Similarly, by definition,$X^{TW}(\theta)$ is the optimal decision over the problem

$$\max_{X(\cdot)}\int_{\underline{\theta}}^{\overline{\theta}}\left[X(\theta) - C(X(\theta),\theta) + \frac{\partial C(X(\theta),\theta)}{\partial \theta}\right]d(\theta)d\theta.$$

We also set the following payoff function

$$\int_{\underline{\theta}}^{\overline{\theta}}\left[X(\theta) - C(X(\theta),\theta) + \underbrace{[(1-p)+pk]}_{\leq 1}\frac{1}{h(\theta)}\frac{\partial C(X(\theta),\theta)}{\partial \theta}\right]f(\theta)d\theta$$

where the principal hypothetically implements an output function which varies only with the type θ of the agent. Coefficient $(1-p)+pk$ reflects the fact that the principal pays the full information rent $U(\theta)$ to the agent when there is no hard evidence and pays a fraction of it $kU(\theta)$ to the supervisor when there is hard evidence on θ. Let $X^P(\theta)$ be the optimal decision over this "Pooling" regime. Then, the equilibrium payoff of the problem is,

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[X^P(\theta) - C(X^P(\theta),\theta) + \underbrace{[(1-p)+pk]}_{\leq 1} \frac{1}{h(\theta)} \frac{\partial C(X^P(\theta),\theta)}{\partial \theta} \right] f(\theta) d\theta. \tag{A.3}$$

From the revealed preference argument, the following two inequalities hold.

$$\begin{aligned}
&(1-p)\int_{\underline{\theta}}^{\overline{\theta}} [X^S(\theta) - C(X^S(\theta),\theta)] f(\theta) d\theta \\
&+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)} \frac{\partial C(X^S(\theta),\theta)}{\partial \theta} f(\theta) d\theta \\
\geq &(1-p)\int_{\underline{\theta}}^{\overline{\theta}} [X^P(\theta) - C(X^P(\theta),\theta)] f(\theta) d\theta \qquad \text{(A.4)} \\
&+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)} \frac{\partial C(X^P(\theta),\theta)}{\partial \theta} f(\theta) d\theta \\
&\underbrace{p\int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] f(\theta) d\theta}_{\text{First Best Expected Total Surplus}} \\
\geq &p\int_{\underline{\theta}}^{\overline{\theta}} [X^P(\theta) - C(X^P(\theta),\theta)] f(\theta) d\theta. \qquad \text{(A.5)}
\end{aligned}$$

Adding them up, we obtain the first comparison result:

$$\begin{aligned}
&(1-p)\int_{\underline{\theta}}^{\overline{\theta}} [X^S(\theta) - C(X^S(\theta),\theta)] f(\theta) d\theta \\
&+\underbrace{p\int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] f(\theta) d\theta}_{\text{First Best Expected Total Surplus}} \\
&+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)} \frac{\partial C(X^S(\theta),\theta)}{\partial \theta} f(\theta) d\theta
\end{aligned}$$

$$\geq (1-p)\int_{\underline{\theta}}^{\overline{\theta}} [X^P(\theta) - C(X^P(\theta),\theta)]f(\theta)d\theta$$

$$+p\int_{\underline{\theta}}^{\overline{\theta}} [X^P(\theta) - C(X^P(\theta),\theta)]f(\theta)d\theta \tag{A.6}$$

$$+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)}\frac{\partial C(X^P(\theta),\theta)}{\partial\theta}f(\theta)d\theta$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} \left[X^P(\theta) - C(X^P(\theta),\theta) + [(1-p)+pk]\frac{1}{h(\theta)}\frac{\partial C(X^P(\theta),\theta)}{\partial\theta}\right] f(\theta)d\theta. \tag{A.7}$$

This result shows that the principal can do better by conditioning the output on whether there is hard evidence $r = \theta$ or not $r = \phi$. When there is no hard evidence with probability $1-p$, the principal can set the output at $X^S(\theta)$ and pay the full information rent $U(\theta)$ to the agent. On the other hand, if the supervisor reports hard evidence $r = \theta$ with probability p, the principal sets the output level at the First best $X^{FB}(\theta)$ and pays wage $kU(\theta)$ to the supervisor. That is, the "Three-tier" regime with supervision (TH) is preferable to the "Pooling" regime (PL).

Next, from the revealed preference argument, the second comparison holds.

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[X^P(\theta) - C(X^P(\theta),\theta) + [(1-p)+pk]\frac{\partial C(X^P(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}\right] f(\theta)d\theta$$

$$\geq \int_{\underline{\theta}}^{\overline{\theta}} \left[X^{TW}(\theta) - C(X^{TW}(\theta),\theta) + [(1-p)+pk]\frac{\partial C(X^{TW}(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}\right] f(\theta)d\theta$$

$$+\underbrace{p(1-k)}_{\geq 0}\underbrace{\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)}\frac{\partial C(X^{TW}(\theta),\theta)}{\partial\theta}f(\theta)d\theta}_{-} \tag{A.8}$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} \left[X^{TW}(\theta) - C(X^{TW}(\theta),\theta) + \frac{1}{h(\theta)}\frac{\partial C(X^{TW}(\theta),\theta)}{\partial\theta}\right] f(\theta)d\theta.$$

In comparison to the Two-tier hierarchy, the principal is better off in the "Pooling" regime, since the distortion on the output is lower $X^{TW}(\theta) \leq X^P(\theta) \leq X^{FB}(\theta)$, $\forall\theta$ due to $(1-p)+pk \leq 1$. That is, the "Pooling" regime is preferable to 'Two-tier' regime with no supervision (TW).

Combining these two comparison results, we obtain the proposition. ∎

A.4. PROOF OF PROPOSITION 5

Proof. Since

$$\begin{aligned}\frac{\partial J^S(X,\theta)}{\partial X} =&(1-p)\left(1-\frac{\partial C(X,\theta)}{\partial X}\right)+\frac{[(1-p)+pk]}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}=0 \text{ at } X=X^S(\theta)\\ \iff& \frac{1}{h(\theta)}\frac{\partial^2 C(X^S(\theta),\theta)}{\partial X\partial\theta}=-\frac{(1-p)}{[(1-p)+pk]}\left(1-\frac{\partial C(X^S(\theta),\theta)}{\partial X}\right).\end{aligned}$$

we find from (13) that

$$\begin{aligned}\frac{\partial J^B(X,\theta)}{\partial X} =&-p\gamma\left(1-\frac{\partial C(X^S(\theta),\theta)}{\partial X}\right)+\frac{p(\beta-\gamma)}{h(\theta)}\frac{\partial^2 C(X^S(\theta),\theta)}{\partial X\partial\theta} \text{ at } X=X^S(\theta)\\ =&-p\gamma\left(1-\frac{\partial C(X^S(\theta),\theta)}{\partial X}\right)-\frac{p(\beta-\gamma)(1-p)}{[(1-p)+pk]}\left(1-\frac{\partial C(X^S(\theta),\theta)}{\partial X}\right)\\ =&-p\left(\gamma+\frac{(\beta-\gamma)(1-p)}{[(1-p)+pk]}\right)\left(1-\frac{\partial C(X^S(\theta),\theta)}{\partial X}\right).\end{aligned}$$

Since $1-\frac{\partial C(X^S(\theta),\theta)}{\partial X}\geq 0$ for $X^S(\theta)\leq X^{FB}(\theta)$, the sign of $\frac{\partial X^B(X^S(\theta),\theta)}{\partial X}$ depends on $-p\left(\gamma+\frac{(\beta-\gamma)(1-p)}{[(1-p)+pk]}\right)$. We easily see that $\gamma\geq\frac{(\gamma-\beta)(1-p)}{[(1-p)+pk]} \iff \frac{(1-p)+pk}{(1-p)}\geq\frac{\gamma-\beta}{\gamma} \iff 1+\frac{pk}{1-p}\geq 1-\frac{\beta}{\gamma}$ holds for any $0\leq p,k\leq 1$, and $\beta,\gamma\geq 0$.

Then, since $\frac{\partial J^B(X^S,\theta)}{\partial X}$ evaluated at $X=X^S(\theta)$, $X^S(\theta)$ cannot be optimal for the behavioral regimes. A marginal decrease in $X(\theta)$ from $X^S(\theta)$ would increase the virtual surplus $J^B(X,\theta)$ of the behavioral regime. Hence, we have $X^B(\theta)\leq X^S(\theta)$.[29]

∎

A.5. PROOF OF COROLLARY 2

Proof. From (13), the derivative $J^B_X(X,\theta)$ is nonincreasing in β (behavioral elements). That is, $J^B_{X\beta}(X,\theta)=\frac{p}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}\leq 0$. Hence, the optimal solution $X^B(\theta)$ is nonincreasing in β. Further, $J^B_{X\gamma}(X,\theta)=-p\left\{\left[1-\frac{\partial C(X,\theta)}{\partial X}\right]+\frac{1}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}\right\}$. We already know that $\frac{\partial J(X,\theta)}{\partial X}=\left[1-\frac{\partial C(X,\theta)}{\partial X}\right]+\frac{1}{h(\theta)}\frac{\partial^2 C(X,\theta)}{\partial X\partial\theta}=0$ at $X=X(\theta)$.

[29] This result can be obtained also from the comparison in virtual marginal cost between two regimes: No Behavioral (TH) and Behavioral (B) regimes.

Then, since $X^B(\theta) \leq X(\theta)$ from proposition 5, we have $\frac{\partial J(X,\theta)}{\partial X} \geq 0$ at $X = X^B(\theta)$. Hence, we have $J^B_{X\gamma}(X,\theta) \leq 0$, which means that the optimal solution $X^B(\theta)$ is nonincreasing in γ. In sum, the output downward distortion is increasing in β and γ. ▌

A.6. PROOF OF PROPOSITION 6.1

Proof. First, the virtual surplus for type θ in the three-tier, no behavioral regime (TH) is

$$\begin{aligned}&p[X^{FB}(\theta) - C(X^{FB}(\theta),\theta) - kU(\theta)] + (1-p)(X(\theta) - C(X(\theta),\theta) - U(\theta))\\ =&(1-p)(X(\theta) - C(X(\theta),\theta)) + p[X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] - [(1-p)+pk]U(\theta).\end{aligned}$$

Hence the maximized expected virtual surplus in the three-tier regime (TH) is, by using the lemma 3,

$$\begin{aligned}&(1-p)\int_{\underline{\theta}}^{\overline{\theta}} [X^S(\theta) - C(X^S(\theta),\theta)]f(\theta)d\theta\\ &+p\underbrace{\int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)]f(\theta)d\theta}_{\text{First Best Expected Total Surplus}}\\ &+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)}\frac{\partial C(X^S(\theta),\theta)}{\partial\theta} f(\theta)d\theta.\end{aligned}$$

Next, the principal's virtual surplus for type θ in the Behavioral regime (B) is

$$p[X^{FB}(\theta) - C(X^{FB}(\theta),\theta) - W_S(\theta)] + (1-p)(X(\theta) - C(X(\theta),\theta) - U(\theta)).$$

By remembering the following coalition-proof constraint with behavioral elements

$$W_S(\theta) - \underbrace{\beta U(\theta)}_{\substack{\text{Shading Loss}\\ \text{by the Agent}}} \geq kU(\theta) - \underbrace{\gamma\left\{[X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] - (X(\theta) - C(X(\theta),\theta) - U(\theta))\right\}}_{\text{Shading Loss by the Principal}}$$

the virtual surplus for type θ in the Behavioral regime (B) is transformed as follows.

$$\begin{aligned}&p[X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] + (1-p)(X(\theta) - C(X(\theta),\theta)) - [(1-p)+pk]U(\theta)\\ &+p\gamma\{[X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] - [X(\theta) - C(X(\theta),\theta)]\} - p(\beta-\gamma)U(\theta).\end{aligned}$$

Now, the expected virtual surplus is written as follows by **the lemma 3**.

$$
\begin{aligned}
&(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X(\theta)-C(X(\theta),\theta)]f(\theta)d\theta \\
&+p\underbrace{\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta}_{\text{First Best Expected Total Surplus}} \\
&+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h9\theta}\frac{\partial C(X(\theta),\theta)}{\partial\theta}f(\theta)d\theta \\
&+p\gamma\left\{\int_{\underline{\theta}}^{\overline{\theta}}[(X^{FB}(\theta)-C(X^{FB}(\theta),\theta))-(X(\theta)-C(X(\theta),\theta))]f(\theta)d\theta\right\} \\
&+p(\beta-\gamma)\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X(\theta),\theta)}{\partial\theta}f(\theta)d\theta.
\end{aligned}
$$

Let $X^B_{CP}(\theta)$ be the optimal output over the maximization problem for type θ

$$
\begin{aligned}
X^B_{CP}(\theta)\in\arg\max_{X(\cdot)}J^B(X,\theta)=&\underbrace{(1-p)(X(\theta)-C(X(\theta),\theta))+\frac{[(1-p)+pk]}{h(\theta)}\frac{\partial C(X(\theta),\theta)}{\partial\theta}}_{\text{Standard Virtual Surplus}J(X,\theta)} \\
&\underbrace{-p\left[\gamma(X(\theta)-C(X(\theta),\theta))-\frac{(\beta-\gamma)}{h(\theta)}\frac{\partial C(X(\theta),\theta)}{\partial\theta}\right]}_{\text{Virtual Loss through Behavioral Elements}}.
\end{aligned}
\tag{A.9}
$$

Then, the maximized expected virtual surplus in the behavioral regime (B) is transformed as follows.

$$
\begin{aligned}
&(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X^B_{CP}(\theta)-C(X^B_{CP}(\theta),\theta)]f(\theta)d\theta \\
&+p\underbrace{\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta}_{\text{First Best Expected Total Surplus}} \\
&+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial\theta}f(\theta)d\theta \\
&+p\gamma\left\{\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta-\int_{\underline{\theta}}^{\overline{\theta}}[X^B_{CP}(\theta)-C(X^B_{CP}(\theta),\theta)]f(\theta)d\theta\right\}
\end{aligned}
\tag{A.10}
$$

$$+p(\beta-\gamma)\int_{\underline{\theta}}^{]\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial\theta}f(\theta)d\theta.$$

Hence, the condition for the principal's equilibrium profit to increase in the behavioral regime (B) relative to the standard three-tier regime (TH) is as follows.

$$p\gamma\left\{\underbrace{\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta}_{\text{First Best Expected Total Surplus}}-\int_{\underline{\theta}}^{\overline{\theta}}[X^B_{CP}(\theta)-C(X^B_{CP}(\theta),\theta)]f(\theta)d\theta\right\}$$

$$+p(\beta-\gamma)\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\underbrace{\frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial\theta}}_{-}f(\theta)d\theta \tag{A.11}$$

$$\geq(1-p)\int_{\underline{\theta}}^{\overline{\theta}}\{[X^S(\theta)-C(X^S(\theta),\theta)]-[X^B_{CP}(\theta)-C(X^B_{CP}(\theta),\theta)]\}f(\theta)d\theta$$

$$+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\left\{\frac{\partial C(X^S(\theta),\theta)}{\partial\theta}-\frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial\theta}\right\}f(\theta)d\theta.$$

The RHS of the inequality is the payoff difference between $X^S(\theta)$ and $X^B_{CP}(\theta)$ coming from the following revealed preference relation:

$$(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X^S(\theta)-C(X^S(\theta),\theta)]f(\theta)d\theta+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^S(\theta),\theta)}{\partial\theta}f(\theta)d\theta$$

$$\geq(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X^B_{CP}(\theta)-C(X^B_{CP}(\theta),\theta)]f(\theta)d\theta+[(1-p)+pk]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial\theta}f(\theta)d\theta. \tag{A.12}$$

The LHS of the inequality is totally the principal's payoff increase through discretely relaxing the coalition incentive constraint by the principal's shading threat $\gamma\geq\beta$. That is, the principal can reduce the reward to the supervisor discretely through his shading threat (γ times aggrievement) to the supervisor, thereby increasing his profit. ▌

Remark on Proposition 6.1:
The supervisor's equilibrium payoff under shading is, from (14),

$$W_S(\theta)-\beta U(\theta)=kU(\theta)-\gamma\{[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]-[X(\theta)-C(X(\theta),\theta)-U(\theta)]\}.$$

Thus, the condition for the supervisor's IR constraint to be satisfied is

$$kU(\theta)-\underbrace{\gamma\{[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]=[X(\theta)-C(X(\theta),\theta)-U(\theta)]\}}_{\text{Shading Loss}}\geq 0,\forall\theta\in[\underline{\theta},\overline{\theta}]. \tag{A.13}$$

This requires that the shading by the principal is not too strong. Hence, a necessary condition under which (1) the principal's equilibrium profit more likely increases by the introduction of the behavioral elements and (2) his IR constraint also holds is $\beta\leq\gamma\leq\frac{U(\theta)}{(\text{FBprofit})-(\text{SBprofit})}k$, more concretely,

$$\beta\leq\gamma\leq\min_{\theta}\frac{U(\theta)}{\underbrace{\{[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]-[X(\theta)-C(X(\theta),\theta)-U(\theta)]\}}_{\text{Aggrievement}}}k,\forall\theta\in[\underline{\theta},\overline{\theta}]. \tag{A.14}$$

A.7. PROOF OF PROPOSITION 6.2

Proof. When $\beta\geq\gamma$, the second term of the LHS of the corresponding inequality in Proposition 7.1 $p\underbrace{(\beta-\gamma)}_{\geq 0}\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\underbrace{\frac{\partial C(X^{B}_{CP}(\theta),\theta)}{\partial\theta}}_{-}f(\theta)d\theta\leq 0$. That is, when $\beta\geq\gamma$, the net positive shading cost by the agent must be compensated for the supervisor by the principal. Only the first term of the LHS is positive, which becomes smaller when p,γ are smaller. This makes the inequality more difficult to hold. ∎

A.8. PROOF OF PROPOSITION 9

Proof.

Step 1 First, we compare the equilibrium payoffs between the 'Collusion-proof' (CP) regime $[X^{B}_{CP}(\theta)\text{w.p1}-p, X^{FB}(\theta)\text{w.p}p]$ and the 'Pooling' (PL) regime $[X^{P}(\theta)\text{w.p1}]$.

By definition, $X^{B}_{CP}(\theta)$ is the optimal output rule over the problem (CP)

$$\max_{X(\cdot)}(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X(\theta)-C(X(\theta),\theta)]f(\theta)d\theta$$
$$+[(1-p)+p(k+\beta)]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X(\theta),\theta)}{\partial\theta}f(\theta)d\theta.$$

By definition, $X^P(\theta)$ is the optimal output rule over the problem (PL)

$$\max_{X(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} \left[X(\theta) - C(X(\theta),\theta) + [(1-p) + p(k+\beta)] \frac{\partial C(X(\theta),\theta)}{\partial \theta} \frac{1}{h(\theta)} \right] f(\theta) d\theta.$$

Hence, from the revealed preference argument, the following holds.

$$\begin{aligned} &(1-p) \int_{\underline{\theta}}^{\overline{\theta}} [X^B_{CP}(\theta) - C(X^B_{CP}(\theta),\theta)] f(\theta) d\theta \\ &+[(1-p) + p(k+\beta)] \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)} \frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial \theta} f(\theta) d\theta \\ \geq &(1-p) \int_{\underline{\theta}}^{\overline{\theta}} [X^P(\theta) - C(X^P(\theta),\theta)] f(\theta) d\theta \\ &+[(1-p) + p(k+\beta)] \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)} \frac{\partial C(X^P(\theta),\theta)}{\partial \theta} f(\theta) d\theta. \end{aligned} \tag{A.15}$$

The following inequality holds by the same revealed preference argument.

$$\underbrace{p \int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] f(\theta) d\theta}_{\text{First Best Expected Total Surplus}} \geq p \int_{\underline{\theta}}^{\overline{\theta}} [X^P(\theta) - C(X^P(\theta),\theta)] f(\theta) d\theta. \tag{A.16}$$

Hence, we have the following inequality.

$$\begin{aligned} &(1-p) \int_{\underline{\theta}}^{\overline{\theta}} [X^B_{CP}(\theta) - C(X^B_{CP}(\theta),\theta)] f(\theta) d\theta \\ &+\underbrace{p \int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] f(\theta) d\theta}_{\text{First Best Expected Total Surplus}} \\ &+[(1-p) + p(k+\beta)] \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)} \frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial \theta} f(\theta) d\theta \\ \geq &(1-p) \int_{\underline{\theta}}^{\overline{\theta}} [X^P(\theta) - C(X^P(\theta),\theta)] f(\theta) d\theta \\ &+p \int_{\underline{\theta}}^{\overline{\theta}} [X^P(\theta) - C(X^P(\theta),\theta)] f(\theta) d\theta \end{aligned} \tag{A.17}$$

$$+[(1-p)+p(k+\beta)]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^P(\theta),\theta)}{\partial\theta}f(\theta)d\theta$$
$$=\int_{\underline{\theta}}^{\overline{\theta}}[X^P(\theta)-C(X^P(\theta),\theta)]f(\theta)d\theta$$
$$+[(1-p)+p(k+\beta)]\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X^P(\theta),\theta)}{\partial\theta}f(\theta)d\theta.$$

Thus, the 'Collusion-proof' regime (CP) $[X^B_{CP}(\theta)\text{w.p}1-p, X^{FB}(\theta)\text{w.p}p]$ is payoff dominant over the 'Pooling' (PL) regime $[X^P(\theta)\text{w.p}1]$ for the principal.

Step 2 Next, we compare the equilibrium payoffs between the 'Pooling' regime (PL) $X^P(\theta)$ and Equilibrium Collusion Regime (EC) $X_{EC}(\theta)$.

By definition, $X_{EC}(\theta)$ is the optimal output rule over the problem

$$\max_{X(\cdot)}\int_{\underline{\theta}}^{\overline{\theta}}\left[X(\theta)-C(X(\theta),\theta)+\frac{\partial C(X(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}\right]f(\theta)d\theta.$$

Then, from the revealed preference argument, the following holds.

$$\int_{\underline{\theta}}^{\overline{\theta}}\left[X^P(\theta)-C(X^P(\theta),\theta)+[(1-p)+p(k+\beta)]\frac{\partial C(X^P(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}\right]f(\theta)d\theta$$
$$\begin{matrix}\geq\\ \leq\end{matrix}\int_{\underline{\theta}}^{\overline{\theta}}\left[X_{EC}(\theta)-C(X_{EC}(\theta),\theta)+\frac{\partial C(X_{EC}(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}\right]f(\theta)d\theta\text{:Eq. Collusion Payoff}$$

We have the following result on the marginal incentives (outputs), by comparing the coefficients of the information rents between Two Regimes.

$$X_{EC}(\theta)\leq X^P(\theta)\text{ if }(1-p)+p(k+\beta)\leq 1\iff\beta\leq 1-k \qquad \text{(A.18)}$$
$$X_{EC}(\theta)\geq X^P(\theta)\text{ if }(1-p)+p(k+\beta)\geq 1\iff\beta\geq 1-k \qquad \text{(A.19)}$$

1. When $X_{EC}(\theta) = X(\theta) \leq X^P(\theta)$ if $(1-p)+p(k+\beta) \leq 1 \iff \beta \leq 1-k$
Combining the results of the above two steps, we obtain

$$
\begin{aligned}
&(1-p)\int_{\underline{\theta}}^{\overline{\theta}} [X^B_{CP}(\theta) - C(X^B_{CP}(\theta),\theta)] f(\theta) d\theta \\
&+p \underbrace{\int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta),\theta)] f(\theta) d\theta}_{\text{First Best Expected Total Surplus}} \\
&+[(1-p)+p(k+\beta)] \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{h(\theta)} \frac{\partial C(X^B_{CP}(\theta),\theta)}{\partial \theta} f(\theta) d\theta \\
&\geq \int_{\underline{\theta}}^{\overline{\theta}} \left[X^P(\theta) - C(X^P(\theta),\theta) + [(1-p)+p(k+\beta)] \frac{\partial C(X^P(\theta),\theta)}{\partial \theta} \frac{1}{h(\theta)} \right] f(\theta) d\theta \\
&\geq \int_{\underline{\theta}}^{\overline{\theta}} \left[X_{EC}(\theta) - C(X_{EC}(\theta),\theta) + \frac{1}{h(\theta)} \frac{\partial C(X_{EC}(\theta),\theta)}{\partial \theta} \right] f(\theta) d\theta .\text{Eq. Collusion Payoff}
\end{aligned}
$$

The principal prefers the Collusion-proof regime (CP) to the Equilibrium Collusion regimes (EC) in terms of his expected payoff when the shading parameter $\beta \leq 1-k$, which is a sufficient condition for the Collusion-proof regime (CP) to be optimal. In this case, the "collusion-proof principle" still holds.

2. When $X_{EC}(\theta) \geq X^P(\theta)$ if $(1-p)+p(k+\beta) \geq 1 \iff \beta \geq 1-k$
The optimal solution $X^B_{CP}(\theta)$ is determined by

$$
\frac{\partial J^B_{CP}(X,\theta)}{\partial X} = \underbrace{[1-p]\left[1 - \frac{\partial C(X,\theta)}{\partial X}\right] + \frac{[(1-p)+pk]}{h(\theta)} \frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}}_{\text{Marginal Vortual Surplus}} + \underbrace{\frac{p\beta}{h(\theta)} \underbrace{\frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}}_{-}}_{\text{Marginal Shading Cost}} = 0.
$$

That is,

$$
\underbrace{\left[1 - \frac{\partial C(X,\theta)}{\partial X}\right] + \frac{\left[1 + \frac{pk}{1-p}\right]}{h(\theta)} \frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}}_{\text{Marginal Virtual Surplus}} + \underbrace{\frac{p\beta}{1-p} \frac{1}{h(\theta)} \underbrace{\frac{\partial^2 C(X,\theta)}{\partial X \partial \theta}}_{-}}_{\text{Marginal Shading Cost}} = 0. \tag{A.20}
$$

Then, as the shading parameter β becomes larger (as $\beta \to +\infty$), the optimal output rule goes to zero, $X^B_{CP}(\theta) \to 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. Then, the potential aggrievement (information rent) for the agent also goes to zero, $U(\theta) \to 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. Hence, the equilibrium payoff of the

Collusion-proof regime with Behavioral elements goes to

$$\begin{aligned}
&\underbrace{(1-p)\int_{\underline{\theta}}^{\overline{\theta}}[X_{CP}^{B}(\theta)-C(X_{CP}^{B}(\theta),\theta)]f(\theta)d\theta}_{\to 0}\\
&+p\underbrace{\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta}_{\text{First Best Expected Total Surplus}}\\
&+[(1-p)+p(k+\beta)]\underbrace{\int_{\underline{\theta}}^{\overline{\theta}}\frac{1}{h(\theta)}\frac{\partial C(X_{CP}^{B}(\theta),\theta)}{\partial\theta}f(\theta)d\theta}_{\to 0}\\
&\to p\underbrace{\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta}_{\text{First Best Expected Total Surplus}}. \qquad \text{(A.21)}
\end{aligned}$$

On the other hand, the payoff of the Equilibrium Collusion regime is independent of β, p

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}}[X_{EC}(\theta)-C(X_{EC}(\theta),\theta)-U(\theta)]d\theta\\
&=\int_{\underline{\theta}}^{\overline{\theta}}\left[X_{EC}(\theta)-C(X_{EC}(\theta),\theta)+\frac{\partial C(X_{EC}(\theta),\theta)}{\partial\theta}\frac{1}{h(\theta)}\right]f(\theta)d\theta \qquad \text{(A.22)}
\end{aligned}$$

Hence, which payoff is greater between (CP) and (EC) at $\beta\to+\infty$ depends on the relative size of

$$\begin{aligned}
&\underbrace{p\int_{\underline{\theta}}^{\overline{\theta}}[X^{FB}(\theta)-C(X^{FB}(\theta),\theta)]f(\theta)d\theta}_{\text{First Best Expected Total Surplus}}\\
&\gtreqless\underbrace{\int_{\underline{\theta}}^{\overline{\theta}}[X_{EC}(\theta)-C(X_{EC}(\theta),\theta)-U(\theta)]f(\theta)d\theta}_{\text{Eq. Collusion Payoff}}
\end{aligned}$$

Case 1

If

$$p < \frac{\underbrace{\int_{\underline{\theta}}^{\overline{\theta}} [X_{EC}(\theta) - C(X_{EC}(\theta), \theta) - U(\theta)] f(\theta) d\theta}_{\text{Eq. Collusion Payoff=Second Best Surplus}}}{\underbrace{\int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] f(\theta) d\theta}_{\text{First Best Expected Total Surplus}}} = p^* \tag{A.23}$$

There exists a cutoff value of shading strength β^* such that for $\beta \geq \beta^* (\geq 1-k)$ "Equilibrium Collusion" Payoff dominates "Collusion-proof" payoff, that is, **Equilibrium Collusion** is optimally chosen by the principal.

Case 2

If

$$p \geq \frac{\underbrace{\int_{\underline{\theta}}^{\overline{\theta}} [X_{EC}(\theta) - C(X_{EC}(\theta), \theta) - U(\theta)] f(\theta) d\theta}_{\text{Eq. Collusion Payoff=Second Best Surplus}}}{\underbrace{\int_{\underline{\theta}}^{\overline{\theta}} [X^{FB}(\theta) - C(X^{FB}(\theta), \theta)] f(\theta) d\theta}_{\text{First Best Expected Total Surplus}}} = p^* \tag{A.24}$$

Equilibrium Collusion is not optimal even for $\beta \to +\infty$, but **Collusion-proof** regime is optimally chosen. (A clear example is $p \to 1$). The point is that **Shut-down** is endogenously chosen in the states of (θ, ϕ), that is, the optimal output rule goes to zero, $X^B_{CP}(\theta) \to 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, and the potential aggrievement (information rent) also goes to zero, $U(\theta) \to 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$.

As p becomes smaller, the states of (θ, ϕ) with probability $1-p$ increase. Then the principal cannot neglect his decision $X^B_{CP}(\theta)$ any more in the supervisory no information state ϕ, in the form of $X^B_{CP}(\theta) \to 0$. However, the cost of collusion-proof constraint, or the shading cost which the principal will eventually bear becomes very large. Since it is too costly, the principal optimally switches to the Equilibrium Collusion Regime which induces $X_{EC}(\theta)$ in both states. ∎

REFERENCES

Amir, Rabah, 2005. Supermodularity and Complementarity in Economics: An Elementary Survey. *Southern Economic Journal* **73, 3**, 636-660.

Bolton, Patrick and Mathias Dewatripont, 2005. Contract Theory. Cambridge, MA: MIT Press.

Coase, Ronald, 1937. The Nature of the Firm. *Economica* **4, 16**, 386-405.

Edlin, Aaron and Chris Shannon, 1998. Strict Monotonicity in Comparative Statics. *Journal of Economic Theory* **81**, 201-219.

Falk, Armin and Urs Fischbacher, 2006. A Theory of Reciprocity. *Games and Economic Behavior* **54, 2**, 293-315.

Fehr, Ernst and Klaus Schmidt, 1999. A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics* **114, 3**, 817-868.

Fehr, Ernst, Oliver Hart, and Christian Zehnder, 2015. How Do Informal Agreements and Renegotiation Shape Contractual Reference Points? *Journal of the European Economic Association* **13, 1**, 1-28.

Fudenberg, Drew and Jean Tirole, 1991. Game Theory. Cambridge, MA: MIT Press.

Hart, Oliver and John Moore, 2008. Contracts as Reference Points. *Quarterly Journal of Economics* **123, 1**, 1-48.

Itoh, Hideshi, 1993. Coalitions, Incentives, and Risk Sharing. *Journal of Economic Theory* **60**, 410-27.

Kofman, Fred and Jacques Lawarree, 1993. Collusion in Hierarchical Agency. *Econometrica* **61, 3**, 629-656.

Kofman, Fred and Jacques Lawarree, 1996. On the Optimality of Allowing Collusion. *Journal of Public Economics* **61, 3**, 383-407.

Laffont, Jean-Jacques and David Martimort, 1997. Collusion under Asymmetric Information. *Econometrica* **65, 4**, 875-911.

Laffont, Jean-Jacques and Jean Tirole, 1991. The Politics of Government Decision-Making: A Theory of Regulatory Capture. *Quarterly Journal of Economics* **106**, 1089-1127.

Laffont, Jean-Jacques and Jean Tirole, 1993. A Theory of Incentives in Procurement and Regulation. Cambridge, MA: MIT Press.

Milgrom, Paul, 1988. Employment Contracts, Influence Activities and Efficient Organization Design. *Journal of Political Economy* **96**, 42-60.

Milgrom, Paul, 2004. Putting Auction Theory to Work. Cambridge University Press: Cambridge.

Mirrlees, James, 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies* **38, 2**, 175-208.

Myerson, Roger, 1981. Optimal auction design. *Mathematics of Operations Research* **6**, 58-73.

Suzuki, Yutaka, 2007. Collusion in Organizations and Management of Conflicts through Job Design and Authority Delegation. *Journal of Economic Research* **12**, 203-241.

Suzuki, Yutaka, 2008. Mechanism Design with Collusive Supervision: A Three-tier Agency Model with a Continuum of Types. *Economics Bulletin* **4, 12**, 1-10.

Suzuki, Yutaka, 2018. Hierarchical Global Pollution Control in Asymmetric Information Environments: A Continuous-type, Three-tier Agency Framework. *Journal of Economic Research* **23**, 1-37.

Suzuki, Yutaka, 2019. A Contract Theory Analysis to Fiscal Relations between the Central and Local Governments in China. *Economic and Political Studies* **7, 3**, 281-313.

Tirole, Jean, 1986. Hierarchies and Bureaucracies: On the role of Collusion in Organizations. *Journal of Law, Economics and Organization* **2**, 181-214.

Tirole, Jean, 1992. Collusion and the Theory of Organizations. In Advances in Economic Theory: The Sixth World Congress. Edited by J-J. Laffont. Cambridge: Cambridge University Press.

Topkis, Donald, 1978. Minimizing a submodular function on a lattice. *Operations Research* **26, 2**, 255-321.

Weitzman, Martin, 1974. Prices vs. quantities. *Review of Economic Studies* **41, 4**, 477-491.

Williamson, Oliver, 1971. The Vertical Integration of Production: Market Failure Considerations. *American Economic Review* **61**, 112-23.

Williamson, Oliver, 1975. Markets and Hierarchies: Analysis and Antitrust Implications. New York, NY: Free Press.