Collusive Supervision, Shading, and Efficient Organization Design◆

Yutaka Suzuki

Hosei University

This Version, September 2012

Preliminary Draft

Abstract

We introduce the recent behavioral contract theory idea, “shading” (Hart and Moore (2007, 2008)) as a component of ex-post haggling (addressed by Coase (1937) and Williamson (1975)) into the collusion model a la Tirole (1986, 1992), thereby constructing a new model of internal hierarchical organization. By combining these two ideas, i.e., *collusion* and *shading*, we can not only enrich the existing collusion model, thereby obtaining a new result on *Collusion-proof* vs. *Equilibrium Collusion*, but also give a micro foundation to ex-post adaptation costs, where we view rent-seeking associated with collusive behavior and ex-post haggling generated from shading as the two sources of adaptation costs. By using this model, we examine the optimal organizational design problem as an optimal response to the trade-off between gross total surplus and ex-post haggling costs. This could help provide a deep understanding of resource allocation and decision process in the internal organization of large firms.

Key Words**:** Collusion, Supervision, Shading, Haggling Cost, Efficient Organization Design

JEL Classification: D23, D82, D86

◆I am grateful to Oliver Hart for his insightful suggestions and advices, and to Harvard University for its stimulating environment and hospitality during my visiting scholarship in 2011-2012. Of course, all opinions and errors are my own.

1. Introduction

In this paper, we introduce the recent behavioral contract theory idea, “shading” (Hart and Moore (2007, 2008)) as a component of ex-post haggling (addressed by Coase (1937) and Williamson (1975)) into the collusion model a la Tirole (1986, 1992), thereby constructing a new model of internal hierarchical organization. By combining these 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 adaptation costs, where we view rent-seeking associated with collusive behavior and ex-post haggling generated from shading as the two sources of adaptation costs. By using this model, we examine the optimal organizational design problem as an optimal response to the trade-off between gross total surplus and ex-post haggling costs. This could lead to a deeper understanding of resource allocation and decision process in the internal organization of large firms.

Transaction Cost Economics (TCE) starting from Coase (1937) and Williamson (1975) has so far emphasized that ex-post haggling and maladaptation drives inefficiencies in large firms. Gibbons (2010) explains that Williamson (2000) emphasizes that maladaptation in the contract execution interval is the principal source of inefficiency[1]. Especially, Gibbons (2005) identifies an “adaptation” theory of the Firm in Williamson (1971), in which hierarchy serves to facilitate “adaptive, sequential decision-making”, and notes that a key theoretical challenge in developing such a theory is to define an environment in which neither ex-ante contracts nor ex-post renegotiation can induce first-best adaptation after uncertainty is resolved.[2] Hart and Moore (2007, 2008) include a new behavioral idea, “shading behaviors” into the incomplete contracting framework and begin the first steps toward building a theory of ex-post inefficiency. Hart and Holmstrom (2010) analyze how the determination of firm scope (Integration vs. Non-integration) is affected by the shading costs in the incomplete contractual setting.

Milgrom (1988) and Milgrom and Roberts (1987) argued that the existence of a principal with discretionary authority can give rise to *influence costs*. They define *influence costs* as “the losses that are suffered when individuals seek to influence the organization’s decision in order to advance their private interests and when the organization adapts to control this behavior”. Since influence activities can be understood as costly activities aimed at persuading a decision maker, we could say that influence activities in their model correspond to haggling a la Coase (1937) and Williamson (1975).

[1] This is in sharp contrast to the Grossman-Hart-Moore Property Right approach’s emphasis on the ex-ante inefficiency based on specific investments (Grossman and Hart (1986), Hart and Moore (1990)).

[2] Thus, as Simon (1951) pointed out, the second-best solution may be to concentrate authority in the hands of a “boss” who then takes (potentially self-interested) decisions after uncertainty is resolved.

Now, an independent literature exists which in a closely related way deals with the issues associated with *collusion* in organizations by using a three-tier agency model, which was first developed by Tirole (1986, 1992) and then exhaustively examined by Laffont and Tirole (1993), Laffont and Martimort (1997,1998) and others. In hierarchical organizations where a supervisor(s) monitors agents for the benefit of the principal, manipulation of information may arise when agents and supervisor(s) collude to conceal the relevant information from the principal. The collusion literature addresses this problem within the framework of triangular or multilateral agency relationships, where participants may contemplate side contracting. Collusion means that within a group of participants, a coalition forms a strategic alliance at the expense of the rest of the group.

This line of research has addressed the possibility of supervisor-agent coalition formation within a three-tier hierarchy, where the principal may wish to monitor an agent and so hires a supervisor to perform the task effectively. However, the supervisor may often be purely self-interested, and willing to accept a payment (bribe) from the agent in return for hiding his observations. The manipulation of information through the collusion between the supervisor and the agent may bring about a large loss for the organization, since inefficient resource allocation may be realized. Hence, the principal may exercise the option to create *collusion-proof contracts* to deter the supervisor's misbehavior. This is a familiar result in the collusion literature following the model of Tirole (1986).

Suzuki (2007) considered the principal-supervisor-*two agents* hierarchy *with supervisory efforts*, and showed that in some cases, the collusion-proof contracts may be the second best solution, but in the other cases, allowing the possibility of vertical collusion and promoting lateral collusion among a subgroup of actors in equilibrium may be welfare enhancing. He investigated the conditions under which each solution is selected as the second best solution, characterized the nature of the incentive schemes, and related the optimal solutions to the problem of authority delegation in organizations, especially from the viewpoint of formal and real authority introduced by Aghion and Tirole (1997).

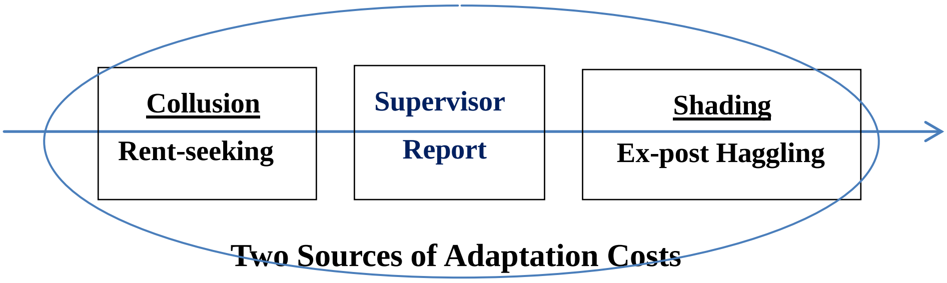
Though Suzuki (2007)'s framework is based on Tirole (1986, 1992) and Laffont and Tirole (1991), the results and the underlying intuition are close to Milgrom (1988) referred to above, which states that efficient organization design counters *influence activities* by limiting the discretion of decision makers, especially for those decisions that have large distributional consequences, but that are otherwise of little consequence to the organization. The main departure from Milgrom (1988) is that Suzuki (2007) includes a more explicit model of the *collusion* game played by the supervisor and agents, while Milgrom (1988) deals with more general but less modeled *influence activities*. Modeling the collusion game explicitly gives us distinct predictions as to *collusion proof* vs. *equilibrium collusion*, decentralization and delegation, and how various forms of collusion lead to inefficiency in the three-tier, contracting problem.

Now, it is important to note that *collusion* in the collusion literature always requires an activity or a behavior for coalition formation e.g., between supervisor and agent. It can be viewed as “rent-seeking behavior” or “influence behavior”, because it is a costly activity aimed at persuading a supervisor, a decision maker in the intermediary position of the organization, in order to advance his private payoffs. In summary, *collusion* and *influence activities* are very closely related, and at the same time both of them are behaviors made *before* an important decision making.

In reality, controversies often occur both *before* and *after* the decision making. For example, in the Faculty Council at universities, we often see partisan formation activities (coalition formation) and enthusiastic persuasion before some decision-making as well as harsh controversy and criticism around execution and enforcement afterwards. Of course, this will be the case with decision-making in the political world. Hence, it would be very sensible to include such ex-ante and ex-post controversies, which may bring about a great deal of inefficiency in large firms.

As a step to generate a theory of ex-post inefficiency as emphasized in Transaction Cost Economics (TCE), 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 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 the contract, he engages in shading. This will lead to ex post controversy and mutual punishment. Note that these shading behaviors are modeled to be made *after* an important decision making.

We introduce this behavioral idea, “shading behavior” (Hart and Moore (2007, 2008)) as a component of ex-post haggling (addressed by Coase (1937) and Williamson (1975, 1985)) into the collusion model a la Tirole (1986, 1992), thereby constructing a new model of internal hierarchical organization. By combining these two ideas, i.e., *collusion* and *shading*, we can not only enrich the existing collusion model, thereby obtaining a new result on the choice of Collusion-proof vs. Equilibrium Collusion regimes, but also 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 shading as the two sources of adaptation costs, as in the figure below. By using this model, we can examine the optimal organizational design problem as an optimal response to the trade-off between gross total surplus and ex-post adaptation costs. This could help provide a deep understanding of resource allocation and decision process in the internal organization of large firms.



Our paper is constructed as follows. In section 2, we present our model: the parties, the timing, the behavioral element, and characterize the first best solution and then the second best, collusion-free solution of the basic three-tier hierarchy a la Tirole (1986,1992). In section 3, we solve the model with collusion but with no behavioral element, and derive the optimal collusion-proof solution. In section 4, we introduce a behavioral element: shading a la Hart and Moore (2007, 2008) into our three-tier model. We obtain a new result on *Collusion-proof* vs. *Equilibrium Collusion* and also examine the optimal organizational design problem as a response to the trade-off between gross total surplus and ex-post adaptation costs. Section 5 concludes this 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. When born, the agent is endowed with a productive parameter θ, $\theta \in \left\{\underline{\theta}, \overline{\theta}\right\}, 0 < \underline{\theta} < \overline{\theta}$, which is private information. He decides how much effort to exert. The effort e is unobservable to third parties. Expending effort e costs the agent $C\left(e\right)$ in disutility, which satisfies $C\left(e\right) > 0, C'\left(e\right) > 0, C'' > 0, \forall e \in \mathbb{R}_+$. For a given productivity level θ and the effort e of the agent, the output is generated as $X = \theta + e$. W is the wage payment the agent receives, and then his utility is described as $W - C\left(e\right) = W - C\left(X - \theta\right)$.We normalize the agent's reservation utility as 0.

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 can, for each θ, provide proof of the fact (θ) with probability p, and with $1-p$, is unable to obtain any information.[3] 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** she makes to the principal on the agent's productivity, and she 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 $h \equiv \Pr\left\{\theta = \overline{\theta}\right\}$. 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 $W\left(X,r\right)$ for the agent and $W_S\left(X,r\right)$ 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 t**he collusion game the agent has all the bargaining power and makes a take-it-or-leave-it offer** to the supervisor. The supervisor can only accept or reject the offer. Specifically, let us assume the following collusion technology: if the agent offers the supervisor a transfer (side payment) t, she benefits up to kt, where $k \in \left[0,1\right]$. The idea is that transfers of this sort, being prevented by the principal, may be hard to organize and are subject to resource losses, whose cost is $\left(1-k\right)t$. We follow the literature in assuming that side-contracts of this sort are possible (see e.g. Tirole 1986,

[3] The supervisor's signal s received from the agent may be informative $s=\theta$ with probability p, or non-informative $s=\phi$ with probability $1-p$.

1992).[4]

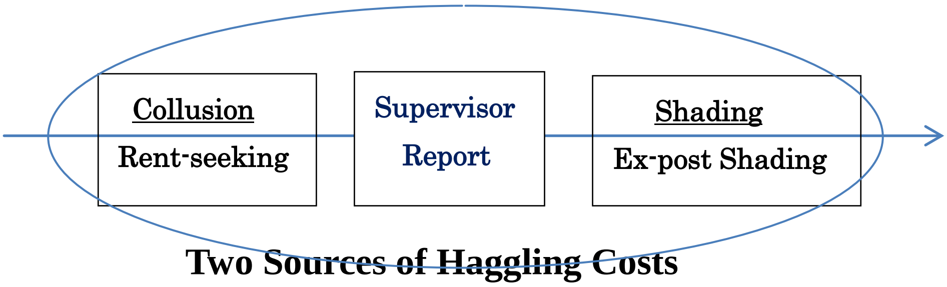
The supervisor then produces a report for the principal. This report is public information.

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. Introduction of Behavioral Element: Shading

We incorporate a behavioral element into the model, based on the "shading" model[5] 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. 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, we try to **give a micro foundation to ex-post** adaptation costs,[6] and understand the ex-post optimal adaptation[7] as an optimal balance resulting from the trade-off between gross total surplus and ex-post adaptation costs associated with the output decisions. We explicitly analyze this in section 4.



Two Sources of Haggling Costs

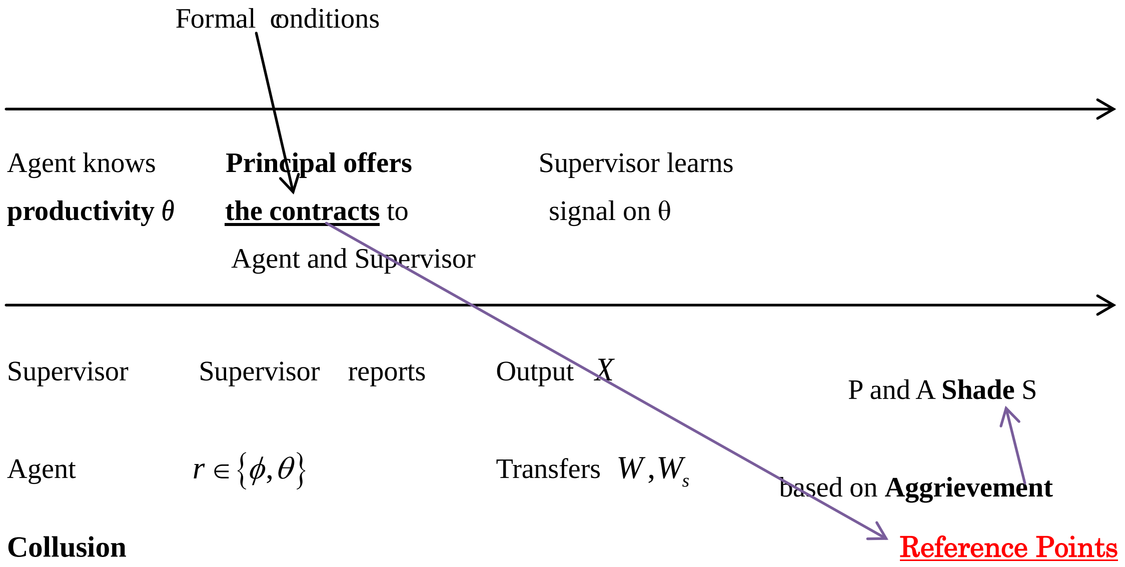
[4] Enforceability of side contracts should have some more theoretical or behavioral foundation. In section 4.2, we introduce a new idea where the behavioral element (Shading) becomes a strong driver that implements Equilibrium Collusion (Side Contract) between the supervisor and the agent.

[5] This 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".

[6] We introduce rent-seeking associated with collusive behavior and ex-post haggling generated from shading as the two sources of adaptation costs.

[7] See Coase (1937), Williamson (1985), and Gibbons (2010).

2.4 Time Line of the Model


Formal conditions

Agent knows **productivity** θ

Principal offers **the contracts** to Agent and Supervisor

Supervisor learns signal on θ

Supervisor Agent **Collusion**

Supervisor reports $r \in \{\phi, \theta\}$

Output X Transfers W, W_s

P and A **Shade** S based on **Aggrievement**

Reference Points

2.5 Symmetric Information (First Best) Solution

Now, let $X\left(\overline{\theta}\right)$ and $X\left(\underline{\theta}\right)$ be the outputs specified for the good-type (high-productivity) agent ($\theta = \overline{\theta}$) and the bad-type (low-productivity) agent ($\theta = \underline{\theta}$), respectively. We write X_H and X_L for $X\left(\overline{\theta}\right)$ and $X\left(\underline{\theta}\right)$, respectively. Defining $W\left(\overline{\theta}\right)$ and $W\left(\underline{\theta}\right)$ similarly, we write W_H and W_L for $W\left(\overline{\theta}\right)$ and $W\left(\underline{\theta}\right)$, respectively. These are the wages specified by the contracts.

The first best solution under symmetric information maximizes the expected profits, subject to the IR (Individual Rationality) constraints, which require that the manager be willing to sign a contract whatever her type. The supervisor has no supervisory role, and so receives the reservation wage 0 for all states of nature. Thus, the problem is:

$$\max_{\{X_H, W_H\}, \{X_L, W_L\}} h\left[X_H - W_H\right] + (1-h)\left[X_L - W_L\right]$$
$$\text{s.t.} \quad W_H - C\left(X_H - \overline{\theta}\right) \geq 0$$
$$W_L - C\left(X_L - \underline{\theta}\right) \geq 0$$

Substituting $W_H = C\left(X_H - \overline{\theta}\right)$ and $W_L = C\left(X_L - \underline{\theta}\right)$ into the objective function results in the expected total surplus maximization:

$$\max_{\{X_H, X_L\}} h\left[X_H - C\left(X_H - \overline{\theta}\right)\right] + (1-h)\left[X_L - C\left(X_L - \underline{\theta}\right)\right]$$

The first order conditions for the optimum are:

$$\left.\begin{array}{l} 1 - \dfrac{\partial C\left(X_H - \overline{\theta}\right)}{\partial X_H} = 0 \\ 1 - \dfrac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} = 0 \end{array}\right\} \Leftrightarrow 1 = \frac{\partial C\left(X_H - \overline{\theta}\right)}{\partial X_H} = \frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L}$$

In the first best optimum, the marginal benefit of output 1 is equal to the marginal cost of output for both types $\overline{\theta}, \underline{\theta}$. Hence, we have $X_H - \overline{\theta} = X_L - \underline{\theta}$ and $e_H^{FB} = e_L^{FB} = e^{FB}$. This means that first best efforts are equal for both types $\overline{\theta}, \underline{\theta}$. We also see that $W_H^{FB} = W_L^{FB} = C\left(e^{FB}\right)$

2.6 Asymmetric Information Solution with No Collusion

Next, under the assumption of **asymmetric information** on θ, we seek the separating contracts, which induce the two types $\overline{\theta}, \underline{\theta}$ to behave differently. For this, the contracts must be incentive compatible.

IC (Incentive Compatibility) requires:

$$W_H - C\left(X_H - \overline{\theta}\right) \geq W_L - C\left(X_L - \overline{\theta}\right) \quad (1a)$$

$$W_L - C\left(X_L - \underline{\theta}\right) \geq W_H - C\left(X_H - \underline{\theta}\right) \quad (1b)$$

(1a) states that the good-type (high-productivity) agent $\left(\theta = \overline{\theta}\right)$ prefers to select the contract intended for him rather than the contract intended for the bad-type (low-productivity) agent $\left(\theta = \underline{\theta}\right)$, i.e., the good-type agent's IC constraint. (1b) states that the bad-type agent $\left(\theta = \underline{\theta}\right)$ prefers to select the contract intended for him rather than the contract intended for the good-type agent $\left(\theta = \overline{\theta}\right)$, i.e., the bad-type agent's IC constraint.

The IR (Individual Rationality) constraints require:

$$W_H - C\left(X_H - \overline{\theta}\right) \geq 0 \quad (2a)$$

$$W_L - C\left(X_L - \underline{\theta}\right) \geq 0 \tag{2b}$$

The first best solutions $\left\{X_H^{FB}, X_L^{FB}\right\} = \left\{\overline{\theta} + e^{FB}, \underline{\theta} + e^{FB}\right\}, W_H^{FB} = W_L^{FB} = C\left(e^{FB}\right)$ are *not* incentive compatible for the good-type agent $\overline{\theta}$, since he has an incentive to tell a lie (mimic/pretend that type $\theta = \underline{\theta}$). Indeed, we can check the incentive of the good type $\overline{\theta}$.

If he tells the truth " $\theta = \overline{\theta}$ ", he obtains $W_H^{FB} - C\left(X_H^{FB} - \overline{\theta}\right) = 0$.

If he says " $\theta = \underline{\theta}$ " (i.e., he lies), he obtains

$W_L^{FB} - C\left(X_L^{FB} - \overline{\theta}\right) = C\left(e^{FB}\right) - C\left(e^{FB} - \left(\overline{\theta} - \underline{\theta}\right)\right) > 0$.

Hence, he has an incentive to tell a lie (mimic/pretend), i.e., *not incentive compatible*.

As is typical in such problems, only the good type's IC (1a) and the bad type's IR (2b) bind at the optimum. From (2b), $W_L = C\left(X_L - \underline{\theta}\right)$. Substituting it into (1a) with equality, we have

$$W_H - C\left(X_H - \overline{\theta}\right) = W_L - C\left(X_L - \overline{\theta}\right) = C\left(X_L - \underline{\theta}\right) - C\left(X_L - \overline{\theta}\right) \tag{3}$$

This is the *information rent* for the good-type (high-productivity) agent $\overline{\theta}$. Hence, the optimization problem can be written as follows

$$\begin{aligned} &\max_{\{X_H, W_H\},\{X_L, W_L\}} h\left[X_H - W_H\right] + (1-h)\left[X_L - W_L\right] \\ &\text{s.t.} \quad W_H - C\left(X_H - \overline{\theta}\right) = C\left(X_L - \underline{\theta}\right) - C\left(X_L - \overline{\theta}\right) \\ &\qquad\; W_L = C\left(X_L - \underline{\theta}\right) \end{aligned}$$

Substituting $W_L = C\left(X_L - \underline{\theta}\right)$ and $W_H = C\left(X_H - \overline{\theta}\right) + \left[C\left(X_L - \underline{\theta}\right) - C\left(X_L - \overline{\theta}\right)\right]$ into the objective function yields

$$\max_{X_H, X_L} \underbrace{h\left[X_H - C\left(X_H - \overline{\theta}\right)\right] + (1-h)\left[X_L - C\left(X_L - \underline{\theta}\right)\right]}_{\text{Expected Total Surplus}} - \underbrace{h\left[C\left(X_L - \underline{\theta}\right) - C\left(X_L - \overline{\theta}\right)\right]}_{\substack{\text{"Information Rent"} \\ \text{for the good type}}}$$

The first order conditions for the optimum are:

$$1 - \frac{\partial C\left(X_H - \overline{\theta}\right)}{\partial X_H} = 0 \Leftrightarrow X_H^* = X_H^{FB}$$

$$\underbrace{(1-h)\left[1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}\right]}_{\substack{\text{Marginal Surplus}\\ \text{for the bad type}}}\underbrace{-h\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]}_{\substack{\text{Marginal Information Rent}\\ \text{for the good type}}}=0$$

$$\Leftrightarrow 1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{h}{1-h}\cdot\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]=0 \qquad (*)$$

From these conditions, we have the following proposition, which is a familiar result in the literature (e.g. Baron and Myerson (1982), Maskin and Riley (1984), and Bolton and Dewatripont (2005))

Proposition 1

In the principal-agent regime with no supervisor, the second-best solution has the properties of

(1)*Efficiency at the top* (for the good-type agent) $X_H^* = X_H^{FB}$

(2)*Downward distortion at the bottom* (for the bad-type agent) $X_L^* < X_L^{FB}$

Proof: As for X_H, the first order condition is the same as the first best case, so $X_H^* = X_H^{FB}$.

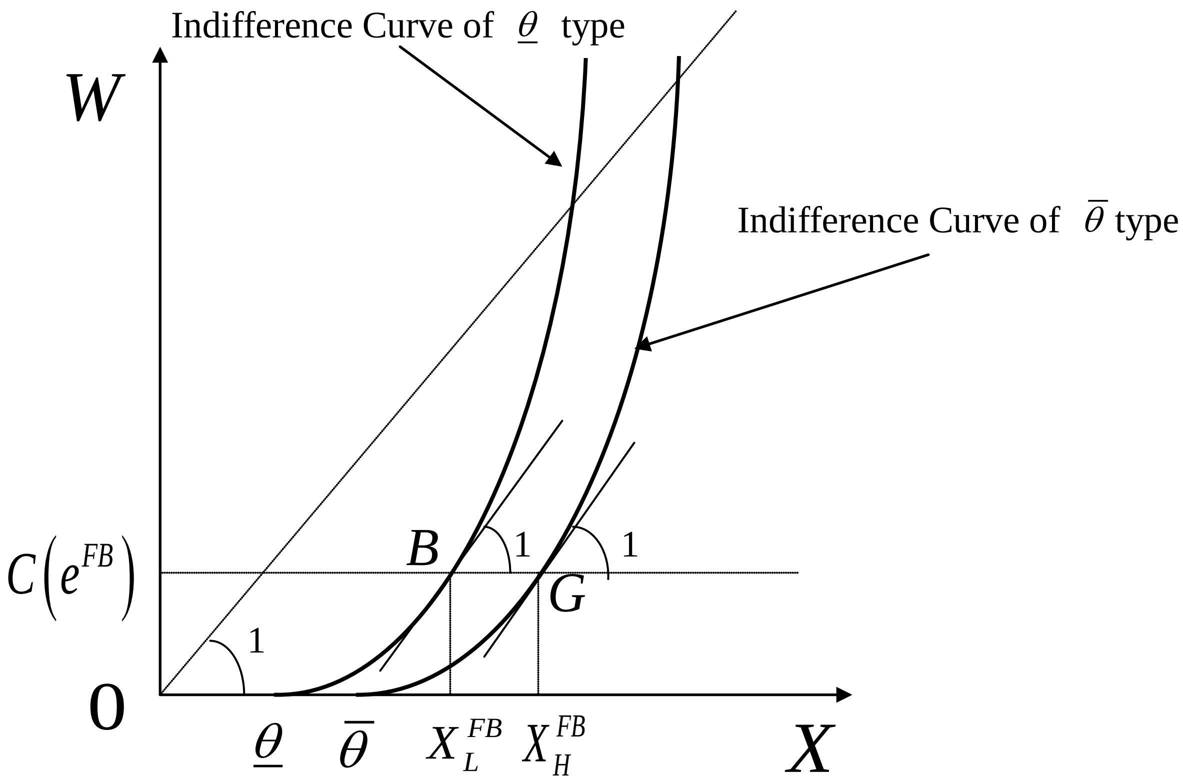
As for X_L, evaluating the first order condition at $X = X_L^{FB}$, we have

$-\frac{h}{1-h}\left[\frac{\partial C\left(X_L^{FB}-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L^{FB}-\bar{\theta}\right)}{\partial X_L}\right]<0$. This means that the principal can raise his virtual payoff by decreasing X_L from the first best level X_L^{FB}. Hence, we have $X_L^* < X_L^{FB}$.■

2.7 Graphical Explanation

Let us explain the argument so far in a graphical manner. First, the payoff function of the type θ agent is $U\left(W,X;\theta\right)=W-C\left(X-\theta\right)$. In order to depict the indifference curve of the type θ agent in the $\left(X,W\right)$ diagram, we totally differentiate both sides of $U_\theta=W-C\left(X-\theta\right)$, and obtain $dW-\frac{\partial C\left(X-\theta\right)}{\partial X}dX=0$. Then, putting it in order, we have the marginal rate of substitution $MRS_{XW}^{\theta} = \left.\frac{dW}{dX}\right|_{U=\text{const}} = \frac{\partial C\left(X-\theta\right)}{\partial X}$. We easily see that the marginal cost of output $\frac{\partial C\left(X-\theta\right)}{\partial X}$ is decreasing in type θ, i.e., the good type $\bar{\theta}$ has a gentler indifference curve (a smaller MRS_{XW}) for any point $\left(X,W\right)$.Remember that the first order conditions for the first best optimum are $1=\frac{\partial C\left(X_H-\bar{\theta}\right)}{\partial X_H}=\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}$. We have $X_H^{FB}-\bar{\theta}=X_L^{FB}-\underline{\theta}$, which means that $e_H^{FB}=e_L^{FB}=e^{FB}$. That is, at the first best solution, $1=C'\left(X_L^{FB}-\underline{\theta}\right)=C'\left(X_H^{FB}-\bar{\theta}\right)=C'\left(e^{FB}\right)$ and $W_H^{FB}=W_L^{FB}=C\left(e^{FB}\right)$. From these facts, we can depict the indifference curves of both types and the first best contracts G and B in the $\left(X,W\right)$ diagram.

Figure1



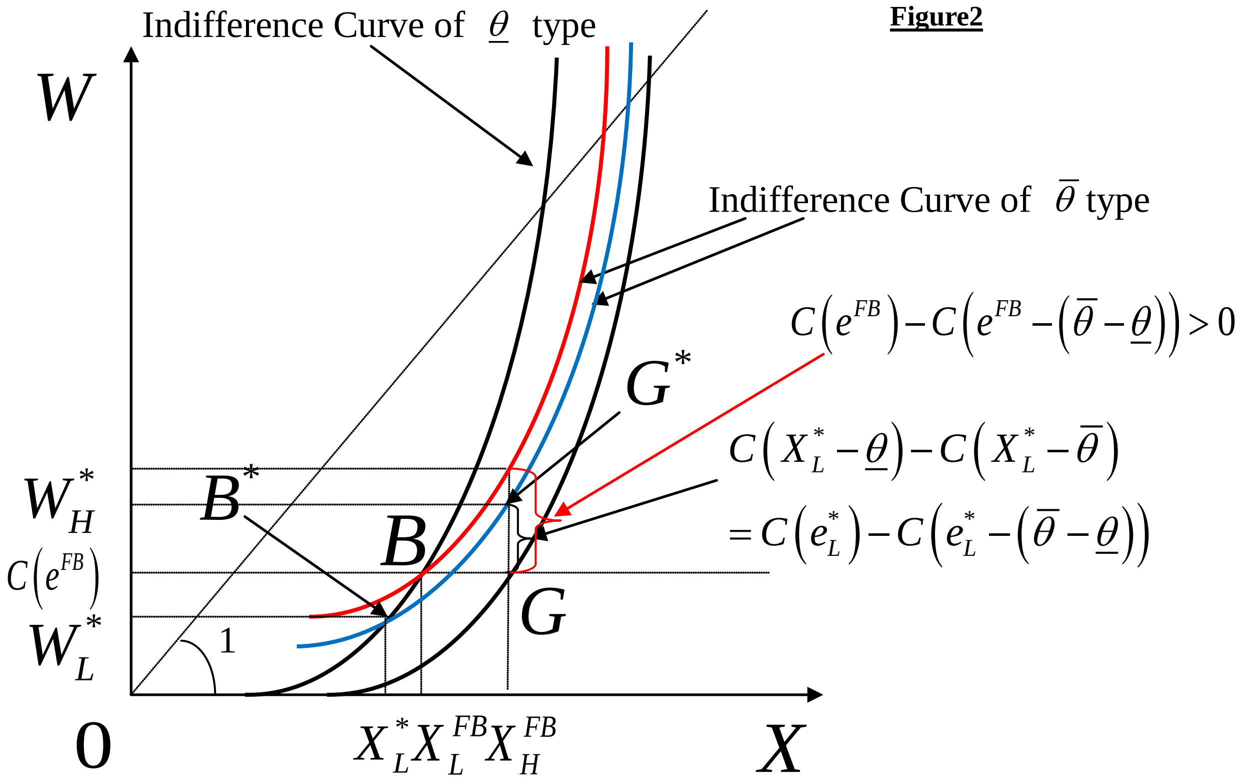
However, the first best solution $G:\left\{X_H^{FB},W_H^{FB}\right\}=\left\{\bar{\theta}+e^{FB},C\left(e^{FB}\right)\right\}$ is *not* incentive compatible for the good type $\bar{\theta}$ under asymmetric information, since he has an incentive to tell a lie (mimic type $\underline{\theta}$) and select $B:\left\{X_L^{FB},W_L^{FB}\right\}=\left\{\underline{\theta}+e^{FB},C\left(e^{FB}\right)\right\}$. Indeed, the good-type agent $\bar{\theta}$ can obtain $W_L^{FB}-C\left(X_L^{FB}-\bar{\theta}\right)=C\left(e^{FB}\right)-C\left(e^{FB}-\left(\bar{\theta}-\underline{\theta}\right)\right)>0$ by telling a lie, instead of $W_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)=0$ by telling the truth.

So, the principal takes the optimal balance between the expected total surplus and the information rent for the good type. As a result, we have the results of (1) *Efficiency at the top* (for the good-type agent $\bar{\theta}$) $X_H^*=X_H^{FB}$ and (2) *Downward distortion at the bottom* (for the bad-type agent $\underline{\theta}$) $X_L^*<X_L^{FB}$. The intuition is that a small reduction in X_L from the first best X_L^{FB} results in a *second-order (marginal) reduction* in total surplus for the bad type $\underline{\theta}$, but generates a *first-order (discrete) reduction* in the good type $\bar{\theta}$'s information rent through relaxing the IC for the good type $\bar{\theta}$ and allowing the principal to reduce W discretely. The optimal wage payments are

$W_L^*=C\left(X_L^*-\underline{\theta}\right)=C\left(e_L^*\right)$ for the bad type $\underline{\theta}$, and

$W_H^*=\underbrace{C\left(X_H^{FB}-\bar{\theta}\right)}_{\text{effort cost}}+\underbrace{C\left(X_L^*-\underline{\theta}\right)-C\left(X_L^*-\bar{\theta}\right)}_{\text{information rent}}$ for the good type $\bar{\theta}$.

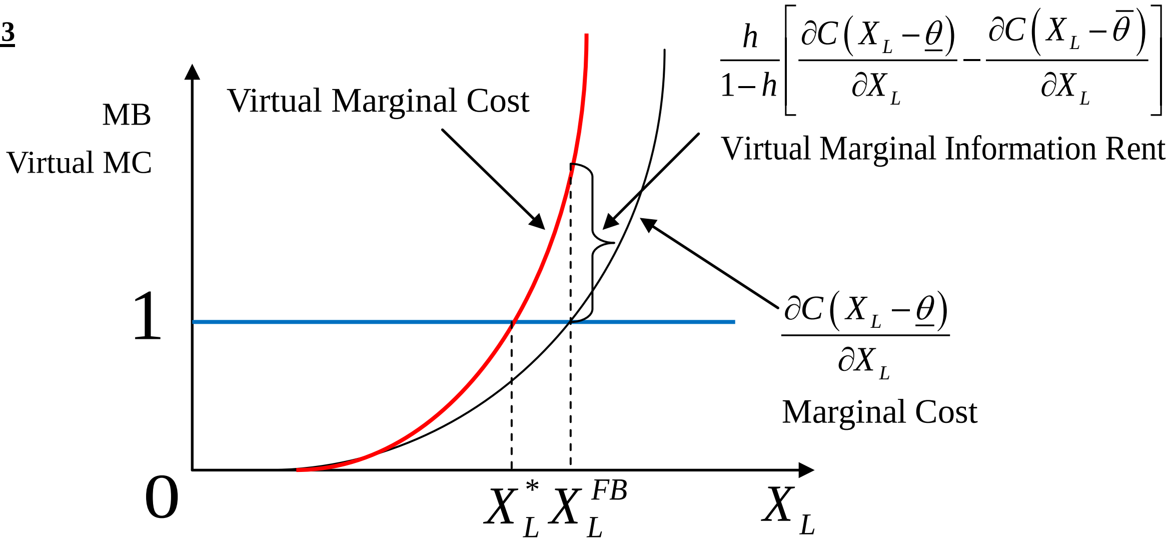
Figure2 shows the result.



The result of $X_L^* < X_L^{FB}$ can be understood by looking at Figure 3, which shows that the optimal solution X_L^* is determined such that the marginal benefit 1 equals the marginal *virtual cost* (the marginal cost $\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}$ plus the virtual marginal information rent

$$\frac{h}{1-h}\left[\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}-\frac{\partial C(X_L-\overline{\theta})}{\partial X_L}\right]).$$

Figure 3



$\frac{h}{1-h}\left[\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}-\frac{\partial C(X_L-\overline{\theta})}{\partial X_L}\right]$
MB
Virtual MC
Virtual Marginal Cost
Virtual Marginal Information Rent
1
$\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}$
Marginal Cost
0
X_L^*
X_L^{FB}
X_L

3. Optimal Collusion-proof Solution under Collusive Supervision and No Shading

Now, we introduce a third player, the supervisor, into the model. The principal has access, at a cost z, to the supervisor who is an internal auditor and can, for each θ, provide proof (evidence) of the fact (θ) with probability p, and with $1-p$, is unable to obtain any information.[8] 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. 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 observes the signal θ. A self-interested supervisor will collude with the agent only if he benefits from such behavior. Specifically, let us assume the following collusion technology: if the agent offers the supervisor a transfer (side payment) t, he benefits up to kt, where $k\in[0,1]$. The idea is that transfers of this sort, being prevented by the principal, may be hard to organize and are subject to resource losses, whose

[8] The supervisor's signal S received from the agent may be informative $s=\theta$ with probability p, or non-informative $s=\phi$ with probability $1-p$.

cost is $(1-k)t$. We follow the literature in assuming that side-contracts of this sort are possible (see, e.g., Tirole 1992).[9]

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 verifiable report from the supervisor that the type information is θ, the principal will have an incentive to *renegotiate* the original contract. The principal can raise her payoff by *eliminating the downward distortion in the bad type* $\underline{\theta}$. [10] Namely, instead of the contract at the no-information phase $\left\{X_L^{NC}, W_L^{NC}\right\}$, the principal will offer the efficient (first best) contract $\left\{X_L^{FB}, W_L^{FB}\right\}$ to the bad-type agent $\underline{\theta}$, and exploit the information rent $U(\overline{\theta}) = C\left(X_L^{NC} - \underline{\theta}\right) - C\left(X_L^{NC} - \overline{\theta}\right)$ from the good-type agent $\overline{\theta}$.[11] If the good-type agent anticipates this modification, since he can benefit from a failure by the supervisor to report his type $\overline{\theta}$ truthfully, he will have an incentive to offer the supervisor the side payment t up to $U(\overline{\theta})$, for which the supervisor benefits up to $kU(\overline{\theta})$, where $k \in [0,1]$. Thus, in order for the principal to induce the true information $\overline{\theta}$ from the supervisor, the following *coalition incentive compatibility constraint* (or truth telling constraint for the supervisor) must be satisfied.

$$W_s(\overline{\theta}) \geq kU(\overline{\theta}) = k\left[C\left(X_L^{NC} - \underline{\theta}\right) - C\left(X_L^{NC} - \overline{\theta}\right)\right]$$

At the optimum, the principal pays to the supervisor $W_s(\overline{\theta}) = kU(\overline{\theta})$ in opposition to the collusive offer by the good type $\overline{\theta}$. The principal also can improve her payoff by increasing X_L^{NC} marginally under the report $r = \phi$, but the information rent $U(\overline{\theta}) = C\left(X_L^{NC} - \underline{\theta}\right) - C\left(X_L^{NC} - \overline{\theta}\right)$ increases for the supervisor and the good-type agent. The increase in the information rent for the supervisor brings about a trade-off for the principal when the supervisor obtains the proof of true information

[9] In section 4.2, we introduce a new idea where the behavioral element (Shading) can become a strong driver that enforces the side contract between the supervisor and the agent. This argument may provide a theoretical or behavioral foundation for the enforceability of side contracts.

[10] This idea is similar to the renegotiation problem from lack of commitment to the long-term contract, which was first considered by Dewatripont (1988)

[11] As is shown below, the output for the good-type agent $\overline{\theta}$ is still the first best X_H^{FB}.

$\theta = \bar{\theta}$, with probability p . Only when the supervisor cannot obtain any information for θ with probability $1-p$, does the principal commit herself to the initial scheme and the standard trade-off between the total surplus and the information rent emerges.

Formally, the expected virtual surplus in the principal-supervisor-agent regime is written as

$$\underbrace{(1-p)}_{\substack{\theta \text{ is not}\\ \text{revealed}}}\left\{\underbrace{h\left[X_H - C\left(X_H-\bar{\theta}\right)\right]+(1-h)\left[X_L - C\left(X_L-\underline{\theta}\right)\right]}_{\text{Expected Total Surplus}} - \underbrace{hU\left(\bar{\theta}\right)}_{\substack{\text{information rent}\\ \text{for the good type}}}\right\}$$

$$+ \underbrace{p}_{\substack{\theta \text{ is}\\ \text{revealed}}}\left\{\underbrace{h\left[X_H^{FB} - C\left(X_H^{FB}-\bar{\theta}\right)\right]+(1-h)\left[X_L^{FB} - C\left(X_L^{FB}-\underline{\theta}\right)\right]}_{\text{(Ex post) First Best Allocative Efficiency}} - \underbrace{hkU\left(\bar{\theta}\right)}_{\substack{\text{information rent}\\ \text{for the supervisor}}}\right\}$$

When the principal determines the contract $\left\{X_L^{NC}, W_L^{NC}\right\}$ for the no-information phase ϕ, she must consider the (expected) information rent for the supervisor $pkU\left(\bar{\theta}\right)$ as well as the (expected) information rent for the good agent $(1-p)U\left(\bar{\theta}\right)$. The principal will optimize the bad-type agent $\underline{\theta}$'s output X_L , in order to mitigate the collusive pressure by the good-type agent when the supervisor observed the signal $\theta = \bar{\theta}$, and to deal with the standard trade-off between the total surplus generated by X_L and the information rent for the good agent $\bar{\theta}$ when the supervisor could not obtain any information for θ. Thus, in this regime, the principal maximizes the following modified *virtual surplus*.

$$\max_{\{X_H, X_L\}} (1-p)\left\{\underbrace{h\left[X_H - C\left(X_H-\bar{\theta}\right)\right]+(1-h)\left[X_L - C\left(X_L-\underline{\theta}\right)\right]}_{\text{Expected Total Surplus}} - \underbrace{hU\left(\bar{\theta}\right)}_{\substack{\text{information rent}\\ \text{for the good type}}}\right\} - p \underbrace{khU\left(\bar{\theta}\right)}_{\substack{\text{information rent}\\ \text{for the supervisor}}}$$

The first order conditions for the optimum are,

$$1-\frac{\partial C\left(X_H-\bar{\theta}\right)}{\partial X_H} = 0 \Leftrightarrow X_H^{NC} = X_H^{FB}$$

$$\underbrace{1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}}_{\text{Marginal Total Surplus}} - \frac{h}{1-h}\underbrace{\left[1+\frac{pk}{1-p}\right]}_{\geq 1}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}} = 0 \qquad (**)$$

We now have the following proposition on the comparison of equilibrium incentives.

Proposition3:

Let X_H^{NC} and X_L^{NC} be the outputs (in the no supervisory information phase ϕ) of the good-type $\overline{\theta}$ and the bad-type $\underline{\theta}$, respectively. Then, we have:

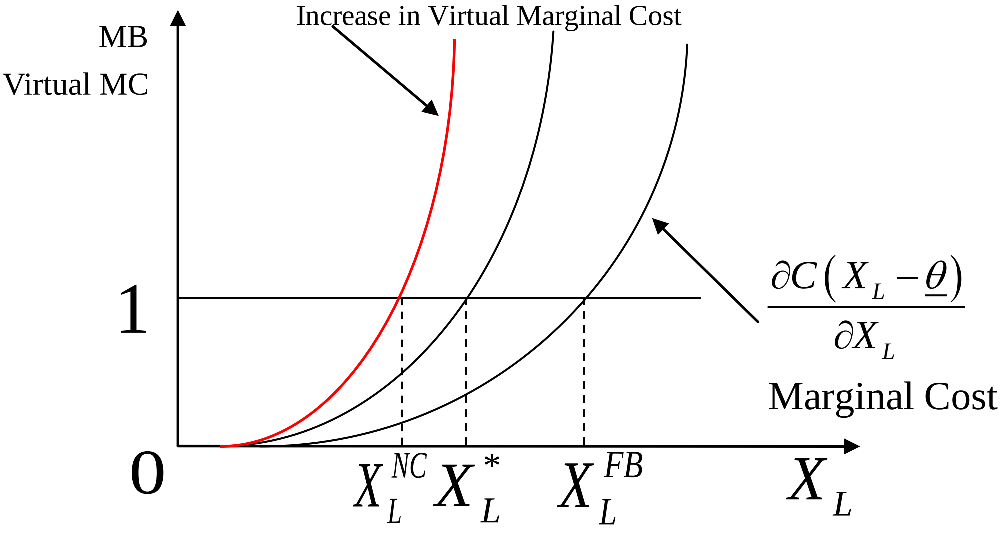
(1) Efficiency at the top (for the good-type agent) $X_H^{NC} = X_H^{FB}$

(2) Downward distortion at the bottom (for the bad-type agent) is aggravated: $X_L^{NC} \leq X_L^{*} \leq X_L^{FB}$

$X_L^{NC} \leq X_L^{*}$ in the above proposition comes from the increase in the virtual cost, i.e. the total and marginal information rents in this regime. Virtual marginal cost increases by $pk/(1-p)$, compared with the standard no-supervisor case. Figure 4 clearly shows this point.

Figure 4

$$\frac{h}{1-h}\underbrace{\left[1+\frac{pk}{1-p}\right]}_{\geq 1}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\overline{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}}$$



Now, we can perform a comparative statics on the optimal solution X_L^{NC} .

Proposition4: Comparative statics on X_L^{NC}

The optimal output X_L^{NC} in this no commitment/renegotiation regime is non-increasing in the parameter p , and non-increasing in the parameter k .

Proof:

The coefficient of the marginal information rent $1+\dfrac{pk}{1-p}$ increases as the parameter p increases. Hence, the virtual marginal information rent (and so the marginal virtual cost) $\dfrac{h}{1-h}\left[1+\dfrac{pk}{1-p}\right]\left[\dfrac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\dfrac{\partial C\left(X_L-\overline{\theta}\right)}{\partial X_L}\right]$ increases as p increases. This brings about the decrease in the optimal output $X_L^{NC}\downarrow$. Similarly, the coefficient of the marginal information rent $1+\dfrac{pk}{1-p}$ increases as the parameter k increases. Hence, the virtual marginal information rent (and so the marginal virtual cost) increases as k increases. This brings about the decrease in the optimal output $X_L^{NC}\downarrow$. ■

4. Efficient Organization Design with Ex-post Adaptation Costs

In this section, we incorporate behavioral elements ala Fehr and Schmidt (1999) into the model.[12] Concretely, we introduce a recent behavioral contract theory idea, "shading behavior" (Hart and Moore (2008)) as a form of ex-post haggling into our collusion model which is based on Tirole (1986, 1992). By combining these two types of models, we try to include a micro foundation into ex-post adaptation costs,[13] and understand the ex-post adaptation as a result of the optimal organizational response to the trade-off between gross total surplus and ex-post adaptation costs associated with the output decisions.

4.1 Shading Model: Observable Collusion

[12] Fehr and Schmidt (1999) is a logical approach to behavioral economics and explains a multitude of evidence. Suzuki (2007) considers a setting where the existence of a behavioral element with a zero-sum structure brings about a strong incentive for vertical collusion in the principal-supervisor-two agent hierarchy, and analyzes the optimal (incomplete) contract design problem.

[13] We introduce rent-seeking associated with collusive behavior and ex-post haggling generated from shading as the two sources of adaptation costs.

We incorporate into the model a formulation of ex-post haggling based on the “shading” model[14] by Hart and Moore (2008), which introduced a new idea that an ex-ante contract provides a reference point for parties’ feelings of entitlement. A party who felt aggrieved in terms of his entitlement shades the party who aggrieved him to the point where his payoff falls by a constant multiplied by the aggrievement level, that is, the former punishes the latter by a constant times the aggrievement level.

In our model, the agent of type $\overline{\theta}$ feels entitled to the information rent (indirect utility) $U\left(\overline{\theta}\right)$ indicated by the initial contract. Nonetheless, the supervisor reported $r=\overline{\theta}$ and aggrieved (disappointed) the agent by exploiting $U\left(\overline{\theta}\right)$. Hence, the agent gets angry and shades (punishes) the supervisor by $\beta U\left(\overline{\theta}\right)$, where β is the parameter of the agent’s shading strength and $\beta \geq 0$. Then, the net payoff of the supervisor when he reports the truth $r=\overline{\theta}$ is $\underbrace{W_s\left(\overline{\theta}\right)}_{\text{Wage Payment}}-\underbrace{\beta U\left(\overline{\theta}\right)}_{\text{shading loss}}$.

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), she 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 she 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 this is *not verifiable*. Only when the principal commits herself to the initial scheme $\left\{X\left(\theta\right),W\left(\theta\right)\right\},\theta\in\left\{\underline{\theta},\overline{\theta}\right\}$ and enforces $X\left(\overline{\theta}\right)=X_H$ for the agent’s report $\hat{\theta}=\overline{\theta}$, can she know the *exact* value of $\theta=\overline{\theta}$, and understand how much she has been aggrieved by the supervisor. Then, she can shade (punish) the supervisor. In summary, this information structure means that collusion (side contracting) between the supervisor and the type $\overline{\theta}$ agent is *observable* but *unverifiable*.

Then, formally, the principal would feel that she had been entitled to $X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)$ since the type information was $\overline{\theta}$. Nonetheless, she could only attain the payoff under asymmetric

[14] This 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”.

information regime between the principal and the agent $\bar{\theta}$, $X_H - C\left(X_H - \bar{\theta}\right) - U\left(\bar{\theta}\right)$, since the supervisor colluded with the agent and hid the information $\bar{\theta}$. In summary, she was aggrieved by

$$\left\{\left[X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right)\right] - \left[X_H - C\left(X_H - \bar{\theta}\right) - U\left(\bar{\theta}\right)\right]\right\}$$

and so she shades (punishes) the supervisor by a constant γ times the aggrievement level

$$\gamma\left\{\left[X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right)\right] - \left[X_H - C\left(X_H - \bar{\theta}\right) - U\left(\bar{\theta}\right)\right]\right\}$$

where γ is the parameter of the principal's shading strength and $\gamma \geq 0$. Thus, we obtain the supervisor's incentive constraint with behavioral assumptions

$$\underbrace{W_s\left(\bar{\theta}\right)}_{\text{wage payment}} - \underbrace{\beta U\left(\bar{\theta}\right)}_{\text{shading loss}} \geq \underbrace{kU\left(\bar{\theta}\right)}_{\text{side payment}} - \underbrace{\gamma\left\{\left[X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right)\right] - \left[X_H - C\left(X_H - \bar{\theta}\right) - U\left(\bar{\theta}\right)\right]\right\}}_{\text{shading loss}}$$

$$\Leftrightarrow \underbrace{W_s\left(\bar{\theta}\right)}_{\text{wage payment}} \geq \underbrace{kU\left(\bar{\theta}\right)}_{\text{side payment}} + \underbrace{\beta U\left(\bar{\theta}\right)}_{\text{shading loss by the agent}}$$
$$\underbrace{-\gamma\left\{\left[X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right)\right] - \left[X_H - C\left(X_H - \bar{\theta}\right) - U\left(\bar{\theta}\right)\right]\right\}}_{\text{shading loss by the principal}}$$

Substituting $W_H = C\left(X_H - \bar{\theta}\right) + U\left(\bar{\theta}\right)$ and

$$W_s\left(\bar{\theta}\right) = kU\left(\bar{\theta}\right) + (\beta - \gamma)U\left(\bar{\theta}\right) - \gamma\left\{\left[X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right)\right] - \left[X_H - C\left(X_H - \bar{\theta}\right)\right]\right\}$$

into the principal's objective function, we have the formulation of virtual surplus for type $\bar{\theta}$

$$p\left(X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right) - \left\{k + (\beta - \gamma)\right\}U\left(\bar{\theta}\right) + \gamma\left\{\left[X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right)\right] - \left[X_H - C\left(X_H - \bar{\theta}\right)\right]\right\}\right)$$
$$+ (1 - p)\left(X_H - C\left(X_H - \bar{\theta}\right) - U\left(\bar{\theta}\right)\right)$$
$$= (1 + \gamma)\, p\left(X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right)\right) - \left[p\left\{k + (\beta - \gamma)\right\} + (1 - p)\right]U\left(\bar{\theta}\right)$$
$$+ \left(1 - (1 + \gamma)\, p\right)\left(X_H - C\left(X_H - \bar{\theta}\right)\right)$$

Then, the *expected* virtual surplus *with behavioral supervisor* can be rewritten as

$$h(1 + \gamma)\, p\left(X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right)\right) - h\left[p\left\{k + (\beta - \gamma)\right\} + (1 - p)\right]U\left(\bar{\theta}\right)$$
$$+ h\left(1 - (1 + \gamma)\, p\right)\left(X_H - C\left(X_H - \bar{\theta}\right)\right)$$
$$+ (1 - h)\left\{p\left(X_L^{FB} - C\left(X_L^{FB} - \underline{\theta}\right)\right) + (1 - p)\left(X_L - C\left(X_L - \underline{\theta}\right)\right)\right\}$$

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

$$\begin{aligned}\max_{X_H,X_L}\ & h(1+\gamma)p\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)-h\left[p\{k+(\beta-\gamma)\}+(1-p)\right]U\left(\bar{\theta}\right)\\ &+h\left(1-(1+\gamma)p\right)\left(X_H-C\left(X_H-\bar{\theta}\right)\right)\\ &+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)\right\}-Z\end{aligned}$$

Since this objective function is additively separable in X_H and X_L, the program can be broken into two:

$$X_H^B\in\arg\max_{X_H} h\left(1-(1+\gamma)p\right)\left(X_H-C\left(X_H-\bar{\theta}\right)\right)$$

$$X_L^B\in\arg\max_{X_L}(1-h)(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-h\left[p\{k+(\beta-\gamma)\}+(1-p)\right]U\left(\bar{\theta}\right)$$

These solutions define the optimal outputs X_L^B, X_H^B.

The First Order Conditions for the optimum are

$$1-\frac{\partial C\left(X_H-\bar{\theta}\right)}{X_H}=0\Leftrightarrow X_H^B=X_H^{FB}$$

$$(1-h)(1-p)\left(1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}\right)-h\left[p\{k+(\beta-\gamma)\}+(1-p)\right]\frac{\partial U\left(\bar{\theta}\right)}{\partial X_L}=0$$

$$\Leftrightarrow 1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{h}{1-h}\left[1+\frac{p}{(1-p)}\{k+(\beta-\gamma)\}\right]\frac{\partial U\left(\bar{\theta}\right)}{\partial X_L}=0$$

$$\Leftrightarrow 1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}=\frac{h}{(1-h)}\left(1+\frac{p}{1-p}\{k+(\beta-\gamma)\}\right)\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]}_{\text{marginal information rent}}\qquad(***)$$

Especially, the principal's program for type $\underline{\theta}$ can be rewritten as

$$\begin{aligned}\max_{X_L} J^B\left(X_L,\underline{\theta}\right)&=\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-\left[1+\frac{p}{(1-p)}\{k+(\beta-\gamma)\}\right]\frac{h}{(1-h)}U\left(\bar{\theta}\right)\\ &=\underbrace{\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-\left[1+\frac{pk}{(1-p)}\right]\frac{h}{(1-h)}U\left(\bar{\theta}\right)}_{\text{Virtual Surplus in No-Commitment Regime}}-\underbrace{\frac{p}{(1-p)}\frac{h}{(1-h)}(\beta-\gamma)U\left(\bar{\theta}\right)}_{\text{Change in Dead Weight Loss through Shading}}\end{aligned}$$

where $(1-h)/h$ is the hazard rate. Therefore we have the following proposition.

Proposition 5:

The optimal solution X_L^B for type $\underline{\theta}$ with behavioral elements is smaller than the solution X_L^{NC} with no behavioral elements, that is, $X_L^B \le X_L^{NC}$ if and only if $\beta \ge \gamma$ and similarly $X_L^B \ge X_L^{NC}$ if and only if $\beta \le \gamma$.

Proof:

In the following formulations of marginal virtual surplus of the two regimes: No Commitment (NC) and Behavioral regimes (B),

$$\frac{\partial J^{NC}\left(X_L,\underline{\theta}\right)}{\partial X_L} = (1-p)\left(1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}\right) - \frac{\left[(1-p)+pk\right]}{h(\theta)}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}}$$

$$\begin{aligned}\frac{\partial J^{B}\left(X_L,\underline{\theta}\right)}{\partial X_L} = {} & (1-p)\left[1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}\right] - \frac{\left[(1-p)+pk\right]}{h(\theta)}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}} \\ & \underbrace{-\frac{(\beta-\gamma)}{h(\theta)}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]}}_{\text{Marginal Information Rent through Shading}}\end{aligned}$$

The optimal solution X_L^{NC} satisfies the first-order condition $\frac{\partial J^{NC}\left(X_L^{NC},\underline{\theta}\right)}{\partial X_L} = 0$. Then,

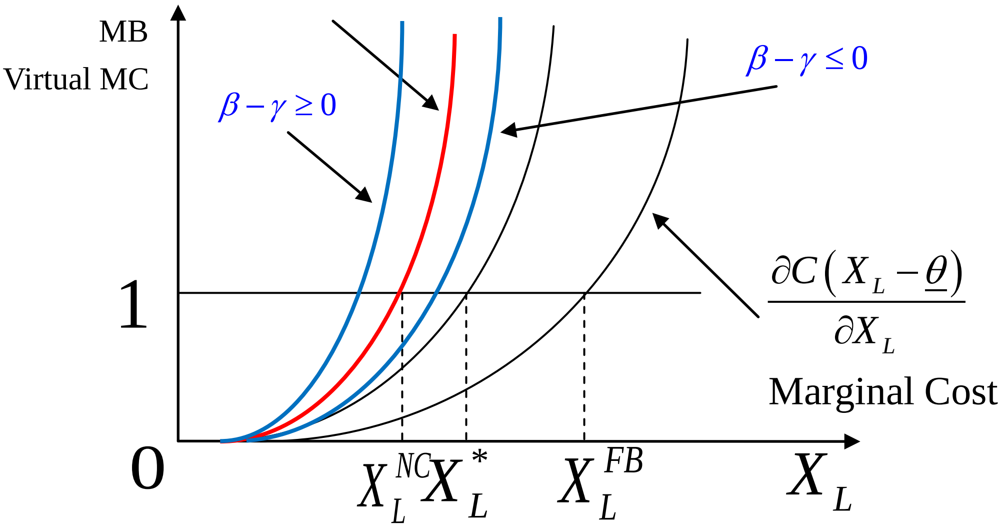
$$\frac{\partial J^{B}\left(X_L^{NC},\underline{\theta}\right)}{\partial X_L} = \underbrace{-\frac{(\beta-\gamma)}{h(\theta)}\underbrace{\left[\frac{\partial C\left(X_L^{NC}-\underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L^{NC}-\bar{\theta}\right)}{\partial X_L}\right]}_{+}}_{\text{Marginal Information Rent through Shading}} \le 0 \;\text{ for }\; \beta \ge \gamma$$

Therefore X_L^{NC} cannot be optimal for the Behavioral regimes (B). A marginal decrease in X_L from X_L^{NC} would increase the virtual surplus $J^B\left(X_L,\underline{\theta}\right)$ of Behavioral regimes (B). Thus, we have $X_L^B \le X_L^{NC}$ for $\beta \ge \gamma$ and vice versa. ■

The following figure clearly shows the point.

$$\frac{h}{1-h}\underbrace{\left[1+\frac{p\left\{k+(\beta-\gamma)\right\}}{1-p}\right]}_{\geq 1}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\overline{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}}$$

Change in Virtual Marginal Cost in (NC) Regime



Theoretical Intuition

Remember that the supervisor’s reward is

$$W_S^B\left(\overline{\theta}\right)=kU\left(\overline{\theta}\right)+(\beta-\gamma)U\left(\overline{\theta}\right)-\gamma\left\{\left[X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)\right]-\left[X_H^B-C\left(X_H^B-\overline{\theta}\right)\right]\right\}$$

where $U\left(\overline{\theta}\right)=C\left(X_L^B-\underline{\theta}\right)-C\left(X_L^B-\overline{\theta}\right)$ is **the information rent** for the agent $\overline{\theta}$ given the output X_L^B, which also implies **the potential for aggrievement for the agent** $\overline{\theta}$

First, when the output X_L^B increases marginally, the information rent $U\left(\overline{\theta}\right)$ goes up. Second, whether the net shading behaviors $(\beta-\gamma)U\left(\overline{\theta}\right)$ increase or not depends on the sign of parameters $\beta-\gamma\geq 0\left(\beta-\gamma\leq 0\right)$ When $\beta\geq\gamma$, the marginal increase in the output X_L^B will bring about not only the standard increase in the information rent $U\left(\overline{\theta}\right)$, but also the marginal in the potential for aggrievement for the agent $\overline{\theta}$, which makes the shading behavior by the agent

$\bar{\theta}$ severer. These effects increase the supervisor's wage $W_s\left(\theta\right)$ **discretely** and generate **a first-order loss** for the principal. Indeed, the increase in $X\left(\theta\right)$ generates **a second-order gain** through the change of optimal solution, but the principal's profit will go down totally (due to the **first-order loss vs. second-order gain**) for $\beta \geq \gamma$ Thus, the optimal solution with the behavioral supervisor X_L^B will fall below the No-commitment (no behavioral supervisor) solution X_L^{NC} , that is $X_L^B \leq X_L^{NC}$. A similar rationale holds for $\beta \leq \gamma$

Relation to the "Influence Activities" and Efficient Organization Design by Milgrom (1988)

The principal can design the optimal output X_L^B to modify equilibrium shading behaviors through controlling the potential for aggrievement, i.e. information rent $U\left(\bar{\theta}\right)$. This 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.

Now, we can perform a comparative statics on the optimal solution X_L^B

Proposition6: *The optimal output* X_L^B *for type* $\underline{\theta}$ *with behavioral supervisor is nonincreasing in parameter* β *(the degree of shading strength by the agent), but nondecreasing in* γ *(the degree of shading strength by the principal)*[15]

Proof: From $\left(**\right)$, the derivative $J_{X_L}^B\left(X_L,\theta\right)$ is nonincreasing in β (behavioral elements). That is,

$$J_{X_L\beta}^B\left(X_L,\theta\right)=-\frac{p}{1-p}\frac{h}{1-h}\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right]\leq 0$$

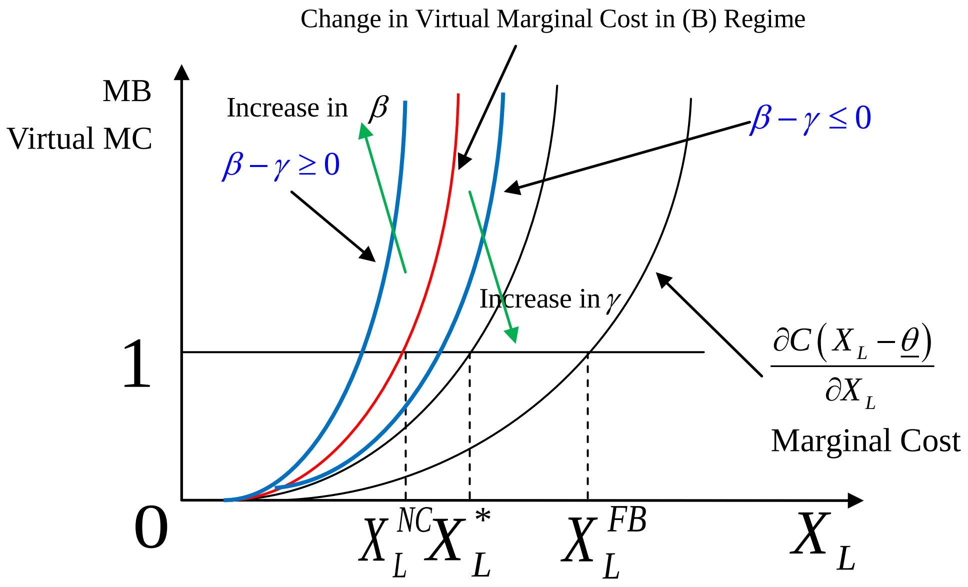
Hence, the optimal solution with behavioral supervisor X_L^B is nonincreasing in β . Similarly,

[15] Comparative statics on p,k is essentially the same as proposition 4.

$$J^{B}_{X_L\gamma}\left(X_L,\theta\right)=\frac{p}{1-p}\frac{h}{1-h}\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\overline{\theta}\right)}{\partial X_L}\right]\geq 0$$

Hence, the optimal solution with behavioral supervisor X_L^B is nondecreasing in γ. ■

$$\frac{h}{1-h}\underbrace{\left[1+\frac{p\left\{k+\left(\beta-\gamma\right)\right\}}{1-p}\right]}_{\geq 1}\underbrace{\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\overline{\theta}\right)}{\partial X_L}\right]}_{\text{Marginal Information Rent}}$$



<u>Proposition7:</u> The principal's equilibrium payoff *decreases* in the regime (B) with behavioral supervisor, in comparison with the regime (NC) without behavioral supervisor, when the shading strength by the agent is greater than that by the principal, i.e., $\beta\geq\gamma$, while on the other hand, the principal's equilibrium payoff *can increase* when the shading strength by the principal is greater than that by the agent $\gamma\geq\beta$.

Proof: The expected virtual surplus in the regime (NC) is

$$h\left\{p\left(X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)-kU^{NC}\left(\overline{\theta}\right)\right)+\left(1-p\right)\left(X_H^{NC}-C\left(X_H^{NC}-\overline{\theta}\right)-U^{NC}\left(\overline{\theta}\right)\right)\right\}$$
$$+\left(1-h\right)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+\left(1-p\right)\left(X_L^{NC}-C\left(X_L^{NC}-\underline{\theta}\right)\right)\right\}$$
$$\text{where } U^{NC}\left(\overline{\theta}\right)=C\left(X_L^{NC}-\underline{\theta}\right)-C\left(X_L^{NC}-\overline{\theta}\right)$$

The <u>maximized</u> expected virtual surplus in the regime (NC) is

$$\begin{aligned}
&hp\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)-h\left[pk+(1-p)\right]U^{NC}\left(\bar{\theta}\right)\\
&+h(1-p)\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)\\
&+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L^{NC}-C\left(X_L^{NC}-\underline{\theta}\right)\right)\right\}
\end{aligned}$$

Next, the expected virtual surplus in the behavioral regime (B) is

$$\begin{aligned}
&h\left\{p\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)-W_S^B\left(\bar{\theta}\right)\right)+(1-p)\left(X_H^B-C\left(X_H^B-\bar{\theta}\right)-U^B\left(\bar{\theta}\right)\right)\right\}\\
&+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L^B-C\left(X_L^B-\underline{\theta}\right)\right)\right\}\\
&\text{where } U^B\left(\bar{\theta}\right)=C\left(X_L^B-\underline{\theta}\right)-C\left(X_L^B-\bar{\theta}\right)
\end{aligned}$$

By remembering the following coalition-proof constraint with behavioral supervisor

$$W_S^B\left(\bar{\theta}\right)-\beta U^B\left(\bar{\theta}\right)\geq kU^B\left(\bar{\theta}\right)-\gamma\left\{\left[X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right]-\left(X_H^B-C\left(X_H^B-\bar{\theta}\right)-U^B\left(\bar{\theta}\right)\right)\right\},$$

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

$$\begin{aligned}
&h(1+\gamma)p\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)-h\left[p\left\{k+(\beta-\gamma)\right\}+(1-p)\right]U^B\left(\bar{\theta}\right)\\
&+h\left(1-(1+\gamma)p\right)\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)\\
&+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L^B-C\left(X_L^B-\underline{\theta}\right)\right)\right\}
\end{aligned}$$

This is transformed as follows.

$$\begin{aligned}
&hp\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)+h\gamma p\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)\\
&-h\left[pk+(1-p)\right]U^B\left(\bar{\theta}\right)-h(\beta-\gamma)U^B\left(\bar{\theta}\right)\\
&+h(1-p)\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)-h\gamma p\left(X_H^{FB}-C\left(X_H^{FB}-\bar{\theta}\right)\right)\\
&+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L^B-C\left(X_L^B-\underline{\theta}\right)\right)\right\}
\end{aligned}$$

Taking the difference of the two maximized expected virtual surpluses, the condition for the expected virtual surplus in the regime (B) to be greater than the one in the regime (NC) is as follows.

$$\begin{aligned}
&0-h(\beta-\gamma)U^B\left(\bar{\theta}\right)\\
&\geq -h\left[pk+(1-p)\right]\left[U^{NC}\left(\bar{\theta}\right)-U^B\left(\bar{\theta}\right)\right] \qquad \left(\mathrm{VS}^B\geq \mathrm{VS}^{NC}\right)(*)\\
&\quad +(1-h)(1-p)\left\{\left(X_L^{NC}-C\left(X_L^{NC}-\underline{\theta}\right)\right)-\left(X_L^B-C\left(X_L^B-\underline{\theta}\right)\right)\right\}
\end{aligned}$$

The RHS of the inequality is positive from the following “Revealed Preference” relation

$$(1-h)(1-p)\left(X_L^{NC}-C\left(X_L^{NC}-\underline{\theta}\right)\right)-h\left[pk+(1-p)\right]U^{NC}\left(\overline{\theta}\right)$$
$$\geq (1-h)(1-p)\left(X_L^{B}-C\left(X_L^{B}-\underline{\theta}\right)\right)-h\left[pk+(1-p)\right]U^{B}\left(\overline{\theta}\right)$$

The LHS of the inequality is non-positive when $\beta \geq \gamma$, and non-negative when $\beta \leq \gamma$ ■

Rationale:

The LHS $-h\left(\beta-\gamma\right)U^{B}\left(\overline{\theta}\right)$ of the inequality $\left(\mathrm{VS}^{B} \geq \mathrm{VS}^{NC}\right)$ is 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 her shading threat (γ times aggrievement) to the supervisor, thereby increasing her profit.[16] Under the information structure where collusion (side contracting) between supervisor and agent is *observable* ex post for the principal but unverifiable, the introduction of behavioral supervisor, together with the fear of being "shaded" by the principal, can *relax* the supervisor's incentive constraint (coalition incentive constraint), and can thereby increase the principal's equilibrium profit, when $\gamma \geq \beta$ (if the positive effect of LHS dominates the negative effect of the RHS).

Now, since $X_H^{B} = X_H^{FB}$ in equilibrium, we have $W_S^{B}\left(\overline{\theta}\right) = kU^{B}\left(\overline{\theta}\right) + \left(\beta-\gamma\right)U^{B}\left(\overline{\theta}\right)$

Then, the supervisor's equilibrium payoff under shading is

$$W_S^{B}\left(\overline{\theta}\right) - \beta U^{B}\left(\overline{\theta}\right) = \left(k-\gamma\right)U^{B}\left(\overline{\theta}\right)$$

Thus, we have the condition for the supervisor's IR constraint to be satisfied $k \geq \gamma$, which means that the shading by the principal is not too strong. It follows that we have the following corollary.

Corollary:

The conditions under which the principal's equilibrium profit can increase by the introduction of the behavioral supervisor and his IR constraint also holds are $\left(\mathrm{VS}^{B} \geq \mathrm{VS}^{NC}\right)$ and $\beta \leq \gamma \leq k$.

4.2 Shading Model: Unobservable Collusion

[16] As an analogy for the moral hazard model with a 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.

Now, suppose that the supervisor's signal $s \in \{\theta, \phi\}$ is 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

$$\underbrace{W_s(\theta)}_{\text{wage payment}} - \underbrace{\beta U(\theta)}_{\text{shading loss}} \geq \underbrace{kU(\theta)}_{\text{side payment}} \Leftrightarrow \underbrace{W_s(\theta)}_{\text{wage payment}} \geq \underbrace{kU(\theta)}_{\text{side payment}} + \underbrace{\beta U(\theta)}_{\text{shading loss}}$$

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.

Proposition8: Suppose that collusion (side contracting) between supervisor and agent is *unobservable* ex post for the principal. Then, only the agent can shade the supervisor, which corresponds to $\beta > 0, \gamma = 0$. Then, the principal's equilibrium payoff is always reduced in the regime with behavioral supervisor, in comparison with that without behavioral supervisor $\beta = \gamma = 0$. That is, "shading" becomes detrimental to organizational design

Proof: When $\beta > 0, \gamma = 0$, the above condition that $\left(\mathrm{VS}^{B} \geq \mathrm{VS}^{NC}\right)$ does not hold. That is, $\mathrm{VS}^{B} < \mathrm{VS}^{NC}$, which means that the maximized expected virtual surplus is smaller in the behavioral regime (B) than in the no behavioral regime (NC). ■

4.2.1 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 $\bar{\theta}$ and the supervisor through the collusion-proof constraint and induces the supervisor's truth telling $r = \bar{\theta}$,and the other 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

$$\underbrace{W_s\left(\bar{\theta}\right)}_{\text{wage payment}} = \underbrace{kU\left(\bar{\theta}\right)}_{\text{side payment}} + \underbrace{\beta U\left(\bar{\theta}\right)}_{\text{shading loss by agent}} = (k+\beta)U\left(\bar{\theta}\right)$$

Then, the expected virtual profit for the principal is

$$h\left\{p\left[X_H^{FB} - C\left(X_H^{FB} - \bar{\theta}\right) - (k+\beta)U\left(\bar{\theta}\right)\right] + (1-p)\left(X_H - C\left(X_H - \bar{\theta}\right) - U\left(\bar{\theta}\right)\right)\right\}$$
$$+(1-h)\left\{p\left(X_L^{FB} - C\left(X_L^{FB} - \underline{\theta}\right)\right) + (1-p)\left(X_L - C\left(X_L - \underline{\theta}\right)\right)\right\}$$

The principal maximizes it over X_H and X_L

$$\max_{X_H} X_H - C\left(X_H - \bar{\theta}\right) \Rightarrow X_H^{CP} = X_H^{FB}$$

$$\max_{X_L} (1-h)(1-p)\left(X_L - C\left(X_L - \underline{\theta}\right)\right) - h\left[p(k+\beta) + (1-p)\right]U\left(\bar{\theta}\right)$$

$$\max_{X_L}\left(X_L - C\left(X_L - \underline{\theta}\right)\right) - \left[1 + \frac{p}{(1-p)}\{k+\beta\}\right]\frac{h}{(1-h)}U\left(\bar{\theta}\right)$$

$$\Leftrightarrow \max_{X_L}\underbrace{\left(X_L - C\left(X_L - \underline{\theta}\right)\right) - \left[1 + \frac{pk}{(1-p)}\right]\frac{h}{(1-h)}U\left(\bar{\theta}\right)}_{\text{Virtual Surplus in No-Commitment (NC) Regime}} \underbrace{-\frac{p}{(1-p)}\frac{h}{(1-h)}\beta U\left(\bar{\theta}\right)}_{\text{Change in Dead Weight Loss through Shading}}$$

First order condition for the optimality on X_L is

$$1 - \frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} = \left[1 + \frac{p}{(1-p)}\{k+\beta\}\right]\frac{h}{(1-h)}\frac{\partial U\left(\bar{\theta}\right)}{\partial X_L}$$

$$\Leftrightarrow 1 - \frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} = \left[1 + \frac{p}{(1-p)}\{k+\beta\}\right]\frac{h}{(1-h)}\left[\frac{\partial C\left(X_L - \underline{\theta}\right)}{\partial X_L} - \frac{\partial C\left(X_L - \bar{\theta}\right)}{\partial X_L}\right] \qquad (***)'$$

Corollary:

The optimal solution X_L^B with behavioral elements under the collusion-proof regime is smaller than

the optimal solution X_L^{NC} with no behavioral elements, that is, $X_L^B \le X_L^{NC}$

Proof: See the proof of Proposition5. This is the case where $\beta > 0, \gamma = 0$

Equilibrium Collusion Regime (EC)

In this regime, when the supervisor observes the proof on $\bar{\theta}$ with probability p, the principal allows the collusion between the agent $\bar{\theta}$ and the supervisor in equilibrium, which means that the supervisor reports $r=\phi$ and the agent $\bar{\theta}$ chooses X_H in exchange for the information rent $U\left(\bar{\theta}\right)$. Then, the principal pays the information rent $U\left(\bar{\theta}\right)$ to the agent $\bar{\theta}$ at the unit transfer price 1. Hence, the expected virtual profit for the principal is

$$h\left\{p\left(X_H-C\left(X_H-\bar{\theta}\right)-U\left(\bar{\theta}\right)\right)+(1-p)\left(X_H-C\left(X_H-\bar{\theta}\right)-U\left(\bar{\theta}\right)\right)\right\}$$

$$+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)\right\}$$

$$=h\left(X_H-C\left(X_H-\bar{\theta}\right)-U\left(\bar{\theta}\right)\right)$$

$$+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)\right\}$$

The principal maximizes it over X_H and X_L

$$\max_{X_H} X_H-C\left(X_H-\bar{\theta}\right)\Rightarrow X_H^{EC}=X_H^{FB}$$

$$\max_{X_L}(1-h)(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-hU\left(\bar{\theta}\right)$$

$$\Leftrightarrow \max_{X_L}\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-\frac{h}{(1-h)}\frac{1}{(1-p)}U\left(\bar{\theta}\right)$$

First order condition for the optimality on X_L is

$$1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}=\frac{h}{(1-h)}\frac{1}{(1-p)}\frac{\partial U\left(\bar{\theta}\right)}{\partial X_L}$$

$$\Leftrightarrow 1-\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}=\frac{h}{(1-h)}\frac{1}{(1-p)}\left[\frac{\partial C\left(X_L-\underline{\theta}\right)}{\partial X_L}-\frac{\partial C\left(X_L-\bar{\theta}\right)}{\partial X_L}\right] \qquad (****)$$

Now, we have the following lemma and proposition.

Lemma: Comparison on equilibrium incentives between Two Regimes

Comparing FOCs on X_L in the two regimes (CP and EC), we have

$$X_L^{EC}\le X_L^{CP} \text{ if and only if } \frac{1}{(1-p)}\ge 1+\frac{p\{k+\beta\}}{1-p}\Leftrightarrow 1-k\ge\beta$$

$$X_L^{EC}\ge X_L^{CP} \text{ if and only if } \frac{1}{(1-p)}\le 1+\frac{p\{k+\beta\}}{1-p}\Leftrightarrow 1-k\le\beta$$

Proposition9:

The principal optimally chooses the Collusion-proof regime (CP) if the shading parameter $\beta \leq 1-k$, and the Equilibrium Collusion Regime (EC) if the shading parameter $\beta \geq 1-k$.

Proof: Substituting the optimal solution $X_H^{CP} = X_H^{FB}$ for the type $\overline{\theta}$, the expected virtual profit for the principal in the Collusion-Proof regime (CP) is written as follows.

$$h\left\{p\left[X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)-(k+\beta)U\left(\overline{\theta}\right)\right]+(1-p)\left(X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)-U\left(\overline{\theta}\right)\right)\right\}$$

$$+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)\right\}$$

$$=(1-h)(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-h\left[(1-p)+p(k+\beta)\right]U\left(\overline{\theta}\right)$$

$$+h\left(X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)\right)+(1-h)p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)$$

$$\text{where } U\left(\overline{\theta}\right)=C\left(X_L-\underline{\theta}\right)-C\left(X_L-\overline{\theta}\right)$$

Similarly, substituting the optimal solution $X_H^{EC} = X_H^{FB}$ for the type $\overline{\theta}$, the expected virtual profit for the principal in Equilibrium Collusion regime (EC) is written as follows.

$$h\left\{p\left(X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)-U\left(\overline{\theta}\right)\right)+(1-p)\left(X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)-U\left(\overline{\theta}\right)\right)\right\}$$

$$+(1-h)\left\{p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)+(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)\right\}$$

$$=(1-h)(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-hU\left(\overline{\theta}\right)$$

$$+h\left(X_H^{FB}-C\left(X_H^{FB}-\overline{\theta}\right)\right)+(1-h)p\left(X_L^{FB}-C\left(X_L^{FB}-\underline{\theta}\right)\right)$$

Hence, which regime can achieve the higher efficiency depends on the comparison of the following two optimal values.

$$VS^{CP}=\max_{X_L}(1-h)(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-h\left[(1-p)+p(k+\beta)\right]U\left(\overline{\theta}\right)$$

$$VS^{EC}=\max_{X_L}(1-h)(1-p)\left(X_L-C\left(X_L-\underline{\theta}\right)\right)-hU\left(\overline{\theta}\right)$$

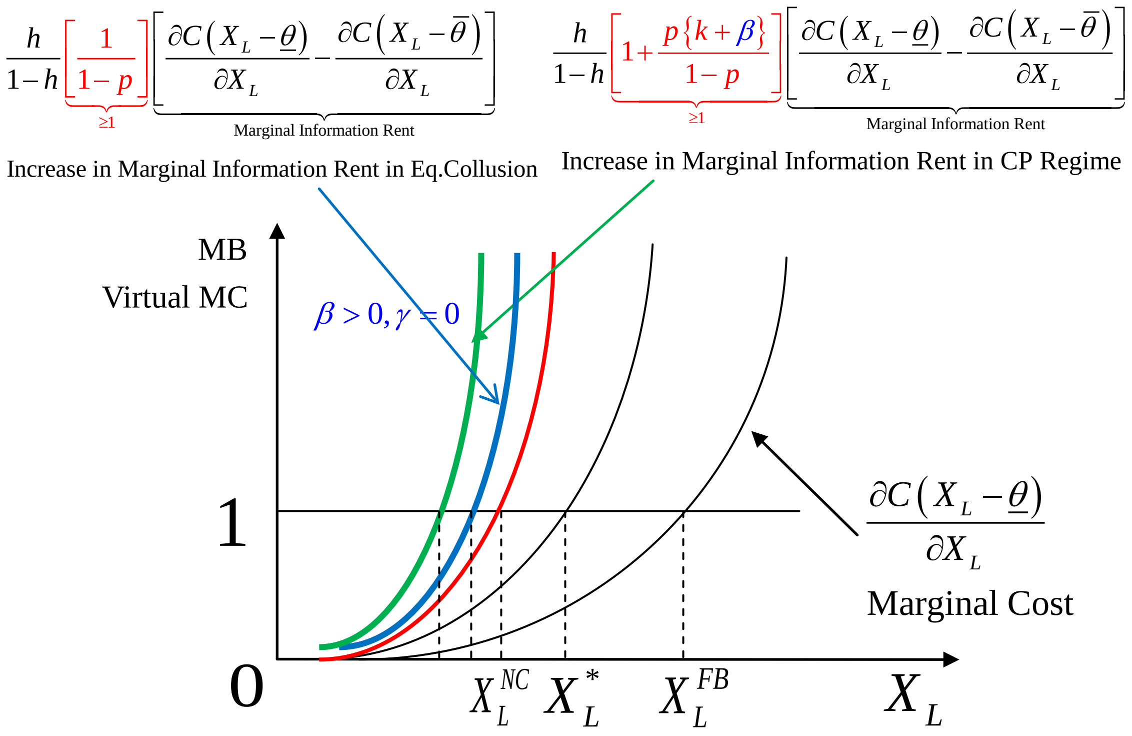
By applying the optimization and envelope theorem, we find that

$$VS^{CP}\geq VS^{EC}\Leftrightarrow(1-p)+p(k+\beta)\leq 1\Leftrightarrow\beta\leq 1-k$$
$$VS^{CP}\leq VS^{EC}\Leftrightarrow(1-p)+p(k+\beta)\geq 1\Leftrightarrow\beta\geq 1-k$$

Putting this together with the above lemma, when the shading strength $\beta \leq 1-k$, we have $X_L^{CP} \geq X_L^{EC} \Leftrightarrow VS^{CP} \geq VS^{EC}$, and so the principal optimally chooses the Collusion-Proof regime

(CP). Similarly, when the shading strength $\beta \geq 1-k$, we have $X_L^{CP} \leq X_L^{EC} \Leftrightarrow VS^{CP} \leq VS^{EC}$, and so the principal optimally chooses the Equilibrium Collusion regime (CP). ■

The following figure represents the case where the Equilibrium Collusion is optimal, that is, $X_L^{CP} \leq X_L^{EC} \Leftrightarrow VS^{CP} \leq VS^{EC}$



$\frac{h}{1-h}\underbrace{\left[\frac{1}{1-p}\right]}_{\geq 1}\underbrace{\left[\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}-\frac{\partial C(X_L-\overline{\theta})}{\partial X_L}\right]}_{\text{Marginal Information Rent}}$
Increase in Marginal Information Rent in Eq.Collusion
$\frac{h}{1-h}\underbrace{\left[1+\frac{p\{k+\beta\}}{1-p}\right]}_{\geq 1}\underbrace{\left[\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}-\frac{\partial C(X_L-\overline{\theta})}{\partial X_L}\right]}_{\text{Marginal Information Rent}}$
Increase in Marginal Information Rent in CP Regime
MB
Virtual MC
$\beta > 0, \gamma = 0$
1
0
$\frac{\partial C(X_L-\underline{\theta})}{\partial X_L}$
Marginal Cost
X_L^{NC}
X_L^{*}
X_L^{FB}
X_L

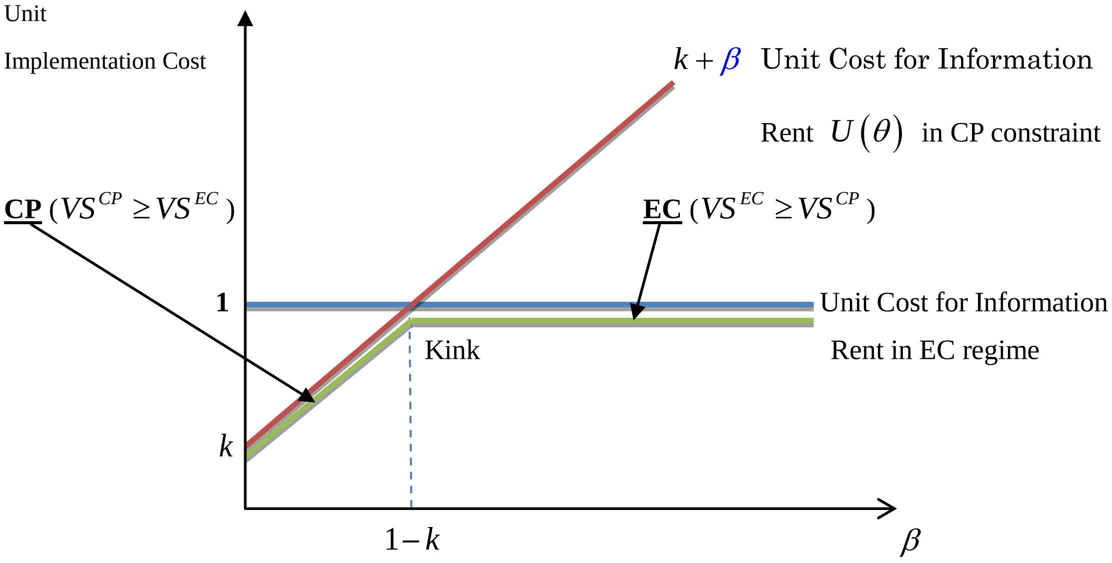
Rationale

As the degree of shading β ("threat" by the agent) increases, the incentive for collusion between the agent $\overline{\theta}$ and the supervisor increases. Thereby, it becomes more costly for the principal to impose collusion-proof schemes and deter collusion, and to induce 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 high productivity agent $\overline{\theta}$ and the supervisor, and then attain the higher efficiency through discretely reducing the ex-post aggrievement and shading by the high productivity agent $\overline{\theta}$.

The figure below shows the essence of the argument. As β increases, it becomes more costly for the principal to impose collusion-proof schemes, deter collusion and then induce truth telling from the supervisor $r = \bar{\theta}$, since the unit cost $k + \beta$ for information rent $U\left(\bar{\theta}\right)$ increases as β goes up in the Collusion-proof (CP) problem. When $k + \beta \geq 1 \Leftrightarrow \beta \geq 1 - k$, it becomes better for the principal to allow collusion between the agent $\bar{\theta}$ and the supervisor in equilibrium, since the principal then pays the information rent $U\left(\bar{\theta}\right)$ to the agent $\bar{\theta}$ just at the unit transfer price 1.

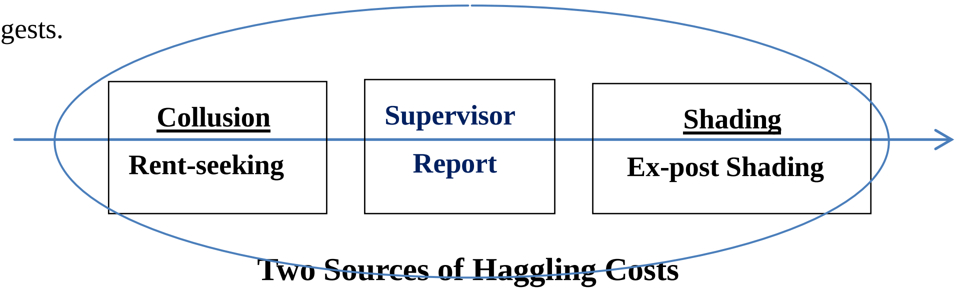
Figure



This is a new idea in the Collusion literature a la Tirole (1986, 1992) in that the increase in shading pressure β (**behavioral element**) strengthens the incentive for collusion, thereby making 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 high productivity agent and the supervisor in equilibrium, and the supervisor reports $r = \phi$ ("I did not observe any information") and the high productivity type $\bar{\theta}$ reveals his type information by self-selecting $\left\{X_H, W_H\right\}$ and obtains the information rent $U\left(\bar{\theta}\right)$

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 high productivity agent $\overline{\theta}$ occurs, since the high productivity agent $\overline{\theta}$ expected to obtain the best reward for himself, that is, the information rent $U\left(\overline{\theta}\right)$, but was aggrieved to have lost it due to the supervisory report $r = \overline{\theta}$. Therefore, the high productivity agent $\overline{\theta}$ shades the supervisor by the shading parameter β times the aggrievement level $U\left(\overline{\theta}\right)$. 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 high productivity agent $\overline{\theta}$ and the supervisor, which may be costly by itself but does not generate any aggrievement for the high productivity agent $\overline{\theta}$, since he can indeed obtain the information rent $U\left(\overline{\theta}\right)$ (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.

As the degree of shading β increases, the incentive for collusion between the high productivity agent $\overline{\theta}$ and the supervisor increases. Thereby, it becomes more costly for the principal to impose collusion-proof schemes and deter collusion, and to induce 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 high productivity agent $\overline{\theta}$.

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 making 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).

5. Conclusion

We introduced the recent behavioral contract theory idea, "shading" (Hart and Moore (2007, 2008)) as a component of ex-post haggling (addressed by Coase (1937) and Williamson (1975)) into the collusion model a la Tirole (1986, 1992), thereby constructing a new model of internal hierarchical organization. By combining these two ideas, i.e., *collusion* and *shading*, we could not only enrich the existing collusion model, thereby obtaining a new result on *Collusion-proof* vs. *Equilibrium Collusion* in that the increase in shading pressure (behavioral element) strengthened the incentive for collusion, thereby making it difficult to implement the collusion-proof (Supervisor's truth telling) incentive schemes, which led to the Equilibrium Collusion, but also gave a micro foundation (an explicit modeling) to ex-post adaptation costs, where we viewed rent-seeking associated with collusive behavior and ex-post haggling generated from aggrievement and shading as the two sources of the costs. 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. We believe that our model could help provide a deep understanding of resource allocation and decision process in the internal organization of large firms.

Suzuki (2012) constructed a *continuous-type*, three-tier agency model with hidden information and collusion a la Tirole (1986, 1992), adopted the First Order (Mirrlees) Approach and the Monotone Comparative Statics method, and analyzed almost the same situation as the current paper more theoretically. The current paper is a *two-type*, more simplified model and places more emphasis on an application to Transaction Cost Economics (TCE) and Efficient Organization Design. In that sense, our two papers are complementary.

REFERENCES

Baron, D. and R, Myerson（1982) "Regulating a Monopolist with Unknown Cost," *Econometrica* 50. 911-930

Bolton, P and M. Dewatripont (2005) *Contract Theory* MIT Press

Coase, R. (1937). "The Nature of the Firm", *Economica*, N.S., 4(16), pp. 386-405.

Dewatripont, M (1988) "Commitment and Information Revelation over Time: The case of Optimal Labor Contracts." *Quarterly Journal of Economics*, 104, 589-619.

Fehr, E. and K. Schmidt (1999) "A Theory of Fairness, Competition and Cooperation" *Quarterly Journal of Economics*, Vol.114, No3, 817-868.

Gibbons, R. (2005) "Four Formal(izable)Theories of the Firm" *Journal of Economic Behavior and Organization,* 58(2): 200–45.

Gibbons, R. (2010) "Transaction-Cost Economics: Past, Present, and Future" *Scandinavian Journal of Economics*, Vol. 112, No. 2. pp. 263-288.

Hart, O. and Holmstrom, B. (2010) “A Theory of Firm Scope”, *Quarterly Journal of Economics*, 125 (2): 483-513.

Hart, O., and J. Moore (2007) “Incomplete Contracts and Ownership : Some New Thoughts” *American Economic Review Papers and Proceedings*. Vol. 97, No. 2, pp. 182-186.

Hart, O. and J. Moore (2008) “Contracts as Reference Points," *Quarterly Journal of Economics*, vol. 123(1), pp. 1-48, 02.

Kofman, F, and J. Lawarree (1993) “Collusion in Hierarchical Agency”, *Econometrica*, Vo1.61, No3, May, 629-656.

Laffont, J-J and D. Martimort (1997) “Collusion under Asymmetric Information” *Econometrica*, Vol.65, No4, 875-911.

Laffont, J-J, and J. Tirole (1988) “The Dynamics of Incentive Contracts,” *Econometrica*, Vol. 56 No.5, 1153-75, September

Laffont, J-J and J. Tirole (1991) “The Politics of Government Decision-Making: A Theory of Regulatory Capture," *Quarterly Journal of Economics.*106, 1089-1127.

Maskin, E and J, Riley (1984) "Monopoly with Incomplete Information," *RAND Journal of Economics,* Vol.15. No.2. 171-196.

Milgrom, P. (1988) “Employment Contracts, Influence Activities and Efficient Organization Design,” *Journal of Political Economy*, 96, 42-60.

Milgrom, P and J.Roberts. (1992) *Economics, Organization and Management*, Prentice-Hall, Englewood Cliffs.

Simon, H. (1951) “A formal theory of the employment relationship,” *Econometrica* Vol.19 293-305

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

Suzuki, Y. (2008)"Mechanism Design with Collusive Supervision: A Three-tier Agency Model with a Continuum of Types,” *Economics Bulletin*, Vol. 4 No. 12. 1-10

Suzuki, Y. (2012) “Collusive Supervision and Organization Design in a Three-tier Agency Model with a Continuum of Types”, mimeo-graphed.

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

Tirole, J. (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.

Williamson, O. (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.

Williamson, Oliver E. (1985) The Economic Institutions of Capitalism. New York: Free Press.