

# Bribery vs. extortion: allowing the lesser of two evils

August 8, 2008

FAHAD KHALIL<sup>†</sup>

JACQUES LAWARRÉE<sup>‡</sup>

SUNGHO YUN<sup>\*</sup>

## Abstract

Both bribery and extortion weaken the power of incentives, but there is a tradeoff in fighting the two since rewards to prevent supervisors from accepting bribes create incentives for extortion. Which is the worse evil? Can a supervisor who can engage in bribery and extortion still be useful in providing incentives? By highlighting the role of team work in forging information, we present a notion of soft information that makes supervision valuable. We show that a fear of inducing extortion may make it optimal to allow bribery, but extortion is never tolerated. Extortion discourages “good behavior” because the agent is subject to it even though he has done the right thing. Bribery, however, helps somewhat in providing incentive because it occurs when the agent is seen as violating the rules. The bribery payment acts as a penalty for “bad behavior”. We find that extortion is a more serious issue when incentives are primarily based on soft information, when the agent has a greater bargaining power while negotiating an illegal payment, or when the agent has weaker outside opportunities. Our analysis provides explanations why developed countries may have an advantage in dealing with the threat of extortion.

**JEL Classification:** D82, L23

**Key words:** Monitoring, Corruption; Collusion, Bribery, Extortion; Framing.

---

<sup>†</sup> Department of Economics, University of Washington, Seattle, WA 98195, [khalil@u.washington.edu](mailto:khalil@u.washington.edu)

<sup>‡</sup> Department of Economics, University of Washington, Seattle, WA 98195 and ECARES, Brussels [lawarree@u.washington.edu](mailto:lawarree@u.washington.edu)

<sup>\*</sup> Department of Economics, Hanyang University, Ansan, Korea, [uwyunsh@hanyang.ac.kr](mailto:uwyunsh@hanyang.ac.kr)

We would like to thank Jin Cao, Juan Carrillo, Anke Kessler, Ivan Png, Troy Scott, Steve Shavell and Jean Tirole for their help and suggestions. Devrim Yavus provided very useful research assistance.

## 1. Introduction

In the design of optimal organizations, the fight against corruption by enforcement officers relies on strong incentives to detect and report violations by agents. Such incentives raise the specter of extortion since rewards to deter bribery may act as inducements to engage in extortion. Consider the case of an enforcer whose role is to detect and report violations by an agent. Offering a reward to the enforcer for turning in the agent will lower his incentive to accept a bribe from that agent. For instance, a driver under the influence of alcohol may attempt to bribe a police officer to let him off the hook for a DUI conviction, but a corrupt officer will find it less profitable to accept a bribe if he can collect a reward when turning in the drunk driver.<sup>1</sup> Now consider the case of an officer catching drivers who run red lights. Again, a reward would lower his incentive to accept a bribe from a driver caught running the light, but the same reward may invite a corrupt officer to claim that the driver ran the light when he did not. Incentive to deter bribery may lead a corrupt officer to extort innocent drivers.

The goal of this paper is to analyze the distinct impact that extortion and bribery have on incentives and exploit those differences in the fight against corruption. In our model, bribery and extortion differ by their effect on an agent when corrupt enforcement agents manipulate evidence to extract money from the agent.<sup>2</sup> The enforcer can manipulate evidence in two different ways: (a) make a favorable report about the agent — this will be called bribery in this paper; (b) make an unfavorable report about the agent — this will be called extortion. We also use the generic term of corruption to describe bribery and extortion. Note that while bribery is mutually beneficial for the corrupt parties, extortion is antagonistic, with one party benefiting at the expense of the other. This difference plays an important role when analyzing the impact of corruption on *ex ante* incentives, and it should be kept in mind in the design of anti corruption measures.

---

<sup>1</sup> The reward can be non-monetary such as good reputation, promotion, etc. Similarly, bribes and extortion payments can take the form of favors to members in an organization.

<sup>2</sup> In the legal literature, there is a debate on the definitions of extortion and bribery based on who initiates the corrupt transaction. For example, Ayres (1997) argues that in an environment where corruption is endemic, an individual initiating a side-payment to an enforcement agent could well be the victim of extortion rather than someone attempting to engage in bribery. See also Lindgren (1993). We are able to abstract from this debate by focusing on whether the corrupt behavior helps or hurts the agent as we are mainly interested in optimal incentives for the agent.

Yet, too often, the popular debate does not distinguish between the two, treating them together as merely illegal or immoral payments to enforcers, which obfuscates fundamental issues.

The intuition that rewards to enforcement agents may also encourage extortion has not played much of a role in the literature on corruption in hierarchies that began with Tirole (1986).<sup>3</sup> In the received models, a supervisor or enforcement agent who can both be bribed and engage in extortion is not useful. Tirole (1986) showed that a corruptible supervisor can still be useful, but, his model and much of the subsequent literature did not feature the effect of extortion since extortion was not a credible threat in these models.

By highlighting the role of team work in forging information, we introduce a notion of soft information that allows us to capture the above tradeoff in a model of extortion in which the supervisor remains useful even when there is no external honest enforcement available. We derive two main results: (i) extortion should always be deterred but bribery should not; (ii) bribery is deterred when information is hard but may be allowed when information is soft. There is an extensive literature in economics dealing with bribery but our result that the threat of extortion makes bribery optimal is new.<sup>4</sup> We also find that the principal is better off when the agent has less bargaining power when negotiating a bribe, and that higher outside opportunities for the agent makes extortion less relevant.

The intuition for our result (i) depends on the fact there is a critical difference in the cost of providing incentives to the agent in the presence of bribery as compared to extortion. Even though both increase incentive cost, extortion discourages “good behavior” because the agent is subject to it even though he has done the right thing. Bribery, however, helps somewhat in providing incentive because it occurs when the agent is seen as violating the rules. The bribery payment acts as a penalty for “bad behavior”. This is in line with the less formal literature that suggests that bribes may

---

<sup>3</sup> There has been surprisingly little attention given to corruption in the economic theory of law enforcement. Shavell’s (2004) authoritative textbook on law and economics has no references to corruption. This was also noted by Polinsky and Shavell (2000) in their comprehensive survey, in which the first item on the agenda for future research is the study of incentives for enforcement agents and the fight against corruption.

<sup>4</sup> See the surveys by Tirole (1992) and Bardhan (1997), and references in Khalil and Lawarree (2006), or Silva et al. (2007) for recent contributions.

have some positive role to play but extortion does not (see Bardhan (1997)). Bribery can help “grease” the incentives in badly run organizations but, as Klitgaard (1988) noted, “Extortion is a particularly debilitating form of corruption.”... “It leads not only to inefficiencies but the alienation of citizens from their government.”

The above suggests extortion is worse than bribery, but it does not say why both should not be deterred. Indeed, in result (ii), we find that even if it is feasible to deter both, it is optimal to allow bribery when information is soft. Most of the existing literature, which relies on hard supervisory information, finds that deterring bribery is optimal.<sup>5</sup> Suppose, as in Tirole (1986) or Laffont and Tirole (1993), that the supervisor either finds hard evidence (positive or negative) or finds no conclusive evidence. With hard evidence, the supervisor can conceal information and pretend she has found no conclusive evidence but she cannot forge evidence. For example, in the case of DUI, a policeman may ignore a tainted blood sample, but cannot create one. It turns out that in this information structure, a threat of extortion is not credible.<sup>6</sup> Therefore, most of the literature following Tirole focused on the problem of bribery in models where extortion is not relevant.<sup>7</sup> Other than special circumstances, the literature largely finds that it is optimal to deter bribery.<sup>8</sup> Therefore, we contribute to this literature by pointing out that if information is soft, the threat of extortion may make it optimal to allow bribery.

In our model, where information is soft for the coalition (i.e., the coalition can forge evidence), the principal has to pay the supervisor a new reward to deter evidence

---

<sup>5</sup> Our focus is on the agency literature that followed the pioneering work by Tirole (1986, 1992) as opposed to the non-agency literature (as reviewed in Bardhan (1997)).

<sup>6</sup> The argument is explained in detail later. If the supervisor has no conclusive evidence, she has no discretion and no bribery or extortion can occur. Consequently, if she has positive evidence about the agent and wants to threaten to extort by concealing it, her threat is not credible. This is because she will not be rewarded if she reports no conclusive evidence.

<sup>7</sup> For instance in Kessler (2000) and Vafai (2005), the information is hard. Baliga (1999) analyzes the case of soft information but extortion does not increase the implementation costs because the mechanism of the game allows the agent to quit when faced with the possibility of extortion. See also Faure Grimaud, Laffont and Martimort (2003) for a model of soft information with asymmetric information between the supervisor and the agent. In Kofman and Lawarree (1993) the information structure allows forging of evidence but rules out extortion by assumption.

<sup>8</sup> Several papers have shown that it may be optimal to allow bribery because of restrictions on contracts. For instance, Kofman and Lawarree (1996), Acemoglu and Verdier (2000) and Auriol (2006) (uncertain auditor type); Che (1995) and Mookherjee and Png (1995) (auditor moral hazard); Strausz (1997a), Olsen and Torsvik (1998), Lambert-Mogiliansky (1998), and Khalil and Lawarree (2006) (renegotiation and no-commitment). See also Cadot (1987) or Carrillo (2000b) on the perverse effects of anti-corruption measures.

forging in addition to the reward for producing incriminating evidence. This new reward makes extortion credible. The tradeoff between bribery and extortion appears when information is soft, and we find that bribery occurs in equilibrium. Note that even though information is soft for the coalition, the principal is not powerless. It can exploit the difference in objectives between the agent and supervisor, which plays a key role in fighting extortion in the model. Recall that bribery is collusive and turns out to be more difficult to fight.

Our paper is also related to a growing literature that examines relationships between Information Technology (IT) adoption and incentives in organizations. Most of this literature has found evidence that harder information provided by IT allows the principal to provide stronger incentives.<sup>9</sup> This is consistent with our model as the hard-information based contracts are likely to be less susceptible to extortion and the agent's incentives can be made stronger and more efficient. An interesting extension would be to find out if police corruption has declined after the introduction of technologies such as video camera in patrol cars or red-light cameras. A recent study, lending support to this view, shows that red-light cameras help prevent traffic stops triggered by extortion motives based on racial considerations (Colb (2001)).

This is consistent with the fact that extortion is mainly a problem in less developed countries relying mostly on soft evidence, while in developed countries hard evidence is more common and it is mainly bribery that makes the news. In the financial world for instance, making information hard can take various forms and be represented by the use of institutions like lawyers, CPAs, auditors, bankruptcy courts, independent directors and legal actions by the shareholders (see the survey paper by La Porta et al. (2000)).

We consider extensions of the model and derive further results. Extortion is a less serious issue when the agent has less bargaining power or stronger outside opportunities. A lesser bargaining power hurts the agent as the supervisor can extract a larger bribe. The bribe is a more effective deterrent and the principal has to give a smaller reward to deter bribery. Since it was this reward that induced extortion, extortion is less of an

---

<sup>9</sup> For instance, Baker and Hubbard (2003) show how the adoption of on-board computers in trucks allowed trucking company to offer better incentive contracts.

issue. Better outside opportunities also make extortion less of an issue as they increase the agent's reservation utility and help protect the agent from the supervisor's extortion attempts. A higher reservation utility forces the principal to increase the risk-averse agent's wage while making it less dependent on the supervisor's report. We show that with strong enough outside opportunities, the agent's wage is independent of the supervisor's report unless it reveals shirking and extortion is no longer a threat for the agent. Again, this seems consistent with evidence that extortion is mainly a problem in less developed countries where agents have weaker outside opportunities.

Besides the non-agency literature reviewed in Bardhan (1997), there have been a few recent models of extortion in agency settings. These papers feature extortion in different settings and with a different focus than ours. Polinsky and Shavell (2001) study an optimal law enforcement problem. Acemoglu and Verdier (2000) study the choice between government intervention to address a market failure and the resulting bureaucratic corruption. Mookherjee (1997) and Hindriks et al. (1999) consider a tax-evasion model. Mookherjee focuses on reforms in public bureaucracies and Hindriks et al. on the redistributive properties of the tax scheme. To deter corruption, all four papers rely on the availability of incorruptible external enforcement agents and the penalties they can impose. Instead, we focus on internal mechanisms to deter bribery and extortion by developing an informational structure that makes a supervisor useful even though she can engage in bribery and extortion and incorruptible external enforcers are absent. In a procurement setting without external enforcement, Auriol (2006) allows for the possibility of extortion and bribery. However, unlike us, the extortion payment is only a redistribution between the agent and the supervisor causing no allocative inefficiency. Thus there are circumstances where extortion may be allowed since it is costly to deter, but bribery is only allowed when the corruptibility of the auditor is not known (Acemoglu and Verdier (2000), and Kofman and Lawarrée (1996)).

## 2. The Setup

We present a standard principal/supervisor/agent hierarchy with a key new feature that makes extortion relevant. The principal (it) is the owner of a firm, the agent (he) is the productive unit in the firm, and the supervisor (she) collects information for the principal. The agent produces an output  $x$  which depends on his level of effort,  $e \in \{0, 1\}$ . If the agent works, that is,  $e = 1$ , he produces  $x_H$  with probability  $\pi$  and  $x_L$  with probability  $1 - \pi$ , where  $x_H - x_L = \Delta x > 0$ , and  $\pi \in (0, 1)$ . If he shirks, that is,  $e = 0$ , he produces  $x_L$  with probability one.<sup>10</sup> While the level of output  $x$  is observed by all parties, the level of effort  $e$  is private information of the agent. The agent's disutility of effort is given by  $\varphi e$ , where  $\varphi > 0$ . The output belongs to the principal, who pays a transfer  $w$  to the agent. We assume that the agent is risk averse with a separable utility function given by  $U(w, e) = u(w) - \varphi e$ , where  $u$  is concave,  $u(0) = 0$ , and satisfies the Inada conditions ( $u'(0) = +\infty$  and  $u'(+\infty) = 0$ ). The principal who is risk-neutral offers a take-it-or-leave-it contract to the agent, who has zero reservation utility.<sup>11</sup> We assume that  $\Delta x$  is large enough that it is always profitable to induce the agent to work, that is, exert  $e = 1$ . The principal's objective is to minimize its expected cost of inducing  $e = 1$ .

In the absence of a supervisor, the contract for the agent could only be based on  $x$ , and the wages would be  $w_L$  when  $x_L$  is produced and  $w_H$  when  $x_H$  is produced.<sup>12</sup> In this model, the optimal contract in the absence of a supervisor — we refer to it as the *second-best contract* — requires that  $w_H^s = u^{-1}(\varphi/\pi)$  and  $w_L^s = 0$ . In other words, the principal compensates the agent only when there is definitive evidence that the agent worked, i.e., when  $x_H$  is realized. The agent does not obtain any rent.

The supervisor's role is to collect information about the agent's effort level and to report it to the principal. Since  $x_H$  can be realized only with  $e = 1$ , there is no reason to use the supervisor following  $x_H$ , and the principal will send the supervisor only when it observes  $x_L$ . Following Tirole (1986), we assume that the supervisor observes the true

---

<sup>10</sup> In section 5, we show that our main results are robust to a more general production function.

<sup>11</sup> We consider the case of a strictly positive reservation utility in section 5.

<sup>12</sup> We assume the principal itself does not have the expertise to monitor the agent. See Strausz (1997b) for a model of collusion comparing monitoring by the principal or delegation of monitoring to the supervisor.

level of effort with probability  $p$  or obtains no conclusive evidence with probability  $1 - p$ , where  $p \in (0, 1)$ . The supervisor's signal  $\sigma$  can take three values:  $\sigma \in \{0, \emptyset, 1\}$ , where  $\emptyset$  denotes that the supervisor does not have conclusive evidence about effort. Therefore, the agent is given a wage  $w_H$  following  $x_H$ , and  $w_r$ , following  $x_L$ , where  $r$  is the supervisor's report with  $r \in \{0, \emptyset, 1\}$ . We assume that the supervisor is costless but the principal may want to pay her a wage  $s$  to deter corruption.<sup>13</sup> The supervisor is risk neutral, and it is common knowledge that the supervisor is corruptible.<sup>14</sup> Without loss of generality, the wage to the supervisor depends only on her own report and is denoted by  $s_r$ . We assume that the supervisor's reservation utility is zero. Both the agent and supervisor are protected by limited liability such that  $w_r \geq 0$  and  $s_r \geq 0$ .<sup>15</sup>

### ***Supervision Technology and Corruption***

The supervisor is corrupt in the sense that she may not always report what she has observed to the principal. She will report the truth only if it is in her interest to do so. In this environment, we identify two types of corrupt behavior, which we define below. The key distinction is whether the corrupt behavior benefits the agent (the case of bribery) or hurts him (the case of extortion).

**Definition 1.** *Bribery* occurs when the supervisor accepts a payment in return for misreporting information to benefit the agent.

**Definition 2.** *Extortion* occurs when the supervisor obtains a payment from the agent by threatening to misreport information to hurt the agent. We say *framing* has occurred if the attempt at extortion fails and the supervisor misreports information that was favorable to the agent.

---

<sup>13</sup> We abstract from supervisor's moral hazard (costly effort) in order to focus on the tradeoff between extortion and bribery. Mookherjee and Png (1995) have shown that bribery may occur in equilibrium if the supervisor exerts an unobservable audit effort.

<sup>14</sup> As noted in the introduction, it may be optimal to allow bribery if the supervisor's corruptibility is private information (see, e.g., Kofman and Lawarrée (1996), Acemoglu and Verdier (2000) and Auriol (2006)). For a dynamic model where the supervisor privately knows her propensity for corruption, see Carrillo (2000a).

<sup>15</sup> Without limited liability, the first best could be reached since  $e = 0$  is off the equilibrium path. When the supervisor reports that  $e = 0$ , the principal can impose an infinite punishment on the agent, and also give a large reward to the supervisor if she is corruptible.

Bribery and extortion are accompanied by side-contracts between the supervisor and the agent whereas framing is not. With bribery, the supervisor and agent forge information to maximize their joint surplus. With extortion (resp. framing), the supervisor acts alone by threatening to suppress (resp. actually suppress) evidence since she is acting against the agent's interest.

We depart from the literature on monitoring that relies on hard information, which mainly captures the idea that it is relatively easy to conceal but very costly to forge information. In reality, there is often an asymmetry in the cost of forging information if the supervisor tries to do it alone or if she has help from the agent. In many circumstances the cost of forging can be significantly lowered with the help of the agent. Consider the previous example of the blood test taken after a car accident. If the police officer or the lab worker colludes with the driver, they can easily substitute another untainted blood sample. This means that information can be more easily manipulated when several people collaborate.<sup>16</sup> In the spirit of the recent literature on communication (Dewatripont-Tirole (2005) or Caillaud-Tirole (2007)), we emphasize that forging evidence, like information transmission, is a team activity and its cost depends on the amount of help from team members. Dewatripont and Tirole argue that the sender and the receiver, working together as a team, can make soft information hard. Because our focus is on fraud, we look at the opposite issue: can hard information be made soft? Information that is hard for the supervisor can become soft for the supervisor-agent coalition.<sup>17</sup>

To keep our model tractable, we incorporate the cost of forging evidence into our model in a binary way as follows: without the agent's cooperation the supervisor cannot forge information by herself (her cost is infinite) but can only conceal it. Her information is hard. If  $\sigma = e$ , she can only report  $r \in \{e, \emptyset\}$ , and if  $\sigma = \emptyset$ , the only possible report is  $r = \emptyset$ . Thus, extortion involves threatening to suppress information favorable to the agent. With the agent's cooperation, the supervisor can forge evidence at zero cost and

---

<sup>16</sup> In financial auditing for instance, the auditee can help the auditor draw "favorable samples."

<sup>17</sup> In our model, the principal (receiver) remains passive in the information transmission process, but the cost of forging depends on whether the supervisor (sender) acts alone or receives help from the agent.

report that the agent has worked regardless of what she observed, i.e., it is possible to have  $r \in \{0, \emptyset, 1\}$  regardless of  $\sigma$ . The information is soft for the coalition.<sup>18</sup>

It may seem counterintuitive that to make extortion by the supervisor relevant, information has to be soft for the *coalition* while it is hard for the supervisor. However, this assumption is critical because supervisory extortion would not be an issue if the information were either soft or hard. If the information were soft for the supervisor, the supervisor would be useless. If the information were hard for both the supervisor and the coalition, extortion would not be relevant.<sup>19</sup> This is because a threat of extortion is credible only if the supervisor is able to collect a reward by suppressing information. Since evidence cannot be forged, the supervisor has no discretion when  $\sigma = \emptyset$ , and there is no need to reward the supervisor when  $\sigma = \emptyset$ . Therefore, the threat of extortion by suppressing evidence is vacuous in a model with hard information as it is the case in many prominent models like Tirole (1986, 1992) or Kessler (2000).<sup>20</sup> In our model, it is the reward to deter forging of information when  $\sigma = \emptyset$ , since information is soft for the coalition, that makes the threat of extortion credible (even though information is hard for the supervisor).

Besides the standard assumption of enforceable side-contracts (see Tirole 1992), we need to make one additional assumption. Since bribery may occur in equilibrium, we need to be explicit in how side transfers are determined. We assume they are determined according to the Nash bargaining solution.<sup>21</sup> We require that extortion or framing be sequentially rational; the supervisor's threat of suppressing information is credible only if she receives a higher utility by suppressing evidence than by revealing it truthfully.

We summarize the model by presenting the timing of moves:

---

<sup>18</sup> We thus assume that it costs the coalition the same whether the misreport is from  $\emptyset$  to 1 or from 0 to 1, but this is not restrictive since misreporting after  $\sigma = 0$  can be deterred without cost in equilibrium. Similarly, assuming that it is equally inexpensive to suppress evidence from 0 or from 1 is not restrictive.

<sup>19</sup> Thus the relevant type of corruption is bribery, which is deterred in equilibrium.

<sup>20</sup> There is a series of papers by Vafai (cited in Vafai (2005)) analyzing extortion under hard information. To make extortion credible Vafai relies on the “prohibitive psychological or emotional cost” of not carrying out a threat and he shows that bribery can be deterred without cost.

<sup>21</sup> Note that the supervisor and the agent negotiate the side-contract under symmetric information. Another strand of the literature considers collusion under asymmetric information (see Laffont and Martimort (1997, 2000), or recently Che and Kim (2006)).

- (1) The principal offers a contract specifying the transfers to the agent as a function of output and the supervisor's report; and the transfers to the supervisor as a function of her report.
- (2) The agent and the supervisor accept/reject the contract.
- (3) The agent decides whether to work ( $e = 1$ ) or shirk ( $e = 0$ ).
- (4) Output  $x$  is realized. If the principal observes  $x_L$ , it sends the supervisor. If it observes  $x_H$ , the game moves to (8).
- (5) The supervisor and the agent observe the signal  $\sigma$ .
- (6) The supervisor and the agent choose whether or not to make a side-contract.
- (7) The supervisor makes a report  $r$ .
- (8) Transfers are realized.

### 3. Tradeoff between Bribery and Extortion

In this section we will argue that rewards to deter bribery will lead to extortion, but that it is feasible to deter both. In section 4, we show that it is optimal to allow bribery but not extortion. First, we briefly present the case where the supervisor is incorruptible.

If the supervisor were incorruptible, the optimal contract would specify that the supervisor will not be paid any reward,  $s_r = 0$ , for all  $r$ . The agent would only be rewarded when there is *definitive* evidence of effort, i.e., if  $x_H$  occurs or if  $x_L$  occurs but the supervisor finds evidence of work ( $r = 1$ ); the agent will be paid zero otherwise. The agent does not obtain any rent and he is equally compensated when  $x_H$  is realized and when  $r = 1$  with  $x_L$ , i.e.,  $w_H = w_1 > 0 = w_\emptyset = w_0$  (see appendix A for details of the *incorruptible-supervisor* contract). Compared with the second-best or no-supervisor case, the agent receives a positive wage more often, and therefore, his wage after  $x_H$  is smaller than under the second best. Given the effort  $e = 1$ , the agent obtains better insurance, and that reduces the principal's expected wage payment relative to the second-best contract.

This contract, however, is vulnerable to bribery. The supervisor is not being rewarded ( $s_r = 0$ ) since she is assumed to be truthful. The agent will bribe the supervisor when she finds no-evidence or evidence of shirking, and help her fabricate evidence to

give a report of work ( $r = 1$ ) so that they can share the higher wage  $w_1$  collected by the agent.

On first sight, this threat of bribery can be combated by introducing a reward for the supervisor when she reports shirking ( $r = 0$ ) or no-evidence ( $r = \emptyset$ ). If the reward is equal to  $w_l$  (i.e.,  $s_0 = s_\emptyset = w_l$ ), there will be no incentive to bribe. The supervisor is turned into a bounty hunter as in, e.g., Tirole (1986) or Kofman and Lawarrée (1993). However, in our framework, this would introduce a new problem of extortion by the supervisor. To see this, note first that  $s_1 = 0$  since there is no perceived threat of a bribe from the agent when  $\sigma = 1$ . Thus, when she has evidence of work, the supervisor will have an incentive to suppress this evidence to obtain the reward  $s_\emptyset > 0$  rather than get  $s_1 = 0$ .<sup>22</sup> This is the tradeoff mentioned in the introduction: strong incentives to deter bribes create scope for a new kind of corruption, namely extortion. As noted above, this tradeoff does not appear when information is hard.

Next we present the contract where the principal deters both bribery and extortion. However, we also show later that this contract is not optimal.

***The least-cost-corruption-proof (LCCP) contract: no bribery or extortion***

It is not clear a priori if it is optimal to deter all types of corruption. In particular, we have already shown that there is a tradeoff in deterring different kinds of corruption – rewards to deter bribery may encourage extortion. It is useful to characterize as a benchmark the least-cost-corruption-proof contract that deters both types of corrupt behavior. The LCCP contract is also a critical step when we derive the optimal contract in the next section. We show in Lemma 2 that the LCCP contract dominates any contract that allows extortion to occur in equilibrium. The main implication of deterring both bribery and extortion is that the principal loses much of the value of retaining a supervisor. It cannot fully utilize the information provided by the supervisor to differentiate the agent’s payments according to realized states. We show later that the LCCP contract is not optimal in general, but it can be under specific conditions, e.g., if

---

<sup>22</sup> Anticipating extortion the agent will refuse to put in high effort (his incentive constraint will be violated). Note also that raising  $s_1$  to  $s_\emptyset$  is problematic since it would encourage the coalition to report  $r = 1$  when  $\sigma = \emptyset$ .

the agent had all the bargaining power when negotiating the side-contract, and if the agent's outside opportunity is high enough (see Section 5).

Before presenting the principal's problem with its traditional incentive and participation constraints, we first need to consider the last stage with bribery and extortion. To prevent bribery the principal will have to ensure that the contract satisfies the Coalition Incentive Compatibility (*CIC*) constraints.

$$(CIC_{\sigma,r}) \quad T_{\sigma} \geq T_r, \quad \text{where } T_{\sigma} = w_{\sigma} + s_{\sigma} \quad T_r = w_r + s_r, \quad \text{for } \sigma, r \in \{0, \emptyset, 1\}.$$

We have six (*CIC*) constraints and these can be satisfied only when  $T_0 = T_{\emptyset} = T_1$ , i.e., the aggregate transfers in every state following  $x_L$  must be the same. This can also be written as:

$$w_0 + s_0 = w_1 + s_1, \quad \Rightarrow \quad s_0 = w_1 + s_1 - w_0 \quad (1)$$

$$w_{\emptyset} + s_{\emptyset} = w_1 + s_1, \quad \Rightarrow \quad s_{\emptyset} = w_1 + s_1 - w_{\emptyset} \quad (2)$$

Since extortion/framing may occur only by suppressing evidence when  $\sigma \in \{0, 1\}$ , the principal will have to ensure that the contract satisfies two additional extortion/framing deterring (*EF*) constraints to prevent extortion/framing. These can be written as:

$$(EF_1) \quad s_1 \geq s_{\emptyset},$$

$$(EF_0) \quad s_0 \geq s_{\emptyset}.$$

If one of the above constraints is not satisfied, the supervisor will choose to either extort or frame the agent, whichever gives her a higher payoff. Note however that only (*EF*<sub>1</sub>) is the relevant constraint for deterring extortion since it deters suppression of positive evidence. The constraint (*EF*<sub>0</sub>) deters suppression of negative information, and bribery is the pertinent issue. Therefore, we will ignore the (*EF*<sub>0</sub>) constraint and just verify *ex post* that it is satisfied by our identified solutions in each case below. We also assume that the agent and the supervisor do not collude when they are indifferent between colluding and not colluding, and the supervisor will not extort when she is indifferent.<sup>23</sup>

---

<sup>23</sup> This is a standard assumption that relies on the fact that the principal can always break the tie with a small extra payment.

Given the (CIC) and (EF) constraints the agent's participation and incentive constraints and the supervisor's participation constraint are the same as those in the incorruptible supervisor case discussed above.<sup>24</sup> Ignoring the participation constraint as it is implied by the limited liability and the incentive constraints, we present the principal's program – denoted by  $P^o$  – which prevents both bribery and extortion/framing:

$$\begin{aligned} & \text{Min } \pi(w_H) + (1 - \pi) [p(w_1 + s_1) + (1 - p)(w_\emptyset + s_\emptyset)] \\ \text{s.t. } & \text{(IC) } \pi u(w_H) + (1 - \pi) [pu(w_1) + (1 - p)u(w_\emptyset)] - \varphi \geq pu(w_0) + (1 - p)u(w_\emptyset) \\ & (1), (2), (EF_1), (EF_0), w_H \geq 0, w_r \geq 0 \text{ and } s_r \geq 0, \text{ where } r \in \{0, \emptyset, 1\} \end{aligned}$$

The solution to this problem is the *least-cost-corruption-proof contract* and it is characterized in the following lemma:

**Lemma 1** *The least-cost-corruption-proof (LCCP) contract has the following features:*

- (i) *If the supervisor's signal is not very accurate ( $p \leq \pi$ ), the contract is equivalent to the second-best or no-supervisor contract of section 3.*
- (ii) *If the supervisor's signal is accurate enough ( $p > \pi$ ), it is optimal to use the supervisor, and the contract to the agent satisfies:*

$$w_H^o > w_1^o = w_\emptyset^o > 0 = w_0^o,$$

where  $\frac{u'(w_1^o)}{u'(w_H^o)} = \frac{1 - \pi}{p - \pi}$ ,  $\pi u(w_H^o) + (p - \pi) u(w_1^o) = \varphi$ , i.e., the agent obtains an ex ante rent.

- The supervisor's contract involves:  $s_1^o = s_\emptyset^o = 0 < s_0^o = w_1^o$   
but the supervisor receives no ex ante rent.<sup>25</sup>
- The principal's expected cost is  $C^o = \pi(w_H^o) + (1 - \pi)w_1^o$ .

**Proof:** See Appendix B.

There are two main findings from this lemma: (a) the threat of extortion restricts the principal's ability to use the supervisor's information, and (b) the supervisor will be used only if she is accurate enough. We explain these below in turn.

<sup>24</sup> The participation constraint is given by  $\pi u(w_H) + (1 - \pi) [pu(w_1) + (1 - p)u(w_\emptyset)] - \varphi \geq 0$ .

<sup>25</sup> Since the agent does not shirk in equilibrium, the signal  $\sigma = 0$  is off the equilibrium path, and the supervisor's rent is zero even though  $s_0 > 0$ .

It is no longer possible to only reward the agent after definitive evidence of work, and the agent who shirks without being caught must also be treated as if he worked ( $w_{\emptyset} = w_I$ ). As we argued earlier, rewards for turning down bribes introduce incentive to extort/frame. In particular, a reward to the supervisor for reporting  $\sigma = \emptyset$  truthfully would encourage the supervisor to extort/frame when  $\sigma = I$ . This incentive is avoided by reducing  $s_{\emptyset}$  to zero, but then the (CIC) requires that  $w_{\emptyset} = w_I$ .

Since the supervisor is not perfectly accurate, the agent gets a high wage  $w_1$  ( $= w_{\emptyset}$ ) with probability  $1 - p$  even when he shirks, which implies that the supervisor may not be useful if she is not accurate enough. This is different from the case of the incorruptible supervisor where she is useful for any  $p > 0$ . If the agent works, he gets  $w_1$  with probability  $(1 - \pi)(p + (1 - p)) = 1 - \pi$ . The net effect on the (IC) can be seen by setting  $w_{\emptyset} = w_I$  and rearranging terms:

$$\pi u(w_H) + (p - \pi)u(w_1) = \varphi.$$

If  $p \leq \pi$ , the agent is more likely to receive the transfer  $w_1$  when he shirks rather than when he works, in which case it would be optimal to set  $w_1 = 0$ . We have  $w_1 = w_{\emptyset} = w_0 = 0$ , and the principal does not rely on the supervisor's report at all, and we also have  $s_r = 0$  for all  $r$ . Thus, the contract is equivalent to the second-best contract.

On the contrary, if  $p > \pi$ , paying a positive  $w_1$  is useful in providing incentive to the agent since he is more likely to receive a positive transfer when he works. However, this is costly to the principal since it also pays a positive  $w_{\emptyset}$  ( $= w_1$ ) and therefore it is optimal to set  $w^o_1 < w^o_H$ . The expected cost for the principal is smaller than under the second best, but higher than the case with an incorruptible supervisor.

Note that it is not the supervisor but the agent who benefits from the supervisor's ability to misreport information under the corruption-proof contract. The reason is as follows; the only way to prevent both bribery and extortion/framing is to give up the informativeness of  $r = \emptyset$  and treat it as if  $r = 1$  in shaping the agent's incentives. Thus the supervisor cannot affect the agent's payoff by misreporting that  $r = \emptyset$  when  $\sigma = 1$ . As a result, she cannot command any rent. The agent who is the potential victim, on the contrary, obtains a higher utility than his reservation level. Otherwise the agent will shirk

and get  $w_1$  ( $= w_\emptyset$ ) with probability  $1 - p$ . Specifically, lemma 1 establishes that  $\pi u(w_H^o) + (p - \pi) u(w_1^o) - \varphi = 0$  and therefore the rent  $\pi u(w_H^o) + (1 - \pi) u(w_1^o) - \varphi$  must be positive.

#### 4. The Optimal Contract: Bribery in Equilibrium

In this section we characterize the optimal contract when the supervisor can engage in both types of corruption. The principal has the fall-back option of offering the second-best or no-supervisor contract and ignore the supervisor's report, but we know that the least-cost corruption-proof contract dominates this contract when  $p > \pi$ , i.e., when she is accurate enough. Therefore, the interesting question is whether it is possible to improve upon the least-cost corruption-proof contract by allowing some type of corruption.<sup>26</sup>

Since we allow for the possibility of corruption to occur in equilibrium, we have to account for payoffs resulting from side contracts. We assume that when the agent and supervisor engage in a side contract, their payoffs are determined by the Nash bargaining solution. For example, if the agent bribes the supervisor to report work ( $r = 1$ ) when there is no evidence ( $\sigma = \emptyset$ ), the coalition will get  $s_1 + w_1$  which they will share. This implies that the agent's payoff when  $\sigma = \emptyset$  and  $r = 1$  is not  $w_1$ , but rather the outcome from Nash bargaining. Therefore, all the computations, and particularly the agent's (IC) constraint, have to be re-derived using the relevant Nash bargaining payoffs. They are presented in detail in the appendix and we only outline the main intuition here in the text. We first prove that extortion will never be allowed:

**Lemma 2:** *Any contract that induces  $e = 1$ , but violates (EF<sub>1</sub>) is strictly dominated by the least-cost corruption-proof contract.*

**Proof:** See Appendix C.

---

<sup>26</sup> Note that if it is possible to improve on the corruption-proof contract, it will be optimal to use the supervisor even when  $p < \pi$ , but for high enough  $p$ .

The intuition for never allowing extortion is that it appears as a penalty after the agent has exerted effort, which increases the cost of providing incentive. Technically (see Appendix C), this is seen from the outcome of the Nash bargaining between the agent and supervisor when  $(EF_1)$  is violated. If  $(EF_1)$  is violated, i.e., if the threat to report  $\emptyset$  when  $\sigma = 1$  is credible, we show that the agent gets the same payoff from the Nash bargaining whether the state is  $\emptyset$  or 1. Therefore, the supervisor's report is not useful in distinguishing between these states and the agent has less incentive to provide effort. As shown in our lemma 1, the least-cost corruption-proof contract does not distinguish between  $\emptyset$  and 1 either but it is less costly to the principal since the supervisor is not rewarded ( $s_1 = s_\emptyset = 0$ ). Therefore the least-cost corruption-proof contract dominates any contract that induces extortion.

We can now present our main result showing that allowing some bribery is indeed optimal, but allowing extortion is not, which is a novel result in the literature.

**Proposition 1:** *It is optimal to use the supervisor if  $p > \pi$ . If the agent does not have all the bargaining power, the optimal contract induces bribery when the signal  $\sigma = \emptyset$ , but it deters extortion and framing, and the optimal contract will have the following features:*

- $w_H^* > w_1^* > 0 = w_\emptyset^* = w_0^*$ ; when  $\sigma = \emptyset$ , the agent obtains  $kw_1^* > 0$ , where  $k < 1$  and  $k$  depends on the agent's relative bargaining power.<sup>27</sup>
- $s_1^* = s_\emptyset^* = 0 < s_0^* = w_1^*$ ; the supervisor obtains  $(1 - k)w_1^* > 0$  when  $\sigma = \emptyset$ .
- The principal's expected cost, denoted by  $C^*$ , is given by

$$C^* = \pi(w_H^*) + (1 - \pi)w_1^*.$$

**Proof:** See Appendix D.

The reason bribery may help is it provides an indirect way to create a variation in the agent's payoff when direct attempts by the principal would induce extortion. Note from our lemma 1 that the only way to deter all corruption is by not utilizing every piece

---

<sup>27</sup> In the Appendix, we define  $w_{1\emptyset}$  as the agent's payoff in state  $\sigma = \emptyset$  as a result of Nash bargaining and reporting  $r = 1$ , and thus  $k = w_{1\emptyset} / w_1^*$ .

of information provided by the supervisor. In particular, the principal can no longer pay the agent only after definitive evidence of work. The agent receives the same compensation when the signal is  $\emptyset$  and 1 even though the supervisor reports truthfully. This raises the cost of providing incentive to the agent since a shirking agent will also obtain a positive compensation when the signal is inconclusive about the true effort. A way to restore some variation in the agent's compensation between the states  $\emptyset$  and 1 is by allowing bribery to occur in state  $\emptyset$ . Suppose a bribe from the agent leads the supervisor to overstate performance in state  $\emptyset$  and report 1. Then the principal will make the same aggregate transfer in both states  $\emptyset$  and 1, but the agent's payoff in state  $\emptyset$  is lowered since he has to pay a bribe to the supervisor, and this lowers the cost of inducing high effort.<sup>28</sup>

The presence of bribery and forgery in equilibrium shows how corruption results in the violation of an important principle of incentives: the agent ought to be paid only when there is definite evidence of work. As reported in Proposition 1, even though the agent's wage is positive only when the principal receives evidence of work (after  $x_H$  or when  $r = 1$ ), the agent receives a positive payoff even when there was no actual evidence of work since forgery occurs in equilibrium. This weakens incentives since the agent can obtain this payoff even when he shirks – the null signal  $\emptyset$  can occur either when working or shirking.<sup>29</sup>

We now discuss why Tirole's bribery-proofness (or collusion-proofness in his terminology) principle fails. Tirole (1986 and 1992) shows that, under some circumstances, there is no loss of generality to derive an optimal contract that is bribery-proof. Since the principal can anticipate the side contracts between the agent and

---

<sup>28</sup> Polinsky and Shavell (2001) find that, depending on parameter values, it may be optimal to allow extortion/framing and deter bribery. Their model is very different from ours and relies on incorruptible external enforcers to detect corruption. More specifically, the principal can choose different probabilities of detecting bribery, framing, and extortion, and also choose different levels of sanctions for each offence. They also introduce another parameter  $\theta$  that determines how likely an innocent agent will be in a position to be framed. The relative values of these parameters may make it optimal to deter bribery and allow extortion/framing. For instance if the parameter  $\theta$  is very small, then allowing extortion/framing is not very costly, and the principal should focus on deterring bribery.

<sup>29</sup> The implication of the above is that  $w_H^* > w_1^*$ , which is different from the case when the supervisor is honest, where  $w_H = w_1$  since both wages were unambiguously indicative of high effort. Allowing bribery entails paying  $w_1$  even when  $\sigma = \emptyset$ , which means that the principal has to pay  $w_1$  with a positive probability even when the agent shirks. Consequently,  $w_1$  is lowered from  $w_H$  in the optimal contract.

supervisor, adequate incentives can be provided not to collude by replicating the payoffs associated with side contracts. However, bribery may occur in equilibrium due to what Tirole has referred to as non-separabilities in the constraints that deter corruption (section 2.5, Tirole 1992). When these constraints are interlinked, satisfying one constraint raises the cost of satisfying another one and it may be too costly to satisfy them all.

In our case it is the interaction between the collusion (*CIC*) and extortion (*EF*) constraints that causes the collusion-proofness principle to fail. To prevent forging of evidence in state  $\sigma = \emptyset$ , and reporting  $r = 1$ , the principal has to increase the reward  $s_\emptyset$ , but this increases the cost of deterring extortion in state  $\sigma = 1$  since the principal has to maintain  $s_1 \geq s_\emptyset$ .<sup>30</sup> As argued above and in the LCCP contract, the only way to prevent both forms of misreporting is to require  $w_1 = w_\emptyset$ , which is very costly in terms of providing incentive to the agent. With such interlinked-constraints, we show that it is cheaper to allow collusion than to fight it. Bribery allows the principal to create a variation in the agent’s payoffs without inducing extortion.

This captures nicely an intuition often mentioned in the applied literature, that allowing bribery can create markets that improves incentives (Bardhan (1997)). Here, the principal relies on the supervisor to extract a bribe from the agent and lower the agent's payoff in state  $\emptyset$ , when it cannot directly do so for fear of encouraging extortion. The latter is also consistent with the widely held belief that extortion is always counter productive since it penalizes agents when they have obeyed rules or done what they are supposed to. Extortion punishes the agent when he has done the “right thing”, while bribery increases the cost of shirking or violating rules.

## 5. Extensions

### 5.1. *Agent’s bargaining power hurts the principal*

When bribery is deterred, the bargaining power of the coalition members does not matter. The principal competes with the agent for the supervisor’s report and the reward given to

---

<sup>30</sup> In state  $\sigma = \emptyset$ , the principal needs to satisfy  $s_\emptyset \geq s_1 + w_1 - w_\emptyset \geq 0$ , which increases the cost of deterring extortion in state  $\sigma = 1$ .

the supervisor must exceed any viable offer from the agent. In our model the bargaining power is relevant since the principal lets bribery occur in equilibrium. In this section, we show that the principal is better off when the supervisor has relatively more bargaining power. The reason is that the supervisor can extract a larger bribe from the agent, which makes the bribe a more effective penalty and allows the principal to improve incentives. Although the principal cannot affect the relative bargaining power by its choice of contract, the contract influences the bargaining outcomes since it determines the threat points and the pie to be shared.

The principal would like to implement a wage differential based on realized states to provide incentive to the agent, which is the agent's stake in bribery. A reward to deter bribery raises the problem of extortion. Hence, the principal implements a *payoff* differential for the agent by inducing bribery, which acts as a penalty on the agent. The agent's bargaining power hinders the principal's ability to use the bribe as a penalty. If the agent had no bargaining power, the bribe would be equal to the stake of bribery, the wage difference, and the threat of extortion would not add any cost in providing incentive. On the other hand, if the agent has all the bargaining power, a bribe is useless in generating a payoff difference since the bribe would be zero or negligible. Then, the principal may as well deter both forms of corruption since it does not gain from inducing bribery (the LCCP contract is optimal).

To see the precise argument, recall from the incorruptible supervisor benchmark that the principal would prefer to make the agent's payoff zero in state  $\emptyset$ . This is because the state  $\emptyset$  is relatively more likely to occur when the agent shirks compared to when he works. In the optimal contract, the agent earns a positive return  $k w_1$  from Nash bargaining in state  $\emptyset$ . As the agent's bargaining power goes down, he earns a smaller return in state  $\emptyset$ , which implies that the *(IC)* becomes slack and this allows the principal to increase its payoff by adjusting the transfers.

If the agent's bargaining power is reduced down to zero, we can argue that extortion would not impose additional cost on the principal as his payoff is identical to what it would have been in the hypothetical case where extortion could be deterred at

zero cost.<sup>31</sup> As the agent's bargaining power goes down, the agent retains a smaller and smaller share of  $w_1$  in state  $\emptyset$  as part of his Nash bargaining outcome. When his bargaining power is zero, his share of  $w_1$  is also zero and the entire  $w_1$  is taken by the supervisor as a bribe and the agent is left with a zero payoff in state  $\emptyset$ . In the hypothetical case where extortion could be deterred at zero cost, the principal does not have to worry about extortion by assumption and can deter bribery by paying  $s_\emptyset = w_1$ . There would be no difference between the optimal contract where the agent has zero bargaining power and the optimal contract if extortion could be deterred at zero cost. Thus we conclude that the threat of extortion introduces additional cost on the principal only if the agent has bargaining power.

At the other extreme, if the agent has all the bargaining power, allowing bribery in equilibrium has no deterrent effect since the agent gets the entire  $w_1$  when they misreport. Therefore, the bribe does not create a variation in the agent's payoff, the *raison d'être* of allowing bribery in the first place. If the agent has all the bargaining power, the principal's payoff is identical to its payoff under the LCCP contract where  $w_1 = w_\emptyset$ . The principal does not gain by allowing bribery, and is as well off as it deters all forms of corruption. Our findings are summarized in proposition 2:

**Proposition 2:** (i) *The principal's payoff increases with the supervisor's bargaining power.* (ii) *At the limit, if the supervisor has all the bargaining power, the principal's payoff is identical to the case where extortion could be deterred at zero cost.* (iii) *At the other limit, if the agent has all the bargaining power, the principal's payoff is identical to the payoff under the least-cost-corruption-proof (LCCP) contract.*

**Proof:** See Appendix E.

## ***5.2 Better outside opportunities make extortion less relevant***

Previously we suggested that more developed counties can rely more intensively on hard evidence and therefore suffer less from extortion. In this section, we provide another

---

<sup>31</sup> This would be the case if, for example, an agent threatened with extortion may avail himself of an efficient appeals process.

possible explanation why extortion is less of a problem in more developed countries. We show that if the agent has better outside opportunities, he is less likely to be the target of extortion. The reason is that the wage of an agent with better outside opportunities has to be raised to satisfy the higher reservation utility. With a risk averse agent, the most efficient way to increase his expected utility is by reducing the variation in the wages on the equilibrium path and relying on the low wage off the equilibrium path to provide incentives. This implies that the agent's wage when the supervisor has no evidence ( $w_{\emptyset}$ ) increases relatively more than the wages in the other states. Intuitively, a risk averse agent with better outside opportunities is less likely to accept a contract in which he may be punished even though he has worked hard.

For a high enough reservation utility, we show that the agent's wage is made independent of the supervisor's report as long as this report does not reveal shirking ( $r = 0$ ). If the supervisor reveals shirking, the agent is punished with a zero wage. This sanction is relatively more severe when the outside opportunities are high. This could be an explanation for why developing countries with weaker outside opportunities for their workers may suffer more from extortion. Our result is also consistent with the argument that economic agents such as bureaucrats with high salaries are less susceptible to corruption. Often such a claim relies on the decreasing marginal utility of income or an efficiency-wage argument. Our argument is different. In our model, as outside opportunities grow, the agent's wage increases but his rent does not. The supervisor's report can be used to reduce the agent's exposure to risk, provided he works, and extortion becomes less of an issue at the same time. We summarize our result in the proposition below.

**Proposition 3:** *If the agent's reservation utility is high enough, extortion is not a relevant issue for the principal.*

**Proof:** See Appendix F.

Technically, we show in the Appendix F that the optimal contract derived by only deterring bribery also deters extortion when the reservation utility is high enough. The reason is that an increase in the agent's reservation utility forces an increase in  $w_{\emptyset}$  in

order satisfy the (*IR*) constraint. However, such an increase would violate the (*IC*) unless  $w_H$  and  $w_1$  are increased as well. The (*CICs*) require the same total payments in each state so the principal gains by not increasing  $w_1$  at the same rate as  $w_\emptyset$  because by doing so it can decrease the reward  $s_\emptyset$ . For a high enough reservation utility, we obtain  $w_\emptyset = w_1$ , which implies that  $s_\emptyset = 0 = s_1$  and extortion ceases to be a relevant threat. The optimal contract is therefore similar to the LCCP contract.

Of course, if the reservation utility is increased further, the wages  $w_\emptyset = w_1$  are increased to the point where  $w_\emptyset = w_1 = w_H$  and the first best is reached. The threat of a large penalty ( $w_0 = 0$ ) if the agent is found shirking is enough to provide the agent an incentive to work.

### ***5.3. Generalizing the production technology: possibility of success after low effort***

One simplifying assumption of our model was that low effort always yielded a low output. In this section we consider the more general case where low effort can also yield a high output, which corresponds to a situation where the agent can get lucky, and we show that our main results generalize. The main findings are that extortion remains a threat after low output, but it is not relevant after high output. When output is low, bribery is allowed and extortion is deterred, but when output is high, both bribery and extortion are deterred.

We outline the extended model and the intuition before presenting the technical details. Suppose the likelihood of producing the high output is  $\pi_1$  when  $e = 1$ , and it is  $\pi_0$  when  $e = 0$ , where  $\Delta\pi = \pi_1 - \pi_0 > 0$ . The payments to the agent and supervisor will depend on the output and the supervisor's report, and they are denoted by  $w_r^j$ , and  $s_r^j$ , where  $j = L, H$ , for the two output levels, and  $r = 0, \emptyset, 1$  are the supervisor's reports.

To grasp the intuition, recall first that so far a high output was an absolute guarantee of high effort, but now a high output could result from a low effort by a lucky agent. Therefore, the principal will want to send the supervisor even after high output. The high output is more likely after a high effort than a low effort. Therefore, given a null signal  $\emptyset$ , it is more likely that a high effort was exerted when the output is high

compared to when the output is low.<sup>32</sup> Consequently, raising the wage  $w_{\emptyset}^H$  (after high output and null report) helps incentives, whereas raising the wage  $w_{\emptyset}^L$  (after low output and null report) hurts incentives. Thus, when facing the threat of bribery, the principal deters bribery by raising  $w_{\emptyset}^H$  all the way to  $w_1^H$  and removes the stake of bribery. This way of fighting bribery does not induce a threat of extortion unlike providing a reward to the supervisor. However, after low output, the principal cannot increase  $w_{\emptyset}^L$  as it would have a negative incentive effect. The alternative method of fighting bribery, a reward to the supervisor, would introduce a threat of extortion as in our main model. Thus, the principal finds it optimal to allow bribery after low output, and we find that our main result generalizes – a fear of inducing extortion can make bribery optimal.

It is instructive to study the agent's incentive constraint if the supervisor were incorruptible. It is given by,

$$(IC) \quad \pi_1 [pu(w_1^H) + (1-p)u(w_{\emptyset}^H)] + (1-\pi_1) [pu(w_1^L) + (1-p)u(w_{\emptyset}^L)] - \varphi \geq \\ \pi_0 [pu(w_0^H) + (1-p)u(w_{\emptyset}^H)] + (1-\pi_0) [pu(w_0^L) + (1-p)u(w_{\emptyset}^L)],$$

which, after rearranging becomes,

$$\pi_1 pu(w_1^H) + \Delta\pi(1-p)u(w_{\emptyset}^H) - \pi_0 pu(w_0^H) + \\ (1-\pi_1)pu(w_1^L) - \Delta\pi(1-p)u(w_{\emptyset}^L) - (1-\pi_0)pu(w_0^L) \geq \varphi$$

The main points of interest are the two wages following the signal  $\emptyset$ , when the supervisor finds no conclusive evidence of effort. It is immediate that the  $w_{\emptyset}^H$  helps incentives (positive coefficient), while  $w_{\emptyset}^L$  hurts incentives (negative coefficient). Therefore, the principal prefers to have a positive  $w_{\emptyset}^H$  but would like to set  $w_{\emptyset}^L = 0$ . The complete contract when the supervisor is incorruptible is presented in Appendix G.1.

Now consider the case where the supervisor may accept a bribe, but extortion is detected at zero cost. Coalitional incentive constraints would imply that the total transfers to the coalition ( $s + w$ ) is constant given the output level as in our main model. Given an output, the principal makes the same total payment regardless of the supervisor's report. Therefore, the principal's incentive to set  $w_{\emptyset}^j$  is be entirely driven by

---

<sup>32</sup> We assume that the null signal is equally likely after a high output or low output.

the (IC). After high output, the principal fights bribery by removing the stake of a bribe ( $w_1^H = w_\emptyset^H > 0$ ), while after low output, it fights bribery by rewarding the supervisor ( $w_1^L = s_\emptyset^L > 0 = w_\emptyset^L$ ) as in our main model. Therefore, it is only after low output that extortion could become an issue if it could not be detected. The details of this contract are given in Appendix G.2.

When extortion cannot be detected, it is straightforward to derive the optimal contract using arguments similar to those to prove proposition 1. We show that our result generalizes to this case where the agent can be lucky after shirking. A threat of extortion can make bribery optimal – the principal finds it optimal to allow bribery when the supervisor finds no conclusive evidence after low output. These results are summarized in the following proposition.

**Proposition 4.** *If the agent can also produce high output with low effort and it is optimal to use the supervisor, then bribery is allowed after low output but deterred after high output; extortion is always deterred.*

**Proof:** The complete proof is available from the authors.

## 6. Conclusion

This paper builds on a key intuition that has not played much of a role in the literature on corruption in hierarchies: rewards to enforcement agents to turn down bribes may also encourage them to engage in extortion. In his seminal paper, Tirole (1986) showed that a corruptible supervisor can still be useful, but his model and much of the subsequent literature did not feature the effect of extortion since extortion was not a credible threat in these models. Highlighting the team aspect of forging information, we introduce an appropriate notion of soft information. This allows us to present a model of extortion in which the supervisor remains useful even when there is no external honest enforcement available.

This tradeoff creates an interlinking of the bribery and extortion constraints in the principal’s maximization problem and causes a failure of Tirole’s collusion-proofness principle. Our main contribution is to show that bribery may be optimal due to the threat

of extortion.<sup>33</sup> It is important to underline that the tradeoff only appears if information is soft. If information is hard, there is no such tradeoff and bribery does not occur in equilibrium. Our results suggest that organizations that must rely on soft information may also need to allow bribery. By making its information “harder” an organization will suffer less from corruption, but making information harder can be costly. For instance, speeding tickets should rely on sophisticated cameras or shareholders ought to be able to appeal auditing reports to reliable and incorruptible experts. Developing countries with less resources and technological abilities, and weak legal environment also have less capability to make information hard and, therefore, we should expect that bribery to be a more pervasive problem. Again the reason is that they do not have the ability to rely on hard information. The fight against corruption should therefore focus on the reliance on hard evidence.

One implication of bribery occurring in equilibrium is to validate in a model the popular notion that bribery can be useful to “grease the wheels” in inefficient organizations. However, this is only a second-best result – bribery is optimal in our model because it allows the principal to cause a variation in the agent’s payoffs when direct payments would only have resulted in introducing extortion, which is a worse problem. Extortion penalizes an agent after “good” behavior, while bribery at least imposes some penalty for “bad” behavior.

Our analysis provides a ranking of different forms of corruption. It demonstrates the significance of relying on hard information and of the availability of honest external enforcement. There is a difference between bribery and extortion since the former relies on cooperation but not the latter. Thus, bribery would not be reported other than by whistleblowers, but extortion may be relatively easier to deter using an appeals process for agents subject to extortion. Of course, if extortion can be deterred using an (honest) appeals process, the principal would be able to deter both forms of corruption. We can make some observations based on our analysis. The expected penalty on the supervisor

---

<sup>33</sup> While there are many reported examples of explicit bribery in the media, an interesting example of allowing collusion/bribery in organizations is a leniency bias in job performance appraisal. Our result provides one rationale for why many organizations which use job performance appraisal as an incentive device may allow a leniency bias. See Bretz et. al (1992) for a survey on studies related to this issue, and Johnson and Liebcap (1989) for an example of leniency in the federal government.

from a successful appeal would allow the principal to create a variation in the wages  $w_1$  and  $w_0$  without inducing extortion. This variation lowers the cost of incentives. Without an appeals process, the principal has to induce bribery to create the desired variation in the agent's payoffs. Thus, the principal will prefer to deter both forms of corruption if the appeals process can provide a large enough expected penalty. Otherwise, both types of corruption will be deterred only if the agent's outside option is high, or if the agent's bargaining strength is high.<sup>34</sup> Neither is likely to be true in a poor country with a weak appeals process. Also, detection of extortion is usually not perfect because extortion reports may be seen as malevolent, for example.<sup>35</sup> It is well known that policing the police is not an easy task, and incorruptible enforcement agents may be scarce and expensive in many contexts. Thus the threat of extortion is likely to be more of a problem in poorer countries and ought to be thwarted, but perhaps at the cost of tolerating some bribery.

---

<sup>34</sup> As seen in section 5.2, if the agent's outside option is high the variation in the two wages is low. We saw in section 5.1, if the agent's bargaining strength is high, bribery is not very effective in creating a variation in wages.

<sup>35</sup> Furnivall (1956) studying bribery and extortion in Burma noted, "Those who gained their ends by bribery naturally made no complaint, and complaints from those who suffered were suspect as malicious. Such evidence as was available mostly came from people who had given bribes and, as accomplices, their evidence, even if admissible, was doubtful. It was difficult and dangerous for any private individual to set the law in motion, and in practice this was hardly possible except by some local or departmental superior of the man suspected of corruption." Klitgaard (1988) discussing tax assessor extortion noted that the appeal process is not straightforward: "In one of the most notorious versions [of extortion] a tax assessor would slap an unrealistically high assessment on the taxpayer. The taxpayer could appeal, but that would take time and effort; furthermore, the taxpayer might not be sure what the 'correct' tax really was."

## Appendices

### Appendix A Incorruptible Supervisor

Suppose the supervisor always reports truthfully what he has observed. The agent's participation  $[\pi u(w_H) + (1 - \pi) [pu(w_1) + (1 - p) u(w_\emptyset)] - \varphi \geq 0]$  and incentive constraints (IC)  $[\pi u(w_H) + (1 - \pi) [pu(w_1) + (1 - p) u(w_\emptyset)] - \varphi \geq pu(w_0) + (1 - p) u(w_\emptyset)]$  are standard. The incentive constraint (IC) can be rewritten as  $\pi u(w_H) + (1 - \pi) pu(w_1) - \pi(1 - p) u(w_\emptyset) - pu(w_0) \geq \varphi$ . Given limited liability, and since zero effort entails zero cost, the incentive constraint will imply that the participation constraint is satisfied in each of the cases we consider. The supervisor's participation constraint is also satisfied due to limited liability. Thus, we ignore both the agent's and the supervisor's participation constraints.

The principal's program when the supervisor is truthful,  $P^I$ , can be written as follows:  $\text{Min } \pi(w_H) + (1 - \pi) [p(w_1 + s_1) + (1 - p) (w_\emptyset + s_\emptyset)]$  s.t. (IC),  $w_H \geq 0$ ,  $w_r \geq 0$  and  $s_r \geq 0$ , where  $r \in \{0, \emptyset, 1\}$ . The optimal levels of transfers are obtained using standard techniques.

### Appendix B Proof of Lemma 1

In the problem  $P^o$  of section 4, we will first ignore the constraint ( $EF_0$ ) and verify later that it is satisfied by the optimal contract. Using (2) to replace  $s_\emptyset$  everywhere, we can rewrite ( $EF_1$ ) as ( $EF_1^b$ ):

$$(EF_1^b) \quad w_\emptyset \geq w_1,$$

The principal's problem is to  $\text{Min } \pi w_H + (1 - \pi) (w_1 + s_1)$  s.t. (IC), ( $EF_1^b$ ), (1) and the non-negativity constraints. Note that once we ignore ( $EF_0$ ), the variable  $s_0$  does not appear anywhere else in the problem except in (1). Therefore, we are free to choose  $s_0$  to satisfy this constraint (1) as long as  $s_0 \geq 0$ . Again using standard techniques, we can derive the optimal levels of transfers reported in lemma 1. ■

### Appendix C Proof of Lemma 2

We proceed in steps. First, we show that the agent receives the same *payoff* from Nash bargaining for  $\sigma \in \{\emptyset, 1\}$  if the constraint ( $EF_1$ ) is violated, but the supervisor earns an *ex ante* rent. We then show that there exists a corruption-proof contract that achieves the same cost but is more costly than the least-cost corruption-proof contract. This proves the claim. Note that the least-cost corruption-proof contract is strictly better since it also pays the agent the same *wage* for  $\sigma \in \{\emptyset, 1\}$  but the supervisor earns no *ex ante* rent.<sup>36</sup>

---

<sup>36</sup> Although we rely on the axiomatic approach, our bargaining outcome can be related to the outcome of a strategic alternative-offers-bargaining model with a risk of breakdown, where time between offers are very small (see e.g., Binmore et al. (1986) or Osborne and Rubinstein (1990)). In our model, the bargaining is about negotiating an illegal side contract which is fraught with uncertainty, e.g., opportunities to interact may disappear abruptly or the principal may require an early report. Therefore, the exogenous risk of a breakdown in negotiations will be the dominant force that drives the parties to an agreement. Moreover, since the principal will set a short deadline for a report from the supervisor, discount rates play a minor role compared to the fear of a breakdown. In such a case, it is appropriate to choose the breakdown point as the disagreement or threat point in the Nash bargaining solution as we have done in the paper (See Binmore et

(i) If  $(EF_1)$  is violated, i.e.,  $s_1 < s_\emptyset$ , then the agent gets identical payoffs for  $\sigma = \emptyset$  or  $\sigma = 1$ ; the same is true for the supervisor. Define  $T_k$ :  $T_k = w_k + s_k$  for  $k = \{0, \emptyset, 1\}$ , and define  $m$  by  $T_m = \max \{T_0, T_\emptyset, T_1\}$ . Then define  $w_{r\sigma}$  and  $s_{r\sigma}$  as the agent and the supervisor's respective payoffs (from Nash bargaining where relevant) when the signal is  $\sigma$  and the supervisor reports  $r$ .

(a) If  $T_m = T_\emptyset$ : Given  $s_1 < s_\emptyset$ , the supervisor will report  $r = \emptyset$  when  $\sigma = \{\emptyset, 1\}$ , and the agent will not find it profitable to bribe the supervisor into announcing  $r = 1$ . Therefore, payoffs will be:  $w_{m1} = w_{m\emptyset} = w_\emptyset$ ;  $s_{m1} = s_{m\emptyset} = s_\emptyset$ .

(b) If  $T_m > T_\emptyset$ : The supervisor reports  $r = m$  and the coalition receives  $T_m$  for  $\sigma = \{\emptyset, 1\}$ . Their payoffs are given by Nash bargaining. Since only the supervisor reports, the threat point is  $r = \emptyset$  for  $\sigma \in \{\emptyset, 1\}$  since  $s_1 < s_\emptyset$ . The bargaining problem is given by

$$\begin{aligned} & \max_{w,s} (u(w) - u(w_\emptyset))^\alpha (s - s_\emptyset)^{1-\alpha} \\ & s.t. \quad w + s = T_m, \end{aligned}$$

where  $\alpha \in (0, 1)$  is the agent's bargaining power.<sup>37</sup>

The solution is denoted by  $w_{m\sigma}$  and  $s_{m\sigma}$  for  $\sigma \in \{\emptyset, 1\}$ . Since the bargaining set and the threat point remain unchanged whether  $\sigma = \emptyset$  or 1, their respective payoffs must also remain unchanged. They are:  $w_{m1} = w_{m\emptyset}$ ;  $s_{m1} = s_{m\emptyset} > 0$  since  $s_\emptyset > s_1 \geq 0$ .

Therefore, from (a) and (b), we have proved that  $w_{m1} = w_{m\emptyset}$  regardless of  $m$ .

(ii) Expected cost of any contract that induces  $e = 1$  but violates  $(EF_1)$ .

Consider the contract denoted by  $\{\hat{w}_H, \hat{w}_r, \hat{s}_r\}$  that induces  $e = 1$ , but violates  $(EF_1)$ ,  $\hat{s}_\emptyset > \hat{s}_1$ . Then the expected cost is:  $\pi (\hat{w}_H) + (1 - \pi) (\hat{T}_m)$  where  $\hat{T}_m = \max \{\hat{T}_0, \hat{T}_\emptyset, \hat{T}_1\}$ , and  $\{\hat{w}_H, \hat{w}_r, \hat{s}_r\}$  satisfy the (IC) constraint:

$$(IC) \quad \pi u(\hat{w}_H) + (1 - \pi) \{p u(\hat{w}_{m1}) + (1 - p) u(\hat{w}_{m\emptyset})\} - \varphi \geq p u(\hat{w}_{m0}) + (1 - p) u(\hat{w}_{m\emptyset}).$$

Define  $\hat{W}_m = \hat{w}_{m1} = \hat{w}_{m\emptyset}$ ,  $\hat{S}_m = \hat{s}_{m1} = \hat{s}_{m\emptyset}$  and simplify (IC):<sup>38</sup>

$$(IC) \quad \pi u(\hat{w}_H) + (p - \pi) u(\hat{W}_m) - \varphi \geq p u(\hat{w}_{m0})$$

Note that  $\hat{S}_m > 0$  since the supervisor receives at least  $\hat{s}_\emptyset$  from Nash bargaining and  $\hat{s}_\emptyset > \hat{s}_1 \geq 0$ .

---

al. (1986) p.183 or Osborne and Rubinstein (1990) p.88).. Finally note that the outside option which acts as a constraint on the bargaining set is not binding in this model. The supervisor will have to make a report even when the parties fail to agree and, therefore, the outside option and the break down point are identical.

<sup>37</sup> In the strategic alternative offers bargaining model, values of  $\alpha \neq \frac{1}{2}$  can still be justified by asymmetry in the bargaining procedure or in the beliefs about the likelihood of breakdown (Binmore et al. (1986), p. 187).

<sup>38</sup> Note that  $s_0$  could be larger or smaller than  $s_\emptyset$  – both cases are captured in  $\hat{w}_{m0}$ .

(iii) Implement  $e = 1$  with a (constructed) corruption-proof contract  $\{w'_H, w'_r, s'_r\}$  that has the same expected cost as  $\{\hat{w}_H, \hat{w}_r, \hat{s}_r\}$ . Construct  $\{w'_H, w'_r, s'_r\}$  by defining:  $w'_H = \hat{w}_H$ ,  $w'_1 = w'_\emptyset = \hat{W}_m$ ,  $w'_0 = 0$ ,  $s'_1 = s'_\emptyset = \hat{S}_m$ , and  $s'_0 = \hat{T}_m$ .

Check that  $\{w'_H, w'_r, s'_r\}$  is indeed corruption-proof and implements  $e = 1$ :

(CIC) is satisfied since  $w'_k + s'_k = \hat{T}_m$ ,  $k \in \{0, \emptyset, 1\}$ , ( $EF_k$ ) is satisfied since  $s'_k \geq s'_\emptyset$ ,  $k \in \{0, 1\}$ , and (IC) is satisfied since  $w'_k$  must satisfy (IC) given that  $\hat{w}_k$  satisfies (IC) where  $k \in \{H, m0, m\emptyset, m1\}$  and given that  $w'_0 \leq \hat{w}_{m\emptyset}$ .

Finally, note that  $\{w'_H, w'_r, s'_r\}$  is not the least-cost corruption-proof contract since  $\hat{S}_m > 0$ , whereas in least-cost corruption-proof contract  $s_1^0 = s_\emptyset^0 = 0$ . Therefore, the least-cost opportunity-proof contract strictly dominates both  $\{w'_H, w'_r, s'_r\}$  and  $\{\hat{w}_H, \hat{w}_r, \hat{s}_r\}$ . ■

## Appendix D Proof of the Proposition 1

The agent-supervisor coalition will choose the report to maximize their joint payoff, which will be  $T_m$ . Note that since we do not impose (CIC) constraints bribery may potentially occur. Then the objective function becomes  $\pi w_H + (1 - \pi) T_m$

From lemma 2 we know that the ( $EF_1$ ) must be satisfied: ( $EF_1$ )  $s_1 \geq s_\emptyset$ .

The (IC) constraint is:  $\pi u(w_H) + (1 - \pi) p u(w_{m1}) - \pi (1 - p) u(w_{m\emptyset}) - p u(w_{m0}) - \phi \geq 0$ ,

where  $w_{r,\sigma}$  denotes the agents payoff from Nash bargaining when the report is  $r$  and the signal is  $\sigma$ . We ignore the constraint ( $EF_0$ ) for now and verify later that it is indeed satisfied by the optimal contract.

We consider three cases depending on whether  $m = 1, \emptyset$ , or  $0$  respectively, and show that case I is optimal.

Case I:  $T_m = T_1$

Min  $\pi w_H + (1 - \pi) T_1$  s.t.

$$(IC) \quad \pi u(w_H) + (1 - \pi) p u(w_1) - \pi (1 - p) u(w_{1\emptyset}) - p u(w_{10}) - \phi \geq 0$$

$$(EF_1) \quad s_1 \geq s_\emptyset$$

We make some observations to simplify the optimization problem.

(a) Note that  $w_{m1} = w_1$  because  $s_1 \geq s_\emptyset$  and  $T_m = T_1$ . The Nash Bargaining Solution (NBS) implies that  $s_{11} = s_1$ , and  $w_{11} = w_1$ .

(b)  $T_0 = T_1$  and  $w_0 = 0$ : To see this, note that  $w_0$  and  $s_0$  only appear in (IC) through  $w_{10}$ . By setting  $s_0 = T_1$  and  $w_0 = 0$  the principal can make  $w_{10} = 0$  and this does not cost the principal anything since  $s_0$  does not appear in the objective function. Given that  $s_0 = T_1$  and  $w_0 = 0$ ,  $T_0 = T_1$ .

Since  $s_0 = T_1$ , we have  $s_0 \geq s_\emptyset$ , and ( $EF_0$ ) is satisfied.

(c)  $w_{\emptyset} = 0$ : To see this, note that  $w_{\emptyset}$  does not appear in objective function and enters only the (IC) through  $w_{1\emptyset}$  via the threat-point payoff of the agent in the Nash bargaining problem. The Nash bargaining problem that determines  $w_{1\emptyset}$  and  $s_{1\emptyset}$  is given by

$$\begin{aligned} \max_{w,s} & (u(w) - u(w_{\emptyset}))^{\alpha} (s - s_{\emptyset})^{1-\alpha} \\ \text{s.t.} & w + s = w_1 + s_1 \end{aligned}$$

It can be shown that a decrease in  $w_{\emptyset}$  decreases  $w_{1\emptyset}$ . Therefore, from the (IC)  $w_{\emptyset} = 0$ .

(d)  $s_{\emptyset} = s_1$ : To see this note that  $s_{\emptyset}$  does not appear in objective function and enters only the (IC) through  $w_{1\emptyset}$  via the threat-point payoff of the supervisor. It can also be shown that an increase in  $s_{\emptyset}$  reduces  $w_{1\emptyset}$ . Therefore, from the (IC) the principal can raise  $s_{\emptyset}$  until (EF<sub>1</sub>) binds and thus  $s_{\emptyset} = s_1$ .

(e)  $s_1 = 0$ : In the Nash bargaining problem,  $s = s_1 + w_1 - w$ . Since  $s_{\emptyset} = s_1$ , the bargaining problem becomes  $\max (u(w))^{\alpha} (w_1 - w)^{1-\alpha}$ , which is independent of  $s_1$ . Therefore,  $s_1$  can be reduced to zero to minimize the objective function.

Given (a), (b), (c), (d), (e) and the binding (IC) constraint, we can write the Lagrangian as follows:

$$L = \pi w_H + (1 - \pi) w_1 - \lambda [ \pi u(w_H) + (1 - \pi) p u(w_1) - \pi(1 - p) u(w_{1\emptyset}) - \varphi ]$$

$$\frac{\partial L}{\partial w_H} = \pi - \lambda \pi u'(w_H) = 0 \quad (\text{d1})$$

$$\frac{\partial L}{\partial w_1} = (1 - \pi) - \lambda [ (1 - \pi) p u'(w_1) - \pi(1 - p) u'(w_{1\emptyset}) \frac{dw_{1\emptyset}}{dw_1} ] = 0 \quad (\text{d2})$$

$$\text{From (d1)} \quad u'(w_H) = \frac{1}{\lambda}; \text{ from (d2)} \quad u'(w_1) = \frac{1}{\lambda p} + \frac{\pi(1-p)}{(1-\pi)p} u'(w_{1\emptyset}) \frac{dw_{1\emptyset}}{dw_1}.$$

Since the bargaining set becomes bigger as  $w_1$  increases, it can be shown that  $\frac{dw_{1\emptyset}}{dw_1} > 0$ ,

and therefore  $u'(w_H) < u'(w_1)$ , which implies  $w_H > w_1$ . The solution is such that  $w_H > w_1 > 0 = s_1 = s_{\emptyset} = w_{\emptyset} = w_0$  and  $s_0 = w_1 = T_1$ . Note that the (CIC) is violated when  $\sigma = \emptyset$  – the coalition is strictly better off by reporting  $r = 1$  or  $r = 0$ .

Case II:  $T_m = T_{\emptyset}$

Min  $\pi w_H + (1 - \pi) T_{\emptyset}$  s.t.

$$\text{(IC)} \quad \pi u(w_H) + (1 - \pi) p u(w_{\emptyset 1}) - \pi(1 - p) u(w_{\emptyset}) - p u(w_{\emptyset 0}) - \varphi \geq 0$$

$$\text{(EF}_1\text{)} \quad s_1 \geq s_{\emptyset}$$

We make some observations to simplify the optimization problem.

(a)  $w_{\emptyset} \geq w_1$ : To see this, note that  $T_{\emptyset} \geq T_1$  and  $s_1 \geq s_{\emptyset}$ .

(b)  $s_0 = T_{\emptyset}$  and  $w_0 = 0$ : To see this note that  $s_0$  and  $w_0$  only appear in (IC) through  $w_{\emptyset 0}$ . By setting  $s_0 = T_{\emptyset}$  and  $w_0 = 0$ , the principal can make  $w_{\emptyset 0} = w_0 = 0$  since  $s_0$  does not appear in the objective function. Given  $s_0 = T_{\emptyset}$  and  $w_0 = 0$ , we have  $T_0 = T_{\emptyset}$ . Note also that (EF<sub>0</sub>) is satisfied since  $s_0 = T_{\emptyset} \geq s_{\emptyset}$ .

(c)  $w_1 = w_\emptyset$ : To see this, note that  $w_1$  only appears in (IC) through  $w_{\emptyset 1}$  via the threat point payoff of the agent. Therefore the principal can increase  $w_{\emptyset 1}$  and relax the (IC) by increasing  $w_1$ . Since  $w_\emptyset \geq w_1$  from (a),  $w_1$  will be increased until  $w_1 = w_\emptyset$ .

(d)  $s_1 = s_\emptyset$ : To see this, note that  $s_1$  only enters (IC) through  $w_{\emptyset 1}$ . The principal can increase  $w_{\emptyset 1}$  by reducing  $s_1$  since  $s_1$  is the threat-point payoff of the supervisor. It can also be shown that a decrease in  $s_1$  reduces  $w_{\emptyset 1}$ . Therefore, from the (IC), the principal can reduce  $s_1$  until  $(EF_1)$  binds and thus  $s_1 = s_\emptyset$ .

(e)  $w_{\emptyset 1} = w_\emptyset = w_1$ : To see this, note that  $s_1 = s_\emptyset$ ,  $w_1 = w_\emptyset$  and  $T_1 = T_\emptyset$ .

(f)  $s_\emptyset = 0$ : given that  $w_{\emptyset 0} = 0$ ,  $s_\emptyset$  only appears in the objective function and therefore can be reduced to zero.

Also, since  $T_\emptyset = T_1 = w_1$ , we can rewrite the minimization problem as

$$\text{Min } \pi w_H + (1 - \pi) w_1 \text{ s.t. (IC) } \quad \pi u(w_H) + (p - \pi) u(w_1) - \varphi \geq 0$$

And the Lagrangian is:  $L = \pi w_H + (1 - \pi) w_1 + \lambda [\pi u(w_H) + (p - \pi) u(w_1) - \varphi]$ .

The FOCs give the optimal  $w_H$  and  $w_1$  for case II:

$$\frac{\partial L}{\partial w_H} = \pi - \lambda \pi u'(w_H) = 0 \quad (\text{d3})$$

$$\frac{\partial L}{\partial w_1} = (1 - \pi) - \lambda (p - \pi) u'(w_1) = 0 \quad (\text{d4})$$

Therefore, we have shown that the optimal contract under case II is the least-cost-corruption-proof contract.

Case III:  $T_m = T_0$

$$\text{Min } \pi w_H + (1 - \pi) T_0$$

$$\text{(IC) } \quad \pi u(w_H) + (1 - \pi) p u(w_{01}) - \pi (1 - p) u(w_{0\emptyset}) - p u(w_0) - \varphi \geq 0$$

$$\text{(EF}_1\text{)} \quad s_1 \geq s_\emptyset$$

We make a few observations to simplify the optimization problem.

(a)  $s_0 = T_0$  and  $w_0 = 0$ : To see this, note that in the NBS  $w_{01}$  and  $w_{0\emptyset}$  are not affected by the distribution of  $T_0$  between  $s_0$  and  $w_0$  as long as  $w_0 + s_0$  remains the same. Note that by reducing  $w_0$ , (IC) can be relaxed and the objective function reduced. Therefore the principal sets  $w_0 = 0$  and  $s_0 = T_0$ . Note that  $(EF_0)$  is also satisfied since  $s_0 = T_0 = T_m \geq s_\emptyset$ .

(b)  $s_1 = s_\emptyset$  and  $w_1 + s_1 = T_0$ : To see this, note that  $s_1$  and  $w_1$  only affect  $w_{01}$ . By decreasing  $s_1$  and increasing  $w_1$ ,  $w_{01}$  can be increased and (IC) relaxed. Therefore,  $s_1$  is reduced until  $(EF_1)$  binds, and thus  $s_1 = s_\emptyset$ . And  $w_1$  is increased until  $w_1 + s_1 = T_0$  since  $T_0$  is  $T_m$ .

(c)  $s_\emptyset = w_\emptyset = 0$ : To see this, note that in the Nash bargaining problem  $s = w_1 + s_1 - w$  since  $T_1 = T_0$ . Since  $s_1 = s_\emptyset$ , the Nash bargaining problem that determines  $w_{0\emptyset}$  becomes

$$\max_w [u(w) - u(w_\emptyset)]^\alpha (w_1 - w)^{1-\alpha}$$

which is independent of  $s_\emptyset$ . Therefore,  $s_\emptyset$  is reduced to zero to relax the (IC) since (EF<sub>1</sub>) binds from (b). Reducing  $s_\emptyset$  allows the principal to reduce  $s_1$  and increases  $w_{01}$  to relax the (IC). From the NBS  $w_{0\emptyset}$  is reduced by decreasing  $w_{\emptyset\emptyset}$  to zero and therefore relaxing the (IC). Finally, since  $s_1 = s_\emptyset = 0$ ,  $w_1 = T_0$ .

We have proved that the optimization problem and thus the solution for case III is identical to case I. Therefore to find the optimal solution, we only need to compare cases I and II which we do now.

(Case I) Min  $\pi w_H + (1 - \pi) w_1$  subject to

$$(IC) \quad \pi u(w_H) + (1 - \pi) p u(w_1) - \pi (1 - p) u(w_{1\emptyset}) - \varphi = 0$$

(Case II) Min  $\pi w_H + (1 - \pi) w_1$  subject to

$$(IC) \quad \pi u(w_H) + (p - \pi) u(w_1) - \varphi = 0$$

Since Nash bargaining implies  $w_{1\emptyset} < w_1$  for  $\alpha < 1$ , the lowest expected cost under case II can be achieved under case I with a slack (IC). Therefore, the optimal contract under case I results in a smaller expected cost than case II. We have proved that case I is optimal, and it will induce bribery when  $\sigma = \emptyset$ . ■

## Appendix E Proof of the Proposition 2

(i) Consider case I in appendix D, which is the relevant case in equilibrium. Recall the agent's (IC) in equilibrium:  $\pi u(w_H) + (1 - \pi) p u(w_1) - \pi (1 - p) u(w_{1\emptyset}) - \varphi = 0$ .

It can easily be verified that, in state  $\sigma = \emptyset$ , the agent's payoff  $w_{1\emptyset}$  from the Nash bargaining solution increases with the agent's bargaining power  $\alpha$ . Therefore, a decrease in  $\alpha$  will make the (IC) slack and increase the principal's payoff.

(ii) We first characterize the optimal contract where extortion is deterred at zero cost. Then we show that the principal's payoff from the optimal contract approaches the principal's payoff from this contract as the agent's bargaining power goes to zero.

(a) *Optimal contract where extortion is deterred at zero cost:* Since bribery is still an issue, Collusion Incentive Compatibility (CIC) constraints must be added to the principal's problem in appendix A but not the (EF) constraints. By plugging  $s_0$  and  $s_\emptyset$  from (1) and (2) into the principal's objective function and constraint (IC), we can set up the principal's problem as to minimize  $\pi(w_H) + (1 - \pi) (w_1 + s_1)$  s.t.  $\pi u(w_H) + (1 - \pi) p u(w_1) - \pi (1 - p) u(w_\emptyset) - p u(w_0) = \varphi$  with the additional non-negativity constraints. Using standard techniques, we can show that the optimal contract where extortion is deterred at zero cost, denoted by  $\omega^b$ , has the following features:  $w_H > w_1 = s_\emptyset = s_0 > 0 = w_\emptyset = w_0 = s_1$ .

(b) *The principal's payoff from the optimal contract as the agent's bargaining power goes to zero:* Consider the optimal contract derived from case I in appendix D. As  $\alpha \rightarrow 0$ , we know from the NBS that  $w_{1\emptyset} \rightarrow 0$  since the agent's threat point  $w_\emptyset = 0$ . Thus the principal's problem from case I in appendix D simplifies to Min  $\pi w_H + (1 - \pi) w_1$  subject

to  $\pi u(w_H) + (1 - \pi) p u(w_1) - \varphi = 0$ . Note that the principal's problems in (a) and (b) are identical once we replace  $s_1$ ,  $w_\emptyset$  and  $w_0$  in (a) with their solution values (i.e. zero).

(iii) *The LCCP contract is optimal if the agent has all the bargaining power*: Consider the optimal contract derived from case I in appendix D. As  $\alpha \rightarrow 1$ , we know from the NBS that  $w_{1\emptyset} \rightarrow w_1$  since the supervisor's threat point  $s_\emptyset = 0$ . Thus the principal's problem from case I in appendix D simplifies to  $\text{Min } \pi w_H + (1 - \pi) w_1$  subject to  $\pi u(w_H) + (p - \pi) p u(w_1) - \varphi = 0$ . The optimal  $w_H$  and  $w_1$  satisfy:

$$\frac{u'(w_1)}{u'(w_H)} = \frac{1 - \pi}{p - \pi}, \text{ and } \pi u(w_H) + (p - \pi) p u(w_1) = \varphi.$$

Note that these conditions are identical to the conditions in lemma 1 that characterize the *LCCP* contract. ■

### Appendix F Proof of the Proposition 3

In this appendix, we explain how our model changes when the agent's reservation utility, denoted by  $\bar{u}$ , is increased above zero. We show that if  $\bar{u}$  is high enough, the least cost contract that deters bribery also deters extortion, which means that the *LCCP* contract is optimal. Consider the principal's problem  $P^0$  from section 4 but assume that extortion can be deterred at zero cost. That is, we can ignore the (*EF*) constraints and characterize the least cost contract that deters bribery when there is no fear of extortion. We show that ignoring the (*EF*) constraints is without loss of generality if the agent's reservation utility is high enough even if extortion could take place.

Note that when  $\bar{u} > 0$ , the limited liability constraints no longer imply the (*IR*). Therefore, in the problem below, we add an (*IR*) to the principal's problem  $P^0$  from section 4 but ignore the (*EF*) constraints.

We show next that if  $\bar{u}$  is high enough, the solution requires  $w_1 = w_\emptyset$ , which implies that the (*EF*<sub>1</sub>) constraint is then redundant. As earlier in appendix B, we ignore (1) and verify later that  $s_0$  satisfies (1). We can also verify that  $s_0 \geq s_\emptyset$  so that (*EF*<sub>0</sub>) is also redundant as was the case earlier. Replacing  $s_\emptyset$  everywhere using (2), we obtain the Lagrangian:

$L = \pi w_H + (1 - \pi) (w_1 + s_1) - \lambda [\pi u(w_H) + (1 - \pi) p u(w_1) - \pi (1 - p) u(w_\emptyset) - p u(w_0) - \varphi] - \mu [\pi u(w_H) + (1 - \pi) p u(w_1) + (1 - \pi) (1 - p) u(w_\emptyset) - \varphi - \bar{u}] - \delta [w_1 + s_1 - w_\emptyset]$  and the following FOCs:

$$\frac{\partial L}{\partial w_H} = \pi - \lambda \pi u'(w_H) - \mu \pi u'(w_H) \geq 0; \quad w_H \left( \frac{\partial L}{\partial w_H} \right) = 0 \quad (\text{f1})$$

$$\frac{\partial L}{\partial w_1} = (1 - \pi) - \lambda (1 - \pi) p u'(w_1) - \mu (1 - \pi) p u'(w_1) - \delta \geq 0; \quad w_1 \left( \frac{\partial L}{\partial w_1} \right) = 0 \quad (\text{f2})$$

$$\frac{\partial L}{\partial w_\emptyset} = \lambda \pi (1 - p) u'(w_\emptyset) - \mu (1 - \pi) (1 - p) u'(w_\emptyset) + \delta \geq 0; \quad w_\emptyset \left( \frac{\partial L}{\partial w_\emptyset} \right) = 0 \quad (\text{f3})$$

$$\frac{\partial L}{\partial w_0} = \lambda p u'(w_0) \geq 0; \quad w_0 \left( \frac{\partial L}{\partial w_0} \right) = 0 \quad (\text{f4})$$

$$\frac{\partial L}{\partial s_1} = (1 - \pi) - \delta \geq 0; \quad s_1 \left( \frac{\partial L}{\partial s_1} \right) = 0 \quad (\text{f5})$$

There are two cases depending on values of  $\lambda$ .

**(Case 1)  $\lambda > 0$ :**

$\lambda > 0$  implies that (IC) is binding and  $w_0 = 0$  from (f4).  $\lambda > 0$  also leads to  $\mu > 0$  from (f3) as long as  $\bar{u} > 0$ . Otherwise,  $w_\emptyset = 0$  from (f3) and this implies that (IR) is violated when  $\bar{u} > 0$ .

(i) *Subcase:  $\delta = 0$ .* We have  $s_1 = 0$  from (f5) and this implies that  $w_1 \geq w_\emptyset$  since  $s_\emptyset = w_1 + s_1 - w_\emptyset \geq 0$  from (2) the non-negativity constraint on  $s_\emptyset$ .

$$\Rightarrow u'(w_H) = \frac{1}{\lambda + \mu} \text{ and } u'(w_1) = \frac{1}{p(\lambda + \mu)} \Rightarrow w_H > w_1.$$

$$\text{(IC) and (IR)} \Rightarrow u(w_\emptyset) = \frac{\bar{u}}{1 - p}$$

(ii) *Subcase:  $\delta > 0$ .* First we have  $\delta < (1 - \pi)$  from (f2) and this implies that  $s_1 = 0$  from (f5). This result with  $\delta > 0$  (so  $s_\emptyset = 0$ ) leads to  $w_1 = w_\emptyset$ .

$$\text{From (f2) + (f3), we have } u'(w_1) = \frac{1 - \pi}{\mu(1 - \pi) + \lambda(p - \pi)} > \frac{1}{\mu + \lambda} \Rightarrow w_H > w_1.$$

**(Case 2)  $\lambda = 0$**

First we must have  $\delta > 0$  from (f3). Otherwise (f3) implies  $\mu = 0$ . This is because,  $\mu > 0$  in (f3) implies that  $u'(w_\emptyset) = 0$ , which would be a contradiction since it requires an unbounded  $w_\emptyset$ , which implies that (IR) is slack ( $\mu = 0$ ). Note that  $\mu = 0$  implies that  $w_H = 0$  and  $w_1 = 0$  from (f1) and (f2) respectively. However, if this is the case, (IC) is violated.

Since  $\delta > 0$ , we have  $s_\emptyset = 0$ . Moreover, we have  $\delta < (1 - \pi)$  from (f2) and this implies that  $s_1 = 0$  from (f5). This result leads to  $w_1 = w_\emptyset$ . Note that  $w_1 > 0$ , since otherwise we have  $u'(w_1)$  unbounded and (f2) would then imply that  $\mu = 0$  since  $\delta < (1 - \pi)$ . But that would imply that  $w_H = 0$  and (IC) would be violated.

From (f2) and (f3), we have  $u'(w_1) = \frac{1}{\mu} \Rightarrow w_H = w_1$ , and we have the first best.

By collecting results from the two cases, we conclude that the collusion-proof contract is extortion-proof for as long as  $\bar{u} \geq \tilde{u}$ . We obtain  $\tilde{u}$  from the subcase (ii) of Case 1, where both  $u(w_1) = \frac{\tilde{u}}{1 - p}$ , and  $u'(w_1) = \frac{1}{p(\lambda + \mu)}$  hold. From (IR) and (IC), we have

$$\hat{u} = \frac{\varphi(1 - p)}{p}, \text{ where we have the first best for } \bar{u} \geq \hat{u}.$$

In the main text we only considered the case where  $\bar{u} = 0$ . For  $\bar{u} > 0$ , we will either be in the case 1(i), 1(ii), or 2. Using an example, we show that all these cases exist and in the cases 1(ii) and 2, extortion is not relevant.

Suppose  $p = \pi = 0.5$ ,  $\varphi = 1$ . We can show that  $\tilde{u} = 1/4$ , and  $\hat{u} = 1$ . An increase in  $\bar{u}$  (above zero) implies an increase in  $w_\emptyset$ . To prevent (IC) from being violated  $w_H$  and  $w_1$  must increase in a proportion that satisfies the FOC  $u'(w_H) = p u'(w_1)$ . However, the rate of increase in  $w_H$  and  $w_1$  will be lower than the one in  $w_\emptyset$ . At a critical point of  $\bar{u}$ , denoted by  $\tilde{u}$ ,  $w_\emptyset$  becomes the same as  $w_1$  and we switch between cases 1(i) and 1(ii). Beyond this point  $\bar{u}$ , we are in case 1(ii) with  $w_1 = w_\emptyset$ , and the value of  $w_1$  grows with  $\bar{u}$  and approaches  $w_H$ . As  $\bar{u}$  becomes even larger, we reach another critical point of  $\bar{u}$ , denoted by  $\hat{u}$ , and the first best is achieved:  $w_H = w_1 = w_\emptyset$  (case 2). ■

## Appendix G. Generalizing the Production Technology

### Appendix G.1. Optimal Contract with an Incorruptible Supervisor

Given limited liability, and since zero effort entails zero cost, the incentive constraint given in the text will imply that the participation constraint is satisfied in each of the cases we consider. The supervisor's participation constraint is also satisfied due to limited liability. Thus, we ignore both the agent's and the supervisor's participation constraints. The principal's objective function is given by  $Min \pi_1 [p(w_1^H + s_1^H) + (1-p)(w_\emptyset^H + s_\emptyset^H)] + (1 - \pi_1)[p(w_1^L + s_1^L) + (1-p)(w_\emptyset^L + s_\emptyset^L)]$ . Standard techniques give us:  $w_1^H = w_1^L > w_\emptyset^H > 0 = w_\emptyset^L = w_0^H = w_0^L = s_1^H = s_1^L = s_\emptyset^H = s_\emptyset^L = s_0^H = s_0^L$ .

### Appendix G.2. Optimal Contract with a Corruptible Supervisor, but where Extortion Deterred at Zero Cost

Suppose now that the supervisor is corruptible, but that extortion is detected and deterred at zero cost. The possibility of bribery introduces [CIC] constraints which will deter misreporting in lieu of a bribe. We assume that the supervisor does not accept a bribe from the agent if she is indifferent.

$$[CIC_{\sigma,r}] \quad T_\sigma^j \geq T_r^j,$$

where  $T_\sigma = w_\sigma + s_\sigma$ ,  $T_r = w_r + s_r$ , for  $\sigma, r \in \{0, \emptyset, 1\}$  and  $j \in \{L, H\}$ .

We have twelve [CIC] constraints and these can be satisfied only when  $T_0^H = T_\emptyset^H = T_1^H$  and  $T_0^L = T_\emptyset^L = T_1^L$ , i.e., the aggregate transfers in every state with the same output must be the same. Those can also be written as:

$$w_0^H + s_0^H = w_1^H + s_1^H, \quad \Rightarrow \quad s_0^H = w_1^H + s_1^H - w_0^H \quad (g1)$$

$$w_\emptyset^H + s_\emptyset^H = w_1^H + s_1^H, \quad \Rightarrow \quad s_\emptyset^H = w_1^H + s_1^H - w_\emptyset^H \quad (g2)$$

$$w_0^L + s_0^L = w_1^L + s_1^L, \quad \Rightarrow \quad s_0^L = w_1^L + s_1^L - w_0^L \quad (g3)$$

$$w_\emptyset^L + s_\emptyset^L = w_1^L + s_1^L, \quad \Rightarrow \quad s_\emptyset^L = w_1^L + s_1^L - w_\emptyset^L \quad (g4)$$

The agent's participation, incentive constraints and the supervisor's participation constraint are the same as those when the supervisor is honest. Thus, the principal's program which prevents collusion,  $P^{CP}$ , can be written as follows:  $Min \pi_1 [p(w_1^H + s_1^H) + (1-p)(w_\emptyset^H + s_\emptyset^H)] + (1 - \pi_1)[p(w_1^L + s_1^L) + (1-p)(w_\emptyset^L + s_\emptyset^L)]$  s.t. (IC), (g1), (g2), (g3), (g4), and the non-negativity constraints. Using (g2) and (g4) to replace  $s_\emptyset^H$  and  $s_\emptyset^L$  everywhere respectively, we can rewrite the constraints  $s_\emptyset^H \geq 0$  and  $s_\emptyset^L \geq 0$  as follows:

$$w_1^H + s_1^H - w_\emptyset^H \geq 0 \quad (g2)'$$

$$w_1^L + s_1^L - w_\emptyset^L \geq 0 \quad (g4)'$$

Note that the variable  $s_0^H$  and  $s_0^L$  do not appear anywhere else in the problem except in (g1) and (g3) respectively. Therefore, we are free to choose  $s_0^H$  and  $s_0^L$  to satisfy constraints (g1) and (g3) as long as  $s_0^H \geq 0$  and  $s_0^L \geq 0$  respectively. Using standard techniques we can now minimize the principal's objective function subject to (IC), (g2)' and (g3)' to derive the optimal transfers:  $w_1^H = w_\emptyset^H = s_0^H > w_1^L = s_\emptyset^L = s_0^L > 0 = w_\emptyset^L = w_0^H = w_0^L = s_1^H = s_1^L$ .

## REFERENCES

- Acemoglu, D. and Verdier T. (2000), "The Choice between Market Failures and Corruption," *American Economic Review*, 90 (1), pp. 194-211.
- Auriol E. (2006). "Corruption in procurement and public purchase," *International Journal of Industrial Organization*, 24 (5), September, pp. 867-885
- Ayres, I. (1997), "The twin faces of judicial corruption: extortion and bribery," 74 *Denver University Law Review*, 1231.
- Baker G. and Hubbard T. (2003), "Make versus Buy in Trucking: Asset Ownership, Job Design, and Information," *American Economic Review*, 93 (3), June, pp. 551-572.
- Baliga, S. (1999), "Monitoring and collusion with soft information," *Journal of Law, Economics and Organization*, 15, pp. 434-440.
- Bardhan, P. (1997), "Corruption and Development", *Journal of Economic Literature*, September.
- Binmore, K.G., Rubinstein, A., Wolinsky, A. (1986), "The Nash bargaining solution in economic modelling", *Rand Journal of Economics*, 17(2), pp.176-188.
- Bretz, R.D., Milkovich, G. T. and Read, W. (1992), "The Current State of Performance Appraisal Research and Practice: Concerns, Directions, and Implications," *Journal of Management*, 18, pp. 321-332.
- Cadot, O. (1987), "Corruption as a Gamble", *Journal of Public Economics* 33, July, pp. 22 - 44.
- Caillaud, B. and Tirole J. (2007), "Consensus Building: How to Persuade a Group," forthcoming in the *American Economic Review*.
- Carrillo, J. (2000a), "Graft, Bribes, and the Practice of Corruption" *Journal of Economics and Management Strategy*, 9 (2), pp. 257-286.

- Carrillo, J. (2000b), "Corruption in Hierarchies" *Annales d'Economie et de Statistique*, (59), pp. 37-62.
- Che Y.-K. (1995), "Revolving Doors and the Optimal Tolerance for Agency Collusion," *RAND Journal of Economics*, 26 (3), pp. 378-97.
- Colb S. (2001) "Stopping a Moving Target," 3 *RUTGERS RACE & THE LAW REVIEW* 191 (Symposium Article),
- Dewatripont, M. and Tirole J., (2005) "Modes of Communication," *Journal of Political Economy*, 113(6), pp. 1217–1238.
- Faure Grimaud A., Laffont J.-J. and Martimort D. (2003) "Collusion, Delegation and Supervision with Soft Information," *Review of Economic Studies*, 70 (2), pp. 253-279.
- Furnivall, J.S. (1956) "Colonial policy and practice; a comparative study of Burma and Netherlands India." Issued in co-operation with the International Secretariat, Institute of Pacific Relations, Cambridge University Press, England.
- Hindriks, J., Keen, M. and Muthoo, A. (1999). "Corruption, extortion and evasion," *Journal of Public Economics*, 74, pp. 395–430.
- Johnson R. and Libecap G. (1989), "Bureaucratic Rules, Supervisor Behavior, and the Effect on Salaries in the Federal Government," *Journal of Law, Economics, and Organization*, 5(1), pp. 53-82.
- Kessler, A.S. (2000), "On monitoring and collusion in hierarchies," *Journal of Economic Theory*, 91, pp. 280-291.
- Khalil, F. and Lawarrée, J. (2006), "Incentives for corruptible auditors in the absence of commitment," *Journal of Industrial Economics*, 54, pp. 269–291.
- Klitgaard, R.(1988), *Controlling Corruption*, University of California Press, Berkeley, California.
- Kofman, F. and Lawarrée, J. (1993), "Collusion in Hierarchical Agency", *Econometrica*, 61, pp. p.629-656.
- Kofman, F. and Lawarrée, J. (1996), "On the Optimality of Allowing Collusion," *Journal of Public Economics*, 61, pp. 383-407.
- Laffont, J.-J., and Martimort, D. (1997), "Collusion under Asymmetric Information," *Econometrica*, 65, 875–911.

- Laffont, J.-J., and Martimort, D. (2000), "Mechanism Design with Collusion and Correlation," *Econometrica*, 68, 309–342.
- Lambert-Mogiliansky, A. (1998), "On optimality of illegal collusion in contracts," *Review of Economic Design*, 3, pp 303-328.
- La Porta, R., Lopez-de-Silanes, F., Shleifer A. and Vishny, R. (2000), "Investor Protection and Corporate Governance," *Journal of Financial Economics*, 58, pp.3-27.
- Laffont, J.-J. and Tirole J. (1993), *A Theory of Incentive in Procurement and Regulation*, MIT Press.
- Lindgren, J. (1993), "The theory, history and practice of the bribery-extortion distinction," 141 *University of Pennsylvania Law Review*, 1695.
- Mookherjee, D. (1997), "Incentive Reforms in Developing Country Bureaucracies: Lessons From Tax Administration," in Annual Bank Conference in Development Economics, Boris Pleskovic and Joseph Stiglitz (Editors), the World Bank, Washington D.C.
- Mookherjee, D. and Png I. (1995) "Corruptible Law Enforcers: How Should They Be Compensated?" *Economic Journal*, 105 (428), pp. 145-59.
- Olsen, T.E. and Torsvik, G. (1998), "Collusion and renegotiation in hierarchies: a case of beneficial corruption," *International Economic Review*, 39, pp. 413-438.
- Osborne, K.J., Rubinstein, A. 1990, *Bargaining and Markets*, Academic Press, San Diego.
- Polinsky, A.M. and Shavell, S. (2000), "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature*, 38(1), 45-76.
- Polinsky, A.M. and Shavell, S. (2001), "Corruption and optimal law enforcement," *Journal of Public Economics*, 81, pp. 1-24.
- Shavell S., 2004, *Foundations of Economic Analysis of Law*, Harvard University Press.
- Silva, E., Kahn, C. and Zhu, Z. (2007), "Crime and Punishment and Corruption: Who needs 'Untouchables'?", *Journal of Public Economic Theory*, 9 (1), pp. 69-87.
- Strausz, R. (1997a), "Collusion and Renegotiation in a principal-supervisor-agent relationship," *The Scandinavian Journal of Economics*, 99, pp. 497-518.
- Strausz, R. (1997b), "Delegation of Monitoring in a Principal Agent Relationship," *Review of Economic Studies* 64, 337-357.

- Tirole, J. (1986), "Hierarchies and bureaucracies: On the role of collusion in organizations," *Journal of Law, Economics, and Organization*, 2, pp. 181-214.
- Tirole, J. (1992), "Collusion and the theory of organizations," in: Laffont, J.-J. (Ed). *Advances in Economic Theory*, Vol.2 (Cambridge University Press), pp.151-206.
- Vafai, K. (2005), "Collusion and Organization Design," *Economica*, 72, pp.17-37.