Allowing the lesser of two evils: bribery or extortion?

Draft November 27, 2006

FAHAD KHALIL[†] JACQUES LAWARRÉE[‡] SUNGHO YUN*

Abstract

Rewards to prevent enforcement agents from accepting bribes create incentives for extortion. We present a model where a supervisor who can engage in bribery and extortion can still be useful in providing incentive. We show that bribery may be allowed, but extortion is never tolerated in the optimal design of organizations. Allowing extortion penalizes good behavior which increases incentive cost; allowing bribery introduces the bribe as a penalty for bad behavior, which helps restore incentives somewhat. As a key modeling insight, we point out the importance of the appropriate notion of soft information. We demonstrate that the fight against corruption should be rooted in making information hard.

JEL Classification: D82, L23
Key words: Monitoring, Corruption; Collusion, Bribery, Extortion; Framing.

[†] Department of Economics, University of Washington, Seattle, WA 98195, khalil@u.washington.edu
[‡] Department of Economics, University of Washington, Seattle, WA 98195 and ECARES, Brussels lawarree@u.wahington.edu
* Department of Economics, Hanyang University, Ansan, Korea, uwyunsh@hanyang.ac.kr

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.[1] 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.

Notice the important difference between the nature of evidence in the DUI case and the red light case, which turns out to be critical in studying the trade-off between deterring bribery and inducing extortion. In the DUI case, a corrupt officer cannot claim that a sober driver is drunk because hard evidence (such as a blood test) is required. In the red light case however, the testimony of the officer may be enough to convict a driver. We will say that the evidence is soft when the officer can manipulate the evidence (e.g., his testimony), either to help a guilty driver in exchange for a bribe or to extort an innocent driver. Evidence that cannot be manipulated will be described as hard evidence, but we allow for hard evidence to be concealed.[2] The distinction between hard and soft evidence is key to analyzing the trade-off between bribery and extortion and it is relevant to many other settings such as financial or tax audits.

[1] The reward can be non-monetary such as good reputation, promotion, etc.
[2] See, e.g., Tirole (1986). We will make the definitions of hard and soft information precise in our model section.

The difference between bribery and extortion relies on the type of evidence manipulation.[3] 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 in this paper. We also use the generic term of corruption to describe bribery and extortion.

In this paper, we present a model which captures the trade-off between deterring bribery and inducing extortion, and find two new 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.

The intuition for our result (i) is straightforward. Both bribery and extortion make it more costly to provide incentive to an agent, but there is a critical difference between bribery and extortion. Extortion penalizes the agent after "good behavior", while bribery penalizes the agent after "bad behavior". Since bribery occurs when a violation is detected, the bribe is a penalty for "bad behavior", and helps somewhat in providing incentive. This is in line with the less formal literature that suggests that bribes may have some positive role to play but extortion does not. Bribery can help "grease" the incentives in badly run organizations. It is also 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[4]. We show that allowing some form of bribery can be a key part of the optimal design of incentives in an organization.

The intuition for our result (ii) can be understood in light of the existing literature. There is an extensive literature in economics dealing with bribery but our result that the threat of extortion makes bribery optimal is new.[5] Our focus is on the agency literature that followed the pioneering work by Tirole (1986, 1992) as opposed to the non-agency

---

[3] Precise definitions are given later in the model section.

[4] 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 (2000)).

[5] Several papers have shown that it may be optimal to allow bribery by putting restrictions on contracts. For instance, Kofman and Lawarree (1996) (uncertain auditor type); Che (1995) and Mookherjee and Png (1995) (auditor moral hazard); Strausz (1997), Olsen and Torsvik (1998), Lambert-Mogiliansky (1998), and Khalil and Lawarree (2006) (renegotiation and no-commitment).

literature (as reviewed in Bardhan (1997)). In the agency literature, there is no trade-off between bribery and extortion (other than a few exceptions noted below), and a chief reason is that this literature relies on hard information.

To see this, consider a standard moral hazard model with a supervisor who monitors the agent's performance ex post. Suppose, as in Tirole (1986), that the supervisor either finds hard evidence (positive or negative) or finds no conclusive evidence. With hard evidence, the supervisor can hide information and pretend she has found no conclusive evidence but she cannot forge evidence. So if the supervisor has no conclusive evidence, she has no discretion and no bribery or extortion can occur. If the supervisor has incriminating evidence, the agent will want to bribe the supervisor to conceal it. However, this can be deterred without inducing a threat of extortion by rewarding the supervisor *only* for producing incriminating evidence. 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. Therefore extortion is not an issue.

In our model also the information for the supervisor is hard. However, and this is key, we assume that the information for the *supervisor-agent coalition* is soft, i.e., the supervisor can forge evidence with the help of the agent. Now, the principal also has to reward the supervisor for not forging evidence, and not just for presenting incriminating evidence. The new reward goes to the supervisor when she reports no conclusive evidence. This reward makes extortion credible when the supervisor has positive evidence and threatens to conceal it. The trade-off between deterring bribery and extortion appears when information is soft, and we find that allowing bribery is optimal.

One important implication of our analysis is that the fight against bribery should be rooted in making information hard. Most of the literature following Tirole has focused on the problem of bribery in models where extortion is not relevant, i.e., not a credible threat.[6] Other than special circumstances, noted in the footnote above, the

[6] 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

literature largely finds that it is optimal to deter bribery. 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. Our result suggests that the trade-off between bribery and extortion can be avoided by making information hard

One of our contributions is to develop a framework where a supervisor is useful despite the presence of extortion and bribery without having to assume the existence of incorruptible enforcers. In the recent literature, two prominent papers also feature extortion but in different settings and with a different focus. Polinsky and Shavell (2001) study an optimal law enforcement problem, while Hindriks et al. (1999) is a tax-evasion model with a focus on the redistributive properties of the tax scheme. To deter corruption, both 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.

The remaining sections of the paper are organized as follows. Section 2 outlines the model. Section 3 presents the benchmark cases and shows that bribery-proof contract is vulnerable to framing. Section 4 finds the optimal regime by comparing various regimes with each other and characterizes the contract. Section 5 concludes the paper.

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 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 π 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. While the level of output x is observed by all parties, the level of effort e

---

supervisor and the agent. In Kofman and Lawarree (1993) the information structure allows forging of evidence but rules out extortion by assumption.

is private information of the agent. The agent's disutility of effort in terms of money 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 Inada's conditions ($u'(0) = +\infty$ and $u'(+\infty) = 0$). The principal who is risk-neutral offers a take-it-or-leave-it contract to the supervisor and the agent. We assume that Δx 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. The optimal contract in the absence of a supervisor — the well-known *second-best contract* — requires that $w^s{}_H = u^{-1}(\varphi/\pi)$ and $w^s{}_L = 0$. 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 level of effort with probability p or obtains no conclusive evidence with probability $1 - p$, where $p \in (0, 1)$. The supervisor's signal σ can take three values: $\sigma \in \{0, \varnothing, 1\}$, where $\varnothing$ 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 $r \in \{0, \varnothing, 1\}$. We assume that the supervisor is costless but the principal may want to pay her a wage s to deter corruption. The supervisor is risk neutral. 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 agent's and the supervisor's reservation utilities are zero, and that they are protected by limited liability such that $w_r \geq 0$ and $s_r \geq 0$.[7]

[7] 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.

Supervision Technology and Corruption: key assumption

We assume that 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:

Definition 1. *Bribery* occurs when one party accepts a payment in return for manipulating information in favor of the other party.

Definition 2. *Extortion* occurs when the supervisor obtains money from the agent by threatening to falsify evidence that is favorable to the agent. We say *framing* has occurred if the attempt at extortion fails and the supervisor falsifies information that is favorable to the agent.

We assume that the supervisor's information is hard: she cannot fabricate the evidence by herself and the only way to manipulate information by herself is to suppress it, i.e. if $\sigma = e$, she can only report $r \in \{e, \varnothing\}$, and if $\sigma = \varnothing$, the only possible report is $r = \varnothing$. Thus, extortion involves threatening to suppress information favorable to the agent.

However, for extortion to be relevant in this framework, we need to make a critical assumption. We assume that the information for the coalition is soft: with the agent's cooperation, the supervisor can make up evidence and report that the agent has worked regardless of what she observed, i.e. it is possible to have $r \in \{0, \varnothing, 1\}$ regardless of σ. We refer to this as *bribery*.

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 only 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. 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 created, the supervisor has no discretion when $\sigma = \varnothing$, and there is no need to reward the supervisor when $\sigma = \varnothing$. 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).

Besides the standard assumption of enforceable side-contracts (see Tirole 1992), we need to make two additional assumptions. First, 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. Second, we assume that the side-contract is signed after the supervisor finds out about the effort of the agent. This timing is natural but, as we shall see, it also turns out to be important when examining the occurrence of corruption. If the enforceable side-contract was signed before the agent takes his effort, the bribery-proof principle would apply and bribery would not occur in equilibrium. We will have more to say on this in section 4.

Bribery and extortion are accompanied by side-contracts between the supervisor and the agent whereas framing is not. With bribery, the supervisor and agent jointly manipulate 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 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:

(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 σ.

(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.

4. Optimal contract with a corrupt supervisor: trade-off between bribery and extortion

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 and 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 both when x_H is realized and when $r = 1$ with x_L. i.e., $w_H = w_1 > 0 = w_\varnothing = w_0$ (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. If the supervisor is corruptible[8], 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 collected by the agent (w_1).

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 ($r = \varnothing$). If the reward is equal to w_1 (i.e., $s_0 = s_\varnothing = w_1$), 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 $s_\varnothing > 0$ rather than get s_1

[8] It is common knowledge that the supervisor is corruptible. For a dynamic model where the supervisor privately knows her propensity for corruption, see Carillo (2000).

= 0.[9] That is, we see the emergence of the trade-off that we alluded to in the introduction, namely, strong incentives to deter bribes creates scope for a new kind of corruption, namely extortion. As noted above, this trade-off would not appear if we had assumed that information is hard as in many prominent models (e.g., Tirole 1986, 1992, Kessler 2000).[10]

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 whether it is optimal to deter all types of corruption. In particular, we have already shown above that rewards for deterring bribery can encourage extortion/framing, which means there is a trade-off in deterring different kinds of corruption. To study this trade-off, 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 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 the agent had all the bargaining power when negotiating the side-contract.

To prevent bribery the principal will have to ensure that the contract satisfies the Coalition Incentive Compatibility (*CIC*) constraints.

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

---

[9] Anticipating extortion the agent will refuse to put in high effort (his incentive constraint will be violated). Note also that raising s_1 to $s_\varnothing$ is not optimal since it would encourage the coalition to report $r = 1$ when $\sigma = \varnothing$.

[10] 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.

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

$$w_0 + s_0 = w_1 + s_1, \qquad \Rightarrow \qquad s_0 = w_1 + s_1 - w_0 \tag{1}$$

$$w_\varnothing + s_\varnothing = w_1 + s_1, \qquad \Rightarrow \qquad s_\varnothing = w_1 + s_1 - w_\varnothing \tag{2}$$

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) $\qquad s_1 \geq s_\varnothing,$

(EF_0) $\qquad s_0 \geq s_\varnothing.$

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_1) is the relevant constraint for deterring extortion since it deters suppression of positive evidence, whereas (EF_0) deters suppression of negative information, where bribery is the pertinent issue. Therefore, we will ignore the (EF_0) 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 or not colluding, and the supervisor will not extort when she is indifferent.

A corruption-proof contract satisfies the (*CIC*) and (*EF*) constraints and deters both bribery and extortion/framing. 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 (see also Appendix A).[11] Thus, the principal's program which prevents both bribery and extortion/framing, denoted by P^o, can be written as follows:

[11] We can ignore the IR constraints as they are implied by the IC and the limited liability constraints.

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

The solution to this problem is called 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 $p \leq \pi$, the contract is equivalent to the second-best or no-supervisor contract of section 3 with $w_r = s_r = 0$ for $r \in \{0, \varnothing, 1\}$, and $w_H = w_H^s$.
(ii) If $p > \pi$, it is optimal to use the supervisor, and the contract to the agent satisfies:

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

$$where \quad \frac{u'(w_1^o)}{u'(w_H^o)} = \frac{1-\pi}{p-\pi}, \quad \pi(w_H^o) + (p-\pi)u(w_1^o) = \varphi,$$

i.e., the agent obtains an ex ante *rent.*

- *The supervisor's contract involves:*

$$s_1^o = s_\varnothing^o = 0 < s_0^o = w_1^o,$$

but the supervisor receives no ex ante *rent.*[12]

- *The principal's expected cost, denoted by C^o, can be written as*

$$C^o = \pi(w^o{}_H) + (1-\pi)w^o{}_1.$$

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.

As we argued earlier, rewards for turning down bribes introduce incentive to extort/frame. In particular, a reward to the supervisor for reporting $\sigma = \varnothing$ truthfully

[12] 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 $s_0 > 0$.

would encourage the supervisor to extort/frame when $\sigma = 1$. In the corruption-proof contract, this incentive is avoided by reducing $s_\varnothing$ to zero, but then the (*CIC*) requires that $w_\varnothing = w_1$. Therefore, 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. The agent gets a high wage w_1 (= $w_\varnothing$) with probability $1 - p$ even when he shirks since the supervisor is not perfectly accurate.

This implies that the supervisor may not be useful if she is not accurate enough, which is different from the case of the incorruptible supervisor where she is useful for any $p > 0$. If the agent works, he gets this payment with probability $(1 - \pi)(p + (1 - p)) = 1 - \pi$. The net effect on the (*IC*) can be seen by setting $w_\varnothing = w_1$ 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 $w_1 = w_\varnothing = w_0 = 0$, 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 $w_\varnothing$ (= w_1) and therefore it is optimal to set $w^o_1 < w^o_{\mathrm{H}}$. 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 manipulate 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 $r = \varnothing$ 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 $r = \varnothing$ 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 (= $w_\varnothing$) with probability $1 - p$.

Optimal Contract

In this section we characterize the optimal contract when the supervisor can engage in both types of corruption. The principal has always 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.[13]

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 ($\sigma = \varnothing$), the coalition will get $s_1 + w_1$ which they will share. This implies that the agent's payoff when $\sigma = \varnothing$ 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 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 in the following lemma that extortion will never be allowed.

Lemma 2: *Any contract that induces e = 1, but violates* (EF_1) *is strictly dominated by the least-cost corruption-proof contract*.

Proof: *See Appendix C*.

The intuition for never allowing extortion is that it appears as a penalty after the agent has done the right thing, i.e., exerted effort. Thus extortion makes it difficult for the principal to reward the agent for his effort and increases the cost of providing

[13] 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.

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 $\varnothing$ when $\sigma = 1$ is credible, we show that the agent gets the same payoff from the Nash bargaining whether the state is $\varnothing$ 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 $\varnothing$ and 1 either but it is less costly to the principal since the supervisor is not rewarded ($s_1 = s_\varnothing = 0$). 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 *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 $\sigma = \varnothing$, but deters extortion and framing, and the optimal contract will have the following features:*

- $w^*_H > w^*_1 > 0 = w^*_\varnothing = w^*_0$; *when $\sigma = \varnothing$, the agent obtains $kw^*_1 > 0$, where $k < 1$ and k depends on the agent's relative bargaining power.*[14]
- $s_1^* = s_\varnothing^* = 0 < s_0^* = w_1^*$; *the supervisor obtains $(1 - k)\,w^*_1 > 0$ when $\sigma = \varnothing$.*
- *The principal's expected cost, denoted by C^*, is given by*

$$C^* = \pi(w^*_H) + (1 - \pi)w^*_1.$$

Proof: *See Appendix D.*

To see why bribery may help, note from our lemma 1 that the only way to deter all corruption is by not utilizing every piece 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 $\varnothing$ 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

[14] In the Appendix, we define $w_{1\varnothing}$ as the agent's payoff in state $\sigma = \varnothing$ as a result of Nash bargaining and reporting $r = 1$, and thus $k = w_{1\varnothing} / w^*_1$.

compensation between the states ∅ and 1 is by allowing bribery to occur in state ∅. Suppose a bribe from the agent leads the supervisor to overstate performance in state ∅ and report 1. Then the principal will make the same aggregate transfer in both states ∅ and 1, but the agent's payoff in state ∅ is lowered since he has to pay a bribe to the supervisor, and this lowers the cost of inducing high effort.[15]

This captures nicely an intuition often mentioned in the applied literature, that allowing bribery can create markets that improves incentives. Here, the principal relies on the supervisor to extract a bribe from the agent and lower the agent's payoff in state ∅, when it cannot directly do so in 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 occurs if the agent shirks or violates rules.

Supervisor's and agent's bargaining power

From Tirole (1986) we know that 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 the supervisor must exceed any viable offer from the agent. In our model, since the principal lets bribery occur in equilibrium, the bargaining power is relevant. 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 who will find bribery less profitable. Consider a case where the agent has little bargaining power. When the supervisor receives a signal ∅, the agent will want to bribe the supervisor to report $r = 1$ hoping to receive the larger wage $w_1 > w_\varnothing = 0$. However, after bribing the supervisor, the agent will only collect a small fraction of w_1. In other words, an agent with little bargaining power receives little

[15] 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 θ 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 θ is very small, then allowing extortion/faming is not very costly, and the principal should focus on deterring bribery.

benefit from bribing because the bribe acts as an effective penalty. Consequently it is easier for the principal to provide incentive to the agent and therefore the principal is better off when the supervisor has more bargaining power.

Perhaps more interestingly, when the agent has close to zero bargaining power, the principal's payoff is similar to its payoffs when extortion is not an issue. In other words, when the supervisor has all the bargaining power, it is as if extortion was not relevant. To see this, let us envision a case where the supervisor can engage in bribery but cannot extort by assumption. We begin by recalling the incorruptible supervisor contract that is vulnerable to bribery. When the supervisor finds negative or no evidence about the agent's effort, the agent will bribe her to report $r = 1$. To prevent bribery, the supervisor should be rewarded for reporting negative or no evidence, i.e., $s_0 = s_\varnothing = w_1$. Raising $s_\varnothing$ would raise the specter of extortion, but remember that we have ruled out extortion by assumption for the sake of the argument. It can be shown that the new contract that deters bribery but ignores extortion – call it ω^b – is characterized by $w_H > w_1 = s_\varnothing = s_0 > 0 = w_\varnothing = w_0 = s_1$. The principal's payoff from this contract is identical to its payoff in the optimal contract of our proposition when the agent's bargaining power is close to zero.[16] We provide the argument below.

In both ω^b and the optimal contract, the principal has to incur the payment w_1, either as reward or wage, in each of the states $\sigma = \varnothing$ or 1. Thus, the principal would be indifferent between the two contracts if the w_1 were identical. We argue next that this is indeed the case when the agent has no bargaining power. Note that in both contracts the agent gets w_1 when $\sigma = 1$, but when $\sigma = \varnothing$, the agent gets zero in ω^b or his coalition share of w_1 in the optimal contract. Since the latter is close to zero if the supervisor has almost all the bargaining power, the incentive cost of the principal to satisfy the agent's (*IC*) constraint is identical to that under ω^b. Therefore, the wage w_1 (and w_H for that matter) is identical in the two cases, and so is the principal's payoff. To deter extortion, the principal has to allow the agent to obtain a part of w_1 in the side contract with the supervisor, but this does not raise the cost of providing incentive when the agent has no bargaining power.

Remark: *The principal's payoff increases as the agent's bargaining power decreases, and the threat of extortion poses no additional cost if the agent has (almost) no bargaining power.*

[16] Note that when the agent's bargaining power is zero exactly, the agent is indifferent between bribing or not. If we assume that he does not, the principal receives the same payoff as in the incorruptible supervisor contract.

5. Conclusion

A key intuition that has not played much of a role in the literature on bribery in hierarchies is that rewards to enforcement agents to turn down bribes may also encourage them to engage in extortion. Part of the problem is in finding an appropriate model in which a supervisor or enforcement agent remains useful even though they can engage in extortion. Tirole (1986) showed that a corruptible supervisor can still be useful. However, 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 introducing an appropriate notion of soft information, we are able to present a model of extortion in which the supervisor remains useful.

We show that the effect of the trade-off implied by the above intuition may imply that allowing bribery is optimal due to the threat of extortion. In many theoretical models of bribery, the principle of collusion proofness applies and bribery does not appear in equilibrium. We show that this result depends on the softness of information. When information is soft, there is a trade-off between bribery and extortion and collusion or bribery appears in equilibrium. If information is hard, there is no such trade-off 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. 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, it must be kept in mind that this is a second-best result. For example, bribery is optimal in our model because it allows the principal to cause a variation in the agent's payoffs when direct payments from the principal 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.

Finally, our result that allowing bribery may be optimal depends on the fact that we do not allow corrupt behavior to be detected ex post. For example, if there were incorruptible enforcement agents available to detect and sanction corrupt behavior, it would be possible to eliminate bribery in equilibrium. However, 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.

Appendix A Incorruptible Supervisor

Suppose the supervisor always reports truthfully what he has observed. The agent's participation and incentive constraints are as follows:

(IR) $\quad \pi u(w_H) + (1-\pi)\,[pu(w_1) + (1-p)\,u(w_\varnothing)] - \varphi \geq 0,$

(IC) $\quad \pi u(w_H) + (1-\pi)\,[pu(w_1) + (1-p)\,u(w_\varnothing)] - \varphi \geq pu(w_0) + (1-p)\,u(w_\varnothing)$

or, $\quad \pi u(w_H) + (1-\pi)\,pu(w_1) - \pi(1-p)\,u(w_\varnothing) - pu(w_0) \geq \varphi\,.$

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 will ignore both the agent's and the supervisor's participation constraints from now on.

The principal's program when the supervisor is truthful, P^t, can be written as follows:

$$Min \quad \pi(w_H) + (1-\pi)\,[p(w_1 + s_1) + (1-p)\,(w_\varnothing + s_\varnothing)]$$

s.t. $\quad (IC)$, $w_H \geq 0$, $w_r \geq 0$ and $s_r \geq 0$, where $r \in \{0, \varnothing, 1\}$.

The principal's problem has the following Lagrangian:

$$L = \pi(w_H) + (1-\pi)\,[p(w_1 + s_1) + (1-p)\,(w_\varnothing + s_\varnothing)]$$
$$- \lambda\,[\pi u(w_H) + (1-\pi)\,pu(w_1) - \pi(1-p)\,u(w_\varnothing) - pu(w_0) - \varphi]$$

with the additional non-negativity constraints where $\lambda \geq 0$ is the Lagrange multiplier.

The Kuhn-Tucker conditions for minimization are:

$$\partial L/\partial w_H = \pi - \lambda\pi\,u'(w_H) \geq 0; \qquad w_H\left(\partial L/\partial w_H\right) = 0, \qquad \text{(a1)}$$

$$\partial L/\partial w_1 = (1-\pi)\,p - \lambda(1-\pi)\,p\,u'(w_1) \geq 0; \qquad w_1\left(\partial L/\partial w_1\right) = 0, \qquad \text{(a2)}$$

$$\partial L/\partial w_\varnothing = (1-\pi)\,(1-p) + \lambda\,\pi(1-p)\,u'(w_\varnothing) \geq 0; \qquad w_\varnothing\left(\partial L/\partial w_\varnothing\right) = 0, \qquad \text{(a3)}$$

$$\partial L/\partial w_0 = \lambda\,p\,u'(w_0) \geq 0; \qquad w_0\left(\partial L/\partial w_0\right) = 0, \qquad \text{(a4)}$$

$$\partial L / \partial s_1 = (1-\pi)\, p \geq 0; \qquad s_1 \left(\partial L / \partial s_1 \right) = 0, \qquad \text{(a5)}$$

$$\partial L / \partial s_\varnothing = (1-\pi)\,(1-p) \geq 0; \qquad s_\varnothing \left(\partial L / \partial s_\varnothing \right) = 0, \qquad \text{(a6)}$$

plus the complementary slackness conditions for the constraints.

From (a3), (a5) and (a6), we have $w_\varnothing = 0$, $s_1 = 0$ and $s_\varnothing = 0$. Since s_0 does not enter the Lagrangian, it can be any non-negative number and the principal's expected cost is independent of s_0.

Now suppose that $\lambda = 0$. From (a1) and (a2), we have $w_H = w_1 = 0$, which violates the constraint (*IC*). The assumption that $\lambda = 0$ leads to a contradiction. Hence $\lambda > 0$ and (*IC*) is binding. Now (a4) implies that $w_0 = 0$.

The result of $\lambda > 0$ also implies that $w_H = w_1 > 0$. First we argue that both wages are positive and then show that they are equal. If $\partial L / \partial w_H > 0$, then $w_H = 0$ and $1-\lambda u'(0) > 0$, but then (*a2*) implies that $1-\lambda u'(w_1) > 0$ since $w_1 \geq 0$ and $u'' < 0$. This would imply that $w_1 = 0$, but having both $w_H = 0$ and $w_1 = 0$ violates (*IC*). So we must have $\partial L / \partial w_H = 0$ and therefore $w_H > 0$. Likewise, $\lambda > 0$ implies that $w_1 > 0$. Therefore, we have $\partial L / \partial w_H = 0$ and $\partial L / \partial w_1^L = 0$. which leads to $\lambda = 1 / u'(w^H) = 1 / u'(w_1^L)$. Finally, using $w_H = w_1$ in (*IC*), we have $w_H = w_1 = u^{-1}\left(\varphi / \pi + (1-\pi)p \right)$; $\quad w_\varnothing = w_0 = 0$..

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_ϕ everywhere, we can rewrite (EF_1) as (EF_1^b) and state the principal's problem as follows:

$$\text{Min } \pi w_H + (1-\pi)(w_1 + s_1),$$

s.t.

$(IC)\qquad \pi\, u(w_H) + (1-\pi)\, pu(w_1) - \pi(1-p)\, u(w_\varnothing) - pu(w_0) \geq \varphi,$

$(EF_1{}^b)\qquad w_\varnothing \geq w_1,$

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

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 $s_0 \geq 0$. We can now set up the following Lagrangian for this problem:

$$
\begin{aligned}
L = {} & \pi(w_H) + (1-\pi)(w_1 + s_1) \\
& - \delta_1 \left[\pi u(w_H) + (1-\pi)\, pu(w_1) - \pi(1-p)\, u(w_\varnothing) - pu(w_0) - \varphi\right] \\
& - \delta_2 (w_\varnothing - w_1),
\end{aligned}
$$

with the additional non-negativity constraints.

The Kuhn-Tucker conditions for minimization are:

$$\partial L/\partial w_H = \pi - \delta_1 \pi\, u'(w_H) \geq 0; \qquad w_H\,(\partial L/\partial w_H) = 0, \qquad \text{(b1)}$$

$$\partial L/\partial w_1 = (1-\pi) - \delta_1 (1-\pi)\, p\, u'(w_1) + \delta_2 \geq 0; \qquad w_1\,(\partial L/\partial w_1) = 0, \qquad \text{(b2)}$$

$$\partial L/\partial w_\varnothing = \delta_1\, \pi(1-p)\, u'(w_\varnothing) - \delta_2 \geq 0; \qquad w_\varnothing\,(\partial L/\partial w_\varnothing) = 0, \qquad \text{(b3)}$$

$$\partial L/\partial w_0 = \delta_1\, pu'(w_0) \geq 0; \qquad w_0\,(\partial L/\partial w_0) = 0, \qquad \text{(b4)}$$

$$\partial L/\partial s_1 = (1-\pi) \geq 0; \qquad s_1\,(\partial L/\partial s_1) = 0, \qquad \text{(b5)},$$

plus the complementary slackness conditions for the constraints.

From (b5), we have $s_1 = 0$ since $(1-\pi) > 0$. This result, (EF_1), and limited liability imply that $s_\varnothing = 0$. Thus, we have $w_1 = w_\varnothing$ from (2).

Now suppose that $\delta_1 = 0$. From (b1) and (b2), we have $w_H = w_1 = 0$, which violates the constraint (IC). The assumption that $\delta_1 = 0$ leads to a contradiction. Hence $\delta_1 > 0$, (IC) is binding.

The result of $\delta_1 > 0$ also implies that $w_H > 0$ because condition (b1) is violated if we assume that $w_H = 0$ and thus $u'(w_H) = \infty$. Therefore, we have $\partial L / \partial w_H = 0$ and $\delta_1 = 1/u'(w_H)$.

Now (b4) implies that $w_0 = 0$, which leads to $w_1 = s_0$ from (1).

Since we showed above that $w_1 = w_\varnothing$, then using condition (b2) and (b3), we have the following condition

$$\partial L / \partial w_1 + \partial L / \partial w_\varnothing = (1 - \pi) - \delta_1 (p - \pi) \; u'(w_1) \geq 0$$

There are two cases to be considered: (i) $p \leq \pi$ and (ii) $p > \pi$. When (i) $p \leq \pi$, $\partial L / \partial w_1 + \partial L / \partial w_\varnothing$ is always strictly positive, which means $w_1 = w_\varnothing = 0$ since at least one of them must be zero. From (IC), we have $w_H = u^{-1}(\varphi / \pi)$. The contract becomes equivalent to the case when the supervisor is not available.

When (ii) $p > \pi$, $\partial L / \partial w_1 + \partial L / \partial w_\varnothing$ must be zero. If we assume that $\partial L / \partial w_1 + \partial L / \partial w_\varnothing > 0$, then we have $w_1 = w_\varnothing = 0$. However, this implies that $\partial L / \partial w_1 + \partial L / \partial w_\varnothing < 0$ since $u'(w_1) = \infty$, which is a contradiction. By solving $\partial L / \partial w_1 + \partial L / \partial w_\varnothing = 0$, we have the following;

$$\frac{u'(w_1)}{u'(w_H)} = \frac{1-\pi}{p-\pi}.$$

The above equation gives us values of w_H and $w_1 = w_\varnothing$ with binding (IC). Finally, $s_0 = w_1$ is given by (1) and note that the ignored constraint (EF_0) is satisfied in each case. ■

Appendix C Proof of Lemma 2

We proceed in steps. First, we show that the agent receives the same *payoff* from Nash bargaining for $\sigma \in \{\varnothing, 1\}$ 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.]

(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 σ 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

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

where $\alpha \in (0, 1)$ is the agent's bargaining power. 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 $\hat{T}_m = max\ \{\hat{T}_0,\ \hat{T}_\phi,\ \hat{T}_1\}$,

and $\{\hat{w}_H, \hat{w}_r, \hat{s}_r\}$ satisfy the (IC) constraint:

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

Define $\hat{W}_m = \hat{w}_{m1} = \hat{w}_{m\varnothing}$, $\hat{S}_m = \hat{s}_{m1} = \hat{s}_{m\varnothing}$ and simplify (IC):[17]

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

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

(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: $w'_H = \hat{w}_H$, $w'_1 = w'_\varnothing = \widehat{W}_m$, $w'_0 = 0$, $s'_1 = s'_\varnothing = \widehat{S}_m$, and $s'_0 = \widehat{T}_m$.

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

(*CIC*) is satisfied since $\quad w'_k + s'_k = \widehat{T}_m$, $\quad k \in \{0, \varnothing, 1\}$,

(EF_k) is satisfied since $\quad s'_k \geq s'_\varnothing$ $\quad k \in \{0, 1\}$, and

(*IC*) is satisfied since $\quad w'_k$ must satisfy (IC) given that $\hat{w}_k$ satisfies (IC) where $k \in \{H, m0, m\varnothing, m1\}$ and given that $w'_0 \leq \hat{w}_{m\varnothing}$.

Finally, note that $\{w'_H, w'_r, s'_r\}$ is not the least-cost corruption-proof contract since $\widehat{S}_m > 0$, whereas in least-cost corruption-proof contract $s_1^0 = s_\varnothing^0 = 0$. 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

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:

π u(w_H) + (1 – π) p u(w_{m1}) – π (1 – p) u($w_{m\varnothing}$) – p u(w_{m0}) – $\varphi \geq 0$,

where $w_{r\sigma}$ denotes the agents payoff from Nash bargaining when the report is r and the signal is σ. 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, $\varnothing$, or 0 respectively, and show that case I is optimal.

Case I: $T_m = T_1$

Min π w_H + (1 – π) T_1

(IC) π u(w_H) + (1 – π) p u(w_1) – π (1 – p) u($w_{1\varnothing}$) – p u(w_{10}) – $\varphi \geq 0$

(EF_1) $s_1 \geq s_\phi$

We make some observations to simplify the optimization problem.

(a) Note that $w_{m1} = w_1$ because $s_1 \geq s_{\varnothing}$ 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

[17] Note that s_0 could be larger or smaller than $s_\varnothing$ – both cases are captured in $\hat{w}_{m0}$.

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 $s_0 \geq s_\varnothing$, and (EF_0) is satisfied.

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

$$\max_{w,s} \left(u(w) - u(w_\varnothing)\right)^\alpha \left(s - s_\varnothing\right)^{1-\alpha}$$
$$s.t. \quad w + s = w_1 + s_1$$

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

(d) $s_\varnothing = s_1$: To see this note that $s_\varnothing$ does not appear in objective function and enters only the (IC) through $w_{1\varnothing}$ via the threat-point payoff of the supervisor. It can also be shown that an increase in $s_\varnothing$ reduces $w_{1\varnothing}$. Therefore, from the (IC) the principal can raise $s_\varnothing$ until (EF_1) binds and thus $s_\varnothing = s_1$.

(e) $s_1 = 0$: In the Nash bargaining problem, $s = s_1 + w_1 - w$. Since $s_\varnothing = s_1$, 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:

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

$$\partial L / \partial w_H = \pi - \lambda\, \pi\, u'(w_H) = 0 \qquad \text{(c1)}$$

$$\partial L / \partial w_1 = (1 - \pi) - \lambda[(1 - \pi)\, p\, u'(w_1) - \pi\, (1 - p)\, u'(w_{1\varnothing}) \frac{dw_{1\varnothing}}{dw_1}\,] = 0 \qquad \text{(c2)}$$

From (c1) $\quad u'(w_H) = \frac{1}{\lambda}$,

From (c2) $\quad u'(w_1) = \frac{1}{\lambda p} + \frac{\pi(1-p)}{(1-\pi)p}\; u'(w_{1\varnothing})\frac{dw_{1\varnothing}}{dw_1}$.

Since the bargaining set becomes bigger as w_1 increases, it can be shown that $\frac{dw_{1\varnothing}}{dw_1} > 0$, and therefore $u'(w_H) < u'(w_1)$, which implies $w_H > w_1$.

The solution is such that $w_H > w_1 > 0 = s_1 = s_\varnothing = w_\varnothing = w_0$ and $s_0 = w_1 = T_1$. Note that the (*CIC*) is violated when $\sigma = \varnothing$ – the coalition is strictly better off by reporting $r = 1$ or $r = 0$.

<u>Case 2: $T_m = T_\varnothing$</u>

Min $\pi w_H + (1 - \pi) T_\varnothing$

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

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

We make some observations to simplify the optimization problem.

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

(b) $s_0 = T_\varnothing$ and $w_0 = 0$: To see this note that s_0 and w_0 only appear in (*IC*) through $w_{\varnothing 0}$. By setting $s_0 = T_\varnothing$ and $w_0 = 0$, the principal can make $w_{\varnothing 0} = w_0 = 0$ since s_0 does not appear in the objective function. Given $s_0 = T_\varnothing$ and $w_0 = 0$, we have $T_0 = T_\varnothing$. Note also that (EF_0) is satisfied since $s_0 = T_\varnothing \geq s_\varnothing$.

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

(d) $s_1 = s_\varnothing$: To see this, note that s_1 only enters (IC) through $w_{\varnothing 1}$. The principal can increase $w_{\varnothing 1}$ 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 $w_{\varnothing 1}$. Therefore, from the (IC), the principal can reduce s_1 until (EF_1) binds and thus $s_1 = s_\varnothing$.

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

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

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

$Min\ \pi\ w_H + (1 - \pi)\ w_1$

(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:

$$\partial L / \partial w_H = \pi - \lambda\ \pi\ u'(w_H) = 0 \qquad \text{(c3)}$$

$$\partial L / \partial w_1 = (1 - \pi) - \lambda\ (p - \pi)\ u'(w_1) = 0 \qquad \text{(c4)}$$

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

<u>Case 3: </u>$T_m = T_0$

$Min\ \pi\ w_H + (1 - \pi)\ T_0$

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

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

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 $w_{0\varnothing}$ 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 $s_0 = T_0 = T_m \geq s_\varnothing$.

(b) $s_1 = s_\varnothing$ 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 $s_1 = s_\varnothing$. And w_1 is increased until $w_1 + s_1 = T_0$ since T_0 is T_m.

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

$$\max_{w} \left[u(w) - u(w_\varnothing)\right]^{\alpha} (w_1 - w)^{1-\alpha}$$

which is independent of $s_\varnothing$. Therefore, $s_\varnothing$ is reduced to zero to relax the (IC) since (EF_1) binds from (b). Reducing $s_\varnothing$ allows the principal to reduce s_1 and increases w_{01} to relax the (IC). From the NBS $w_{0\varnothing}$ is reduced by decreasing $w_\varnothing$ to zero and therefore relaxing the (*IC*). Finally, since $s_1 = s_\varnothing = 0$, $w_1 = T_0$.

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 $\pi\, w_H + (1 - \pi)\, w_1$

(IC) $\qquad \pi\, u(w_H) + (1 - \pi)\, p\, u(w_1) - \pi\, (1 - p)\, u(w_{1\varnothing}) - \varphi = 0$

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

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

Since Nash bargaining implies $w_{1\varnothing} < w_1$ 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 $\sigma = \varnothing$. ■

References

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.

Carrillo, J. (2000), "Graft, Bribes, and the Practice of Corruption" *Journal of Economics and Management Strategy*, 9 (2), pp. 257-286.

Che Y.-K. (1995), "Revolving Doors and the Optimal Tolerance for Agency Collusion," *RAND Journal of Economics*, 26 (3), pp. 378-97.

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.

Hindriks, J. Keen, M. and Muthoo, A. (1999). "Corruption, extortion and evasion," *Journal of Public Economics*, 74, pp. 395–430.

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.

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.

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.

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 Reivew*, 39, pp. 413-438.

Polinsky, A.M. and Shavell, S. (2001), "Corruption and optimal law enforcement," *Journal of Public Economics*, 81, pp. 1–24.

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

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.