Bribery vs. extortion: allowing the lesser of two evils

April 27, 2007

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

Abstract

Rewards to prevent supervisors from accepting bribes create incentives for extortion. This raises the question whether a supervisor who can engage in bribery and extortion can 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. Even though both increase incentive cost, 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. 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 extortion may be less of a problem in developed countries.

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

We would like to thank Jin Cao, Juan Carrillo, Anke Kessler, Ivan Png, 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.[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 forge 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 forged will be described as hard evidence, but we allow for hard evidence to be concealed.[2] As we explain below, the distinction between hard and soft evidence is key to analyzing the trade-off between bribery and extortion, and it is also relevant to many other settings such as financial or tax audits.

In our model, bribery and extortion differ by how evidence is reported when attempting to extract money from an agent. The enforcer can forge or conceal evidence

[1] 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.

[2] See, e.g., Tirole (1986). We will make the definitions of hard and soft information precise in the model section.

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. In the legal literature, there is a debate on the definitions of extortion and bribery based on who initiates the corrupt transaction.[3] We abstract from this debate because our focus is on whether the corrupt behavior helps or hurts the agent as we are mainly interested in optimal incentives for the agent.

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. 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 capture the above trade-off in a model of extortion in which the supervisor remains useful even when there is no external honest enforcement available. This is our first contribution. Our model allows us to 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.[4] 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 penalizes the agent after

[3] 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).

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

"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 (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. The intuition can be understood in light of the existing literature, even though most of this literature finds that deterring bribery is optimal.[5] We explain below why the information structure used in the current literature fails to capture the trade-off between extortion and bribery, while our information structure succeeds by emphasizing the role of teamwork in forging information.

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 conceal 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.

[5] 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)).

By assuming that information is hard, the previous literature has mainly emphasized that it may be relatively easy (even costless) to conceal or ignore information but that it is prohibitively costly to forge it. 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.

In this paper, we emphasize that forging of information is a team activity by the supervisor and the agent. A supervisor alone may find it very costly to forge information by herself but very cheap when she can collaborate with the agent. Information that is hard for the supervisor can become soft for the supervisor-agent coalition. Our approach is in the spirit of Dewatripont and Tirole (2005) who emphasize that conveying evidence is a team activity. 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? We consider a situation where the team members (supervisor and agent) would find it prohibitively costly to forge information alone but very inexpensive if they can work together.

To deter the coalition from forging evidence, the principal has to pay the supervisor a new reward in addition to the reward for producing incriminating evidence. The new reward goes to the supervisor when she reports no conclusive evidence. It also makes extortion credible when the supervisor has positive evidence and threatens to conceal it. The trade-off between bribery and extortion appears when information is soft, and we find that bribery occurs in equilibrium.

Technically, bribery occurs because of non-separabilities in the constraints that deter corruption (Tirole (1992)). We find that the constraints that would be imposed to deter bribery are interlinked with the constraints imposed to deter extortion. We show that it is more profitable to allow bribery than to deter both forms of corruption.

Soft information is not often used in models of supervision since the supervisor would become useless if she could forge evidence alone. In our model only the coalition can forge evidence, not the supervisor alone. Notice that the agent will not agree to help forge unfavorable evidence. Hence, soft information is not useless because the principal can exploit the incongruence between the team members' objectives (see Dewatripont and Tirole (2005)).

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, the literature largely finds that it is optimal to deter bribery.[7] 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.

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 to 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

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

[7] Several papers have shown that it may be optimal to allow bribery because of 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).

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.

To be sure, there is a pre-existing literature on extortion. Besides the non-agency literature (see Bardhan (1997)), there are two main types of models of extortion in agency settings. In the literature on so-called 'red tape', officials harass citizens by setting up bureaucratic hurdles to extract money. In such models, the principal delegates to the bureaucrat the ability to design part of the incentive scheme, for instance by deciding how much red tape to impose (see, e.g., Banerjee (1997), and Guriev (2004)). Our paper belongs to the other type of models, where the supervisor only has an information gathering role. In this literature, two recent papers feature extortion in different settings and with a different focus than ours. Polinsky and Shavell (2001) study an optimal law enforcement problem, while Hindriks et al. (1999) consider 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.

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 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.[8] 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 φe, 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.[9] 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. 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 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

[8] In section 5, we show that our main results are robust to a more general production function.

[9] We consider the case of a strictly positive reservation utility in section 5.

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 $s_r \geq 0$.[10]

Supervision Technology and Corruption: key assumption

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 misreporting information in favor of the other party.

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

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 the spirit of the recent literature on communication (Dewatripont-Tirole (2005) or Caillaud-Tirole (2007)), we emphasize that forging information is a team activity, and the cost of forging depends on the amount of help from team members. As argued in the introduction, it can be relatively easy to forge when the supervisor can enroll the help of the agent, but very expensive if the supervisor tries to do it alone.[11] In our model, the supervisor cannot forge information by herself but can only conceal it. Her information is hard. If $\sigma = e$, she can only report $r \in \{e, \varnothing\}$, and if $\sigma = \varnothing$, the only

[10] 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.

[11] In financial auditing for instance, the auditee can help the auditor draw "favorable samples."

possible report is $r = \varnothing$. Thus, extortion involves threatening to suppress information favorable to the agent. With the agent's cooperation, the supervisor can forge 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. The information is soft for the coalition.

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. 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 $\sigma = \varnothing$, 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 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. 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.

3. Trade-off 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., $w_{\mathrm{H}} = w_1 > 0 = w_\varnothing = w_0$ (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[12], 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 ($r =\varnothing$). 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

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

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 $s_1 = 0$.[13] 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 (e.g., Tirole (1986), (1992), and Kessler (2000)).[14]

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 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 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).

[13] 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 problematic since it would encourage the coalition to report $r = 1$ when $\sigma = \varnothing$.

[14] 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.

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.

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

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:

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

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

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

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

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. The constraint (EF_0) deters suppression of negative information, and 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 and not colluding, and the supervisor will not extort when she is indifferent.[15]

[15] 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:

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

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

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

We can now present the principal's program – denoted by P^o – which prevents both bribery and extortion/framing.[16]

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

$$\text{s.t.} \quad (IC), (1), (2), (EF_1), (EF_0), w_H \geq 0, w_r \geq 0 \text{ and } s_r \geq 0,$$

$$\text{where } r \in \{0, \varnothing, 1\}$$

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:

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

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

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.*[17]

- *The principal's expected cost is* $C^o = \pi(w^o{}_H) + (1 - \pi)w^o{}_{1.}$

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

[17] 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 $s_0 > 0$.

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.

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. 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 would encourage the supervisor to extort/frame when $\sigma = 1$. This incentive is avoided by reducing $s_{\varnothing}$ to zero, but then the (*CIC*) requires that $w_{\varnothing} = w_1$.

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, 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 $w_{\varnothing} = w_1$ 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 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 $r = \varnothing$ 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 $1 - p$.

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.[18]

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 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_1) *is strictly dominated by the least-cost corruption-proof contract*.

Proof: See Appendix C.

[18] 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 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 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 ∅ when $\sigma = 1$ is credible, we show that the agent gets the same payoff from the Nash bargaining whether the state is ∅ 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 ∅ 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 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 $\sigma = \varnothing$, 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.*[19]
- $s_1^* = s_\varnothing^* = 0 < s_0^* = w_1^*$; *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.

[19] 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 $k = w_{1\varnothing} / w^*_1$.

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 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 compensation between the states $\varnothing$ and 1 is by allowing bribery to occur in state $\varnothing$. Suppose a bribe from the agent leads the supervisor to overstate performance in state $\varnothing$ and report 1. Then the principal will make the same aggregate transfer in both states $\varnothing$ and 1, but the agent's payoff in state $\varnothing$ is lowered since he has to pay a bribe to the supervisor, and this lowers the cost of inducing high effort.[20]

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. The principal can anticipate the side-contracts between the agents and give adequate incentives 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

[20] 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.

evidence in state $\sigma = \varnothing$, and reporting $r = 1$, the principal has to increase the reward $s_\varnothing$, but this increases the cost of deterring extortion in state $\sigma = 1$ since the principal has to maintain $s_1 \geq s_\varnothing$.[21] As argued above and in the LCCP contract, the only way to prevent both forms of misreporting is to require $w_1 = w_\varnothing$, 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 $\varnothing$, 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.

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 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. 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.

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*

[21] In state $\sigma = \varnothing$, the principal needs to satisfy $s_\varnothing \geq s_1 + w_1 - w_\varnothing \geq 0$, which increases the cost of deterring extortion in state $\sigma = 1$.

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 $\varnothing$. This is because the state $\varnothing$ 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 kw_1 from Nash bargaining in state $\varnothing$. As the agent's bargaining power goes down, he earns a smaller return in state $\varnothing$, 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.[22] As the agent's bargaining power goes down, the agent retains a smaller and smaller share of w_1 in state $\varnothing$ 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 $\varnothing$. 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 $s_\varnothing = w_1$. 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.

[22] For example because there is a very efficient appeals process that the agent can utilize if he is extorted.

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 $w_1 = w_\varnothing$. 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 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 ($w_\varnothing$) 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 $w_{\varnothing}$ 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 (*CIC*s) require the same total payments in each state so the principal gains by not increasing w_1 at the same rate as $w_{\varnothing}$ because by doing so it can decrease the reward $s_{\varnothing}$. For a high enough reservation utility, we obtain $w_{\varnothing} = w_1$, which implies that $s_{\varnothing} = 0 = s_1$ 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 $w_{\varnothing} = w_1$ are increased to the point where $w_{\varnothing} = w_1 = w_H$ 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 π_1 when $e = 1$, and it is π_0 when $e = 0$, where $\Delta = \pi_1 - \pi_0 > 0$. 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 $r = 0$, $\varnothing$, and 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 $\varnothing$, it is more likely that a high effort was exerted when the output is high compared to when the output is low.[23] Consequently, raising the wage $w_{\varnothing}^{H}$ (after high output and null report) helps incentives, whereas raising the wage $w_{\varnothing}^{L}$ (after low output and null report) hurts incentives. Thus, when facing the threat of bribery, the principal deters bribery by raising $w_{\varnothing}^{H}$ 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 $w_{\varnothing}^{L}$ 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.

[23] We assume that the null signal is equally likely after a high output or low output.

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

$$(IC) \qquad \pi_1\,[pu(w_1^H) + (1-p)\,u(w_\varnothing^H)] \;+ (1-\pi_1)\,[pu(w_1^L) + (1-p)\,u(w_\varnothing^L)] - \varphi \geq$$
$$\pi_0\,[pu(w_0^H) + (1-p)\,u(w_\varnothing^H)] \;+ (1-\pi_0)\,[pu(w_0^L) + (1-p)\,u(w_\varnothing^L)],$$

which, after rearranging becomes,

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

The main points of interest are the two wages following the signal ∅, when the supervisor finds no conclusive evidence of effort. It is immediate that the $w_\varnothing^H$ helps incentives (positive coefficient), while $w_\varnothing^L$ hurts incentives (negative coefficient). Therefore, the principal prefers to have a positive $w_\varnothing^H$ but would like to set $w_\varnothing^L = 0$. The complete contract when the supervisor is incorruptible is derived 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 $w_\varnothing^j$ is be entirely driven by the (IC). After high output, the principal fights bribery by removing the stake of a bribe ($w_1^H = w_\varnothing^H > 0$), while after low output, it fights bribery by rewarding the supervisor ($w_1^L = s_\varnothing^L > 0 = w_\varnothing^L$) 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 derived 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. 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 trade-off 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.[24] It is important to underline that the trade-off only appears if information is soft. 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, 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

[24] 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.

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. More specifically, 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.

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. For example, if there were incorruptible enforcement agents available to detect and sanction corrupt behavior at a low enough enforcement cost, it would be possible to eliminate bribery in equilibrium. Note that 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. Still detection of extortion is usually not perfect because, e.g., extortion reports may be seen as malevolent.[25] Again we find that developed countries with well-developed legal and institutional structures are more likely to be able to thwart extortion, and extortion may have a more serious impact in the developing world. 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.

[25] 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."

Appendix A Incorruptible Supervisor

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

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

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

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

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:

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

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

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:

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

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

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

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

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

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

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 $\partial L \big/ \partial w_H > 0$, 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 $\partial L \big/ \partial w_H = 0$ and therefore $w_H > 0$. Likewise, $\lambda > 0$ implies that $w_1 > 0$. Therefore, we have $\partial L \big/ \partial w_H = 0$ and $\partial L \big/ \partial w_1^L = 0$. which leads to $\lambda = 1 \big/ u'(w^H) = 1 \big/ u'(w_1^L)$. Finally, using $w_H = w_1$ in (*IC*), we have $w_H = w_1 = u^{-1}\left(\varphi \big/ \pi + (1-\pi)p \right)$; $\quad w_\varnothing = w_0 = 0..$

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*$_1^{\text{b}}$) $\qquad w_\varnothing \geq w_1,$

(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:

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

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 $\sigma \in \{\emptyset, 1\}$ 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.]

(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

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

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 (*EF₁*).

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 $\hat{T}_m = max\ \{\hat{T}_0,\ \hat{T}_\varnothing,\ \hat{T}_1\}$,

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

(IC) π u($\hat{w}_H$) + (1 – π){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):[26]

(IC) π u($\hat{w}_H$) + (p – π) u($\hat{W}_m$) – $\varphi \geq$ p u($\hat{w}_{m0}$)

Note that $\hat{S}_m > 0$ since the supervisor receives at least $\hat{s}_\varnothing$ 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: $w'_H = \hat{w}_H$, $w'_1 = w'_\varnothing = \hat{W}_m$, $w'_0 = 0$, $s'_1 = s'_\varnothing = \hat{S}_m$, 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 ∈ {0, ∅, 1},

(*EF$_k$*) is satisfied since $s'_k \geq s'_\varnothing$ k ∈ {0, 1}, and

(*IC*) is satisfied since w'_k must satisfy (IC) given that $\hat{w}_k$ satisfies (IC) where k ∈ {H, *m0*, *m*∅, *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 $\hat{S}_m > 0$, whereas in least-cost corruption-proof contract $s^0_1 = s^0_\varnothing = 0$. Therefore, the least-cost opportunity-proof contract strictly dominates both $\{w'_H, w'_r, s'_r\}$ and $\{\hat{w}_H, \hat{w}_r, \hat{s}_r\}$. ■

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

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) $s_1 \geq s_\phi$.

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, ∅, 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 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:

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

$$\partial L / \partial w_H = \pi - \lambda\, \pi u'(w_H) = 0 \qquad \text{(d1)}$$

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

From (d1) $\quad u'(w_H) = \dfrac{1}{\lambda}$, From (d2) $\quad u'(w_1) = \dfrac{1}{\lambda p} + \dfrac{\pi(1-p)}{(1-\pi)p}\ u'(w_{1\varnothing})\dfrac{dw_{1\varnothing}}{dw_1}$.

Since the bargaining set becomes bigger as w_1 increases, it can be shown that $\dfrac{dw_{1\varnothing}}{dw_1} > 0$, 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 $r = 0$.

Case II: $T_m = T_\varnothing$

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

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

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

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) $\pi\ u(w_H) + (p - \pi)\ u(w_1) - \varphi \geq 0$

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:

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

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

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

Case III: $T_m = T_0$

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

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

(EF_1) $s_1 \geq s_{\varnothing}$

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

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

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

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

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 $\sigma = \varnothing$. ■

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:

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

It can easily be verified that, in state $\sigma = \varnothing$, the agent's payoff $w_{1\varnothing}$ from the Nash bargaining solution increases with the agent's bargaining power α. Therefore, a decrease in α 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 $s_\varnothing$ from (1) and (2) into the principal's objective function and constraint (*IC*), we can set up the following Lagrangian for this problem:

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

with the additional non-negativity constraints.

The Kuhn-Tucker conditions for minimization are:

$$\partial L/\partial w_H = \pi - \mu\pi u'(w_H) \geq 0; \qquad w_H(\partial L/\partial w_H) = 0, \qquad \text{(e1)}$$

$$\partial L/\partial w_1 = (1-\pi) - \mu\,(1-\pi)\, p\, u'(w_1) \geq 0; \qquad w_1\,(\partial L/\partial w_1) = 0, \qquad \text{(e2)}$$

$$\partial L/\partial w_\varnothing = \mu\, \pi(1-p)\, u'(w_\varnothing) \geq 0; \qquad w_\varnothing\,(\partial L/\partial w_\varnothing) = 0, \qquad \text{(e3)}$$

$$\partial L/\partial w_0 = \mu\, pu'(w_0) \geq 0; \qquad w_0\,(\partial L/\partial w_0) = 0, \qquad \text{(e4)}$$

$$\partial L/\partial s_1 = (1 - \pi) \geq 0; \qquad s_1\,(\partial L/\partial s_1) = 0, \qquad \text{(e5)}$$

plus the complementary slackness conditions for the constraints.

From (e5), we have $s_1 = 0$.

Now suppose that $\mu = 0$. From (e1) and (e2), we have $w_H = w_1 = 0$, which violates the constraint (*IC*). The assumption that $\mu = 0$ leads to a contradiction. Hence $\mu > 0$ and (*IC*) is binding.

Now (e3) and (e4) imply that $w_\varnothing = w_0 = 0$, which leads to $s_0 = s_\varnothing = w_1$ from (1) and (2) respectively.

The result of $\mu > 0$ also implies that $w_H > 0$ because condition (1) is violated if we assume that $w_H = 0$ and thus $u'(w_H) = \infty$. Likewise, $\mu > 0$ implies that $w_1 > 0$. Therefore, we have $\partial L/\partial w_H = 0$ and $\partial L/\partial w_1 = 0$. By solving $\partial L/\partial w_H = 0$ and $\partial L/\partial w_1 = 0$ simultaneously, we have that the optimal w_H and w_1 are given by the following:

$$\frac{u'(w_1)}{u'(w_H)} = \frac{1}{p}, \qquad \text{(e6)}$$

and the binding (*IC*):

$$\pi\, u(w_H) + (1 - \pi)\, p\, u(w_1) - \varphi = 0. \qquad \text{(e7)}$$

Thus, the optimal contract where is extortion is deterred at zero cost, denoted by ω^b, has the following features: $w_H > w_1 = s_\varnothing = s_0 > 0 = w_\varnothing = w_0 = s_1$.

(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 $w_{1\varnothing} \rightarrow 0$ since the agent's threat point $w_\varnothing = 0$. Thus the principal's problem from case I in appendix D simplifies to:

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

subject to

$$\pi\, u(w_H) + (1 - \pi)\, p\, u(w_1) - \varphi = 0$$

And the optimal w_H and w_1 satisfy:

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

Note that these conditions are identical to (e6) and (e7) that characterize the contract that would be offered if extortion is deterred at zero cost.

(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 $w_{1\varnothing} \rightarrow w_1$ since the supervisor's threat point $s_\varnothing = 0$. Thus the principal's problem from case I in appendix D simplifies to:

Min $\pi w_{\mathrm{H}} + (1\text{-}\pi)\, w_1$

subject to

$\pi\, u(w_{\mathrm{H}}) + (p - \pi)$ p $u(w_1) - \varphi = 0$

And the optimal w_{H} and w_1 satisfy:

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

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:

$$\text{Min } \pi w_H + (1-\pi)[p(w_1+s_1) + (1-p)(w_\varnothing + s_\varnothing)]$$

s.t.

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

(*IR*) $\quad \pi u(w_H) + (1-\pi)\, pu(w_1) + (1-\pi)(1-p)\, u(w_\varnothing) \geq \varphi + \bar{u},$

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

(2) $\quad s_\varnothing = w_1 + s_1 - w_\varnothing,$

and the non-negativity constraints.

We show next that if $\bar{u}$ is high enough, the solution requires $w_1 = w_\varnothing$, which implies that the (EF_1) 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 $s_0 \geq s_\varnothing$ so that (EF_0) is also redundant as was the case earlier. Replacing $s_\varnothing$ everywhere using (2), we obtain the Lagrangian:

$$\begin{aligned} L = {} & \pi\, w_H + (1-\pi)\,(w_1 + s_1) \\ & - \lambda\,[\pi\, u(w_H) + (1-\pi)\, p\, u(w_1) - \pi\,(1-p)\, u(w_\varnothing) - p\, u(w_0) - \varphi] \\ & - \mu\,[\pi\, u(w_H) + (1-\pi)\, p\, u(w_1) + (1-\pi)\,(1-p)\, u(w_\varnothing) - \varphi - \bar{u}\,] \\ & - \delta[\, w_1 + s_1 - w_\varnothing] \end{aligned}$$

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

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

$$\partial L/\partial w_\varnothing = \lambda\, \pi(1\text{–}p)u'(w_\varnothing) - \mu(1\text{–}\pi)(1\text{–}p)\, u'(w_\varnothing) + \delta \geq 0; \qquad w_\varnothing\left(\partial L/\partial w_\varnothing\right) = 0 \qquad \text{(f3)}$$

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

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

There are two case depending on values of λ.

(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, $w_\varnothing = 0$ 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 $w_1 \geq w_\varnothing$ since $s_\varnothing = w_1 + s_1 - w_\varnothing \geq 0$ from (2) the non-negativity constraint on $s_\varnothing$.

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

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

(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 $s_\varnothing = 0$) leads to $w_1 = w_\varnothing$.

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

(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 $u'(w_\varnothing) = 0$, which would be a contradiction since it requires an unbounded $w_\varnothing$, 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 $s_\varnothing = 0$. Moreover, we have $\delta < (1 - \pi)$ from (f2) and this implies that $s_1 = 0$ from (f5). This result leads to $w_1 = w_\varnothing$. 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.

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

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$_1$) = $\dfrac{\tilde{u}}{1-p}$, and u'(w$_1$) = $\dfrac{1}{p(\lambda+\mu)}$ hold. From (IR) and (IC), we have $\hat{u} = \dfrac{\varphi\,(1-p)}{p}$, 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 $\tilde{u}$ = ¼, and $\hat{u}$ = 1. An increase in $\bar{u}$ (above zero) implies an increase in w$_\varnothing$. 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$_\varnothing$. At a critical point of $\bar{u}$, denoted by $\tilde{u}$, w$_\varnothing$ 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$_\varnothing$, 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$_\varnothing$ (case 2). ■

Appendix G. Generalizing the Production Technology

Appendix G.1. Optimal Contract with an Incorruptible Supervisor

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

$$(IR)\quad \pi_1\,[pu(w_1^H) + (1-p)\,u(w_\varnothing^H)] + (1-\pi_1)\,[pu(w_1^L) + (1-p)\,u(w_\varnothing^L)] - \varphi \geq 0$$

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

or,
$$\pi_1\,pu(w_1^H) + \Delta\pi(1-p)\,u(w_\varnothing^H) - \pi_0\,pu(w_0^H)$$
$$+ (1-\pi_1)pu(w_1^L) - \Delta\pi(1-p)\,u(w_\varnothing^L) - (1-\pi_0)\,pu(w_0^L) \geq \varphi$$

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:

$$\begin{aligned} Min \quad & \pi_1\,[p(w_1^H + s_1^H) + (1-p)(w_\varnothing^H + s_\varnothing^H)] \\ & \qquad\qquad + (1-\pi_1)[p(w_1^L + s_1^L) + (1-p)(w_\varnothing^L + s_\varnothing^L)] \end{aligned}$$

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

The principal’s problem has the following Lagrangian:

$$\begin{aligned} L = {} & \pi_1\,[p(w_1^H + s_1^H) + (1-p)\,(w_\varnothing^H + s_\varnothing^H)] \\ & + (1-\pi_1)\,[p(w_1^L + s_1^L) + (1-p)\,(w_\varnothing^L + s_\varnothing^L)] \\ & - \lambda\,[\pi_1\,pu(w_1^H) + \Delta\pi(1-p)\,u(w_\varnothing^H) - \pi_0\,pu(w_0^H) \\ & + (1-\pi_1)pu(w_1^L) - \Delta\pi(1-p)\,u(w_\varnothing^L) - (1-\pi_0)\,pu(w_0^L) - \varphi] \end{aligned}$$

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

The Kuhn-Tucker conditions for minimization are:

$$\partial L / \partial w_1^H = \pi_1 p - \lambda\pi_1\,p\,u'(w_1^H) \geq 0; \qquad w_1^H\,(\partial L / \partial w_1^H) = 0, \qquad \text{(g1)}$$

$$\partial L / \partial w_1^L = (1-\pi_1)\,p - \lambda(1-\pi_1)\,p\,u'(w_1^L) \geq 0; \qquad w_1^L\,(\partial L / \partial w_1^L) = 0, \qquad \text{(g2)}$$

$$\partial L / \partial w_\varnothing^H = \pi_1\,(1-p) - \lambda\,\Delta\pi(1-p)\,u'(w_\varnothing^H) \geq 0; \qquad w_\varnothing^H\,(\partial L / \partial w_\varnothing^H) = 0, \qquad \text{(g3)}$$

$$\partial L / \partial w_\varnothing^L = (1-\pi_1)\,(1-p) + \lambda\,\Delta\pi(1-p)\,u'(w_\varnothing^L) \geq 0; \; w_\varnothing^L\,(\partial L / \partial w_\varnothing^L) = 0, \qquad \text{(g4)}$$

$$\partial L / \partial w_0^H = \lambda\,\pi_0 p\,u'(w_0^H) \geq 0; \qquad w_0^H\,(\partial L / \partial w_0^H) = 0, \qquad \text{(g5)}$$

$$\partial L / \partial w_0^L = \lambda\,(1-\pi_0)\,p\,u'(w_0^L) \geq 0; \qquad w_0^L\,(\partial L / \partial w_0^L) = 0, \qquad \text{(g6)}$$

$$\partial L / \partial s_1^H = \pi_1\,p \geq 0; \qquad s_1^H\,(\partial L / \partial s_1^H) = 0, \qquad \text{(g7)}$$

$$\partial L / \partial s_1^L = (1-\pi_1)\,p \geq 0; \qquad s_1^L\,(\partial L / \partial s_1^L) = 0, \qquad \text{(g8)}$$

$$\partial L / \partial s_\varnothing^H = \pi_1\,(1-p) \geq 0; \qquad s_\varnothing^H\,(\partial L / \partial s_\varnothing^H) = 0, \qquad \text{(g9)}$$

$$\partial L / \partial s_\varnothing^L = (1-\pi_1)\,(1-p) \geq 0; \qquad s_\varnothing^L(\partial L / \partial s_\varnothing^L) = 0, \qquad \text{(g10)}$$

plus the complementary slackness conditions for the constraints.

From (g4), (g7), (g8), (g9) and (g10), we have $w_\varnothing^L = 0$, $s_1^H = 0$, $s_1^L = 0$, $s_\varnothing^H = 0$ and $s_\varnothing^L = 0$. 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 (g1), (g2) and (g3), we have $w_1^H = w_1^L = w_\varnothing^H = 0$, which violates the constraint (*IC*). The assumption that $\lambda = 0$ leads to a contradiction. Hence $\lambda > 0$ and (*IC*) is binding. Now (g5) and (g6) imply that $w_0^H = w_0^L = 0$.

The result of $\lambda > 0$ also implies that $w^H = w_1^L > w_\varnothing^H > 0$. First we argue that those wages are positive and then show that $w^H = w_1^L > w_\varnothing^H$. If ${}^{\partial L}\!/\!{}_{\partial w_1^H} > 0$, then $w_1^H = 0$ and $1 - \lambda u'(0) > 0$, but then (g2) and (g3) imply that ${}^{\partial L}\!/\!{}_{\partial w_1^L} > 0$ and ${}^{\partial L}\!/\!{}_{\partial w_\varnothing^H} > 0$ respectively since $w_1^L \geq 0$, $w_\varnothing^H \geq 0$ and $u'' < 0$. This would imply that $w_1^L = w_\varnothing^H = 0$, but having $w_1^H = w_1^L = w_\varnothing^H = 0$ violates (*IC*). So we must have ${}^{\partial L}\!/\!{}_{\partial w_1^H} = 0$. Likewise, $\lambda > 0$ implies that ${}^{\partial L}\!/\!{}_{\partial w_1^L} = 0$. Therefore, we have $\lambda = {}^{1}\!/\!{}_{u'(w_1^H)} = {}^{1}\!/\!{}_{u'(w_1^L)}$. Now suppose ${}^{\partial L}\!/\!{}_{\partial w_\varnothing^H} > 0$. Then we have $w_\varnothing^H = 0$, and $w_1^H = w_1^L > 0$ must hold to satisfy (IC). The assumption of ${}^{\partial L}\!/\!{}_{\partial w_\varnothing^H} > 0$ also implies that $\lambda u'(w_1^H)\, \pi_1(1-p) > \lambda \Delta\pi(1-p)\, u'(0)$ since $1 = \lambda u'(w_1^H)$ from ${}^{\partial L}\!/\!{}_{\partial w_1^H} = 0$. But $\pi_1 u'(w_1^H)$ is always smaller than $\Delta\pi u'(0)$ since $u'(0) = +\infty$. Therefore, we have ${}^{\partial L}\!/\!{}_{\partial w_\varnothing^H} = 0$, which leads to $\frac{u'(w_\varnothing^H)}{u'(w_1^H)} = \frac{\pi_1}{\Delta\pi} > 1$. Now we have $w_1^H = w_1^L > w_\varnothing^H > 0$. Finally, the values of w_1^H, w_1^L and $w_\varnothing^H$ are determined such that both of $\frac{u'(w_\varnothing^H)}{u'(w_1^H)} = \frac{\pi_1}{\Delta\pi}$ and $pu(w_1^H) + \Delta\pi(1-p)\, u(w_\varnothing^H) = \varphi$ are satisfied. Collecting our results gives us: $w_1^H = w_1^L > w_\varnothing^H > 0 = w_\varnothing^L = w_0^H = w_0^L = s_1^H = s_1^L = s_\varnothing^H = s_\varnothing^L = s_0^H = s_0^L$.

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.

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

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

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

$$w_0^H + s_0^H = w_1^H + s_1^H, \quad => \quad s_0^H = w_1^H + s_1^H - w_0^H \tag{g11}$$

$$w_\varnothing^H + s_\varnothing^H = w_1^H + s_1^H, \quad => \quad s_\varnothing^H = w_1^H + s_1^H - w_\varnothing^H \tag{g12}$$

$$w_0^L + s_0^L = w_1^L + s_1^L, \quad => \quad s_0^L = w_1^L + s_1^L - w_0^L \tag{g13}$$

$$w_\varnothing^L + s_\varnothing^L = w_1^L + s_1^L, \quad => \quad s_\varnothing^L = w_1^L + s_1^L - w_\varnothing^L \tag{g14}$$

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:

$$Min \quad \pi_1[p(w_1^H + s_1^H) + (1-p)(w_\varnothing^H + s_\varnothing^H)] + (1-\pi_1)[p(w_1^L + s_1^L) + (1-p)(w_\varnothing^L + s_\varnothing^L)]$$

s.t. (*IC*), (g11), (g12), (g13), (g14), and the non-negativity constraints.

Using (g12) and (g14) to replace $s_\varnothing^H$ and $s_\varnothing^L$ everywhere respectively, we can rewrite the constraints $s_\varnothing^H \geq 0$ and $s_\varnothing^L \geq 0$ as follows:

$$w_1^H + s_1^H - w_\varnothing^H \geq 0 \tag{g12}'$$

$$w_1^L + s_1^L - w_\varnothing^L \geq 0 \tag{g14}'$$

Note that the variable s_0^H and s_0^L do not appear anywhere else in the problem except in (g11) and (g13) respectively. Therefore, we are free to choose s_0^H and s_0^L to satisfy constraints (g11) and (g13) as long as $s_0^H \geq 0$ and $s_0^L \geq 0$ respectively. We can now set up the following Lagrangian for this problem:

$$\begin{aligned}
L = {} & \pi_1 (w_1^H + s_1^H) + (1 - \pi_1)(w_1^L + s_1^L) \\
& - \mu_1 [\pi_1 pu(w_1^H) + \Delta\pi(1 - p)\, u(w_\varnothing^H) - \pi_0 pu(w_0^H) \\
& + (1 - \pi_1) pu(w_1^L) - \Delta\pi(1 - p)\, u(w_\varnothing^L) - (1 - \pi_0)\, pu(w_0^L) - \varphi] \\
& - \mu_2 (w_1^H + s_1^H - w_\varnothing^H) \\
& - \mu_3 (w_1^L + s_1^L - w_\varnothing^L)
\end{aligned}$$

with the additional non-negativity constraints.

The Kuhn-Tucker conditions for minimization are:

$$\frac{\partial L}{\partial w_1^H} = \pi_1 - \mu_1\, \pi_1\, p\, u'(w_1^H) - \mu_2 \geq 0; \qquad w_1^H \left(\frac{\partial L}{\partial w_1^H}\right) = 0, \tag{g15}$$

$$\frac{\partial L}{\partial w_1^L} = (1 - \pi_1) - \mu_1 (1 - \pi_1)\, p\, u'(w_1^L) - \mu_3 \geq 0; \qquad w_1^L \left(\frac{\partial L}{\partial w_1^L}\right) = 0, \tag{g16}$$

$$\frac{\partial L}{\partial w_\varnothing^H} = - \mu_1\, \Delta\pi(1 - p)\, u'(w_\varnothing^H) + \mu_2 \geq 0; \qquad w_\varnothing^H \left(\frac{\partial L}{\partial w_\varnothing^H}\right) = 0, \tag{g17}$$

$$\frac{\partial L}{\partial w_\varnothing^L} = \mu_1\, \Delta\pi(1 - p)\, u'(w_\varnothing^L) + \mu_3 \geq 0; \qquad w_\varnothing^L \left(\frac{\partial L}{\partial w_\varnothing^L}\right) = 0, \tag{g18}$$

$$\frac{\partial L}{\partial w_0^H} = \mu_1\, \pi_0\, pu'(w_0^H) \geq 0; \qquad w_0^H \left(\frac{\partial L}{\partial w_0^H}\right) = 0, \tag{g19}$$

$$\frac{\partial L}{\partial w_0^L} = \mu_1 (1 - \pi_0)\, pu'(w_0^L) \geq 0; \qquad w_0^L \left(\frac{\partial L}{\partial w_0^L}\right) = 0, \tag{g20}$$

$$\frac{\partial L}{\partial s_1^H} = \pi_1 - \mu_2 \geq 0; \qquad s_1^H \left(\frac{\partial L}{\partial s_1^H}\right) = 0, \tag{g21}$$

$$\frac{\partial L}{\partial s_1^L} = (1 - \pi_1) - \mu_3 \geq 0; \qquad s_1^L \left(\frac{\partial L}{\partial s_1^L}\right) = 0, \tag{g22}$$

plus the complementary slackness conditions for the constraints.

First, we show that $\mu_2 > 0$ and constraint (g12)′ is binding. Suppose that $\mu_2 = 0$. Then we have $s_1^H = 0$ from (g21) and $\mu_1 = 0$ from (g17), which leads to that $w_1^H = 0$ from (g15). These results imply that $w_{\varnothing}^{H} = 0$ from (g12)′. There are two cases depending on value of μ_3. Suppose (i) $\mu_3 = 0$. Then we have $w_1^{L} = 0$ from (g16), which violates (*IC*). Now suppose (ii) $\mu_3 > 0$. This implies that constraint (g14)′ is binding and $w_{\varnothing}^{L} = 0$ from (g18), which in turn implies that $w_1^{L} = s_1^{L} = 0$. (*IC*) is violated again. The assumption that $\mu_2 = 0$ leads to a contradiction.

Now we argue that the result of $\mu_2 > 0$ leads to $\mu_1 > 0$ and (IC) is binding. Suppose $\mu_1 = 0$ given that $\mu_2 > 0$. From (g17), we have $w_{\varnothing}^{H} = 0$, which implies that $w_1^{H} = s_1^{H} = 0$ since constraint (g12)′ is binding. There are also two cases depending on value of μ_3. Suppose (i) $\mu_3 = 0$. From (g16), we have $w_1^{L} = 0$, which violates (*IC*). Now suppose (ii) $\mu_3 > 0$, which implies that constraint (g14)′ is binding and $w_{\varnothing}^{L} = 0$ from (g18), which in turn implies that $w_1^{L} = s_1^{L} = 0$. (*IC*) is violated again. The assumption that $\mu_1 = 0$ leads to a contradiction.

Now (g18), (g19) and (g20) imply that $w_{\varnothing}^{L} = w_0^{H} = w_0^{L} = 0$.

The result of $\mu_1 > 0$ also implies that $w_1^{H} > 0$ because condition (g15) is violated if we assume that $w_1^{H} = 0$ and thus $u'(w_1^{H}) = \infty$. Likewise, $\mu_1 > 0$ also implies that $w_1^{L} > 0$. Therefore, we have $\partial L / \partial w_1^{H} = \partial L / \partial w_1^{L} = 0$. By plugging $\mu_2 = \pi_1 - \mu_1\, \pi_1\, p\, u'(w_1^{H})$ from $\partial L / \partial w_1^{H} = 0$ into (g17), we have the following:

$$\partial L / \partial w_{\varnothing}^{H} = \pi_1 - \mu_1\, \pi_1\, p\, u'(w_1^{H}) - \mu_1\, \Delta\pi(1-p)\, u'(w_{\varnothing}^{H}) \geq 0.$$

If we assume that $w_{\varnothing}^{H} = 0$ and thus $u'(w_1^{H}) = \infty$, above condition is violated. Therefore, we have $w_{\varnothing}^{H} > 0$ and $\partial L / \partial w_{\varnothing}^{H} = 0$. The result of $w_1^{L} > 0$ implies that $\mu_3 = 0$. If we assume that $\mu_3 > 0$ and thus (g14)′ is binding, then we have $w_1^{L} = 0$ because $w_{\varnothing}^{L} = 0$, which leads to a contradiction.

By plugging $\partial L / \partial w_1^{H} = 0$ and $\partial L / \partial w_1^{L} = 0$ into (g21) and (g22) respectively, we have $\partial L / \partial s_1^{H} > 0$ and $\partial L / \partial s_1^{L} > 0$, which lead to that $s_1^{H} = s_1^{L} = 0$.

The results that $s_1^{H} = 0$ and (g12)′ is binding imply that $w_1^{H} = w_{\varnothing}^{H}$.

From (g17), we have $\mu_2 = \mu_1\Delta\pi(1-p)\, u'(w_{\varnothing}{}^{H})$ since $\partial L / \partial w_{\varnothing}^{H} = 0$. By plugging this into $\partial L / \partial w_{1}^{H} = 0$, we have $\pi_1 - \mu_1\, [\pi_1 p\, u'(w_1{}^{H}) + \Delta\pi(1-p)\, u'(w_{\varnothing}{}^{H})] = \pi_1 - \mu_1\, [\pi_1 p + \Delta\pi(1-p)]\, u'(w_1{}^{H}) = 0$ since $w_1{}^{H} = w_{\varnothing}{}^{H}$. This and $\partial L / \partial w_{1}^{L} = 0$ imply the following:

$$\frac{u'(w_1^L)}{u'(w_1^H)} = \frac{\Delta\pi + \pi_0 p}{\pi_1 p} > 1 \qquad \text{(g23)}$$

Therefore, $w_1{}^{H} = w_{\varnothing}{}^{H} > w_1{}^{L}$.

Finally, the values of $w_1{}^{H}$, $w_1{}^{L}$ and $w_{\varnothing}{}^{H}$ are determined such that both of (g23) and $[\pi_1\, p + \Delta\pi(1-p)]\, u(w_1{}^{H}) + (1-\pi_1)pu(w_1{}^{L}) = \varphi$ are satisfied. Collecting our results gives us: $w_1{}^{H} = w_{\varnothing}{}^{H} = s_0{}^{H} > w_1{}^{L} = s_{\varnothing}{}^{L} = s_0{}^{L} > 0 = w_{\varnothing}{}^{L} = w_0{}^{H} = w_0{}^{L} = s_1{}^{H} = s_1{}^{L}$.

References

Ayres, I. (1997), “The twin faces of judicial corruption: extortion and bribery,” 74 *Denver University Law Review*, 1231.

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.

Banerjee, A. (1997), “A Theory of Misgovernance”, *Quarterly Journal of Economics*, 112(4), pp. 1289-1332.

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.

Caillaud, B. and Tirole J., “Consensus Building: How to Persuade a Group,” Mimeo, January 2007.

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.

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.

Guriev, S. (2004), “Red Tape and Corruption”, *Journal of Development Economics*, 73, pp. 489-504.

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.

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.

Lindgren, J. (1993), “The theory, history and practice of the bribery-extortion distinction,” 141 *University of Pennsylvania Law Review*, 1695.

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.

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

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. (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.