Equilibrium Extortion, Bargaining Power, and the Tradeoff between Informational and Corruption Rents

Troy J. Scott, RTI International
Equilibrium Extortion, Bargaining Power, and the Tradeoff between Informational and Corruption Rents*

Troy J. Scott†

June 8, 2014

Abstract

The paper considers a principal–supervisor–agent hierarchy where the supervisor is self-interested and able to manipulate information. The supervisor’s self interest may motivate him to accept a bribe in exchange for a report that is overly favorable to the agent or to extort the agent by demanding payment for a favorable report to which the agent is fairly entitled. When attempting to extort the agent, the supervisor may become irrationally attached to his bargaining position, including threats that if carried out would incur a loss ex post; realizing this, the agent may accede to extortion. In spite of this possibility for strictly positive extortion payments, incentives are improved – in that the information rent is lowered – by reducing the bargaining power of the agent, relative to that of the supervisor. However, the expected cost of these payments – the corruption rent collected by the supervisor – is passed on to the principal through the agent’s participation constraint, and this rent is lowered by increasing the agent’s relative bargaining power. Thus there is a tradeoff between informational and corruption rents: Increasing the agent’s relative bargaining power increases the information rent and decreases the corruption rent, and the overall effect on the value of the optimal contract tends to follow a simple rule: Strong supervisors are preferred when supervision is less effective, and strong agents are preferred when supervision is more effective.

JEL classification: D73, D82, D86

Keywords: Adverse selection; Supervision; Information Rent; Corruption Rent; Collusion; Bribery; Extortion; Bargaining Power

---

*I thank Jacques Lawarrée and Fahad Khalil for invaluable advice and suggestions. I am also grateful to John T. Scott, S. Nageeb Ali, Lan Shi, Sumant Rai, and seminar discussants at the University of Washington and Dartmouth College.

†RTI International: 3040 East Cornwallis Road; Research Triangle Park, NC 27709: tjscott@rti.org*
1 Introduction

Two recent surveys are far from alone in demonstrating the prevalence of petty corruption in the everyday lives of many of the world’s citizens and the staggering magnitude of the problem when payments are aggregated: The United Nations Office on Drugs and Crime reports that one in two adults in Afghanistan paid at least one bribe to a public official in 2009, with payments totaling $2.5 billion – nearly a quarter of the country’s licit GDP. Transparency International reports that 10 percent of people surveyed worldwide paid bribes in 2008 (the figure rises to 26 percent in Sub-Saharan Africa, 28 percent in Newly Independent States, and 40 percent in the Middle East and North Africa). On average, bribes amounted to seven percent of respondents’ annual income worldwide.

Although payments are generically referred to as bribes, it is clear in these accounts that much of what transpires is extortion. In order to elicit “bribes,” tax auditors threaten to disallow legitimate deductions and police officers threaten to overlook evidence of suspects’ innocence.

Law enforcement, regulatory agencies, tax bureaus and other government bureaucracies feature prominently in these and other accounts of corruption. Each of these institutions might be viewed as the middle rung of a supervisory hierarchy, justified in large part by its ability to overcome an informational asymmetry between top and bottom. Bribery and extortion in these settings – corruption that is information-based – can be expected to alter incentives and drive a wedge between the nominal and real effects of institutions that greatly shape social interaction and economic activity.

The present paper offers a stylized analysis of these issues in a principal/supervisor/agent model with adverse selection. The agent has private information about his type, which makes it costly for the principal to implement an efficient allocation rule and brings up the familiar tradeoff between efficiency and

---

2 Transparency International (2009) Global Corruption Barometer. World Bank Enterprise Surveys suggest that corruption is also pervasive at the firm level – much of which it would be understatement to call petty.
3 See also Klitgaard (1988), Olken and Barron (2009), and Polgreen (2010).
information rent. The supervisor is sometimes able to discover the agent’s type, which typically makes supervision helpful to the principal. However, a supervisor who finds unfavorable information may collude with the agent and agree to make instead a favorable report in exchange for a bribe, and a supervisor who finds favorable information may extort the agent by threatening to suppress the information (i.e., threatening to frame the agent) unless the agent pays him to report truthfully.

The paper sets out to construct a model in which the principal may tolerate either bribery or extortion, or both, in equilibrium. Therefore, the determination of side transfers – bribes and extortion payments – is a key issue. In order to construct a tractable model, a simple Nash bargaining program, which allows for bargaining power to be unequally divided between the supervisor and the agent, is used to determine transfers. The supervisor is given an added advantage in cases where he has a choice of action if negotiations break down: Irrational actions are improbable but not impossible. In particular, the supervisor may threaten to withhold a favorable report – even if actually doing so would require the supervisor to forego a reward – and the agent must be concerned about the small possibility that the supervisor is just crazy enough to do it.

A parameter \( \gamma \) is introduced to capture the degree of difficulty for the supervisor/agent coalition in manipulating information. The ease of manipulation, \( 1 - \gamma \), is analogous to Laffont’s (1990) index, \( \delta \), of the softness of the supervisor’s information. In law enforcement, \( \gamma \) might capture the availability of sophisticated forensic evidence – e.g., the fraction of police labs that can type blood or match a bullet to a gun. In auditing, one might think of \( \gamma \) as the efficacy of auditing standards and \( 1 - \gamma \) as the probability that the supervisor/agent coalition finds a loophole it can exploit.\(^4\)

1.1 Contributions

The paper’s main contribution to theory is to identify a tradeoff, with respect to the division of bargaining power between the supervisor and agent, between the standard

---

4 See de la Merced and Sorkin (2010) for a tale of one such loophole and Norris (2010) for an account of the IRS’s effort to close another. La Porta et al. (2000) discuss the determinants of hard information in financial auditing.
information rent paid to the agent and the corruption rent paid to the supervisor through elicit bribes and extortion payments. Increasing the agent’s bargaining power increases the information rent and decreases the corruption rent. This tradeoff exists whenever bribes are paid in equilibrium, but it is easily overlooked because the information rent is typically the dominant problem and the principal therefore always prefers to decrease the bargaining power of the agent relative to the supervisor (Kofman and Lawarrée (1996), Khalil, Lawarrée, and Yun (2010)). The intuition is that, for an agent contemplating underreporting his productivity, the prospect of paying a large bribe to the supervisor if he is caught has an incentive effect similar to that of paying a fine to the principal, and the closer the amount of the bribe to amount of the fine the better for incentives. But this is only half the story. Yes, raising the equilibrium bribe reduces the information rent of the agent, but it also raises the corruption rent of the supervisor, and the dominance of the informational rent is not robust. The introduction of even a small difficulty for the supervisor/agent coalition in forging information ($\gamma > 0$, however small) allows for the possibility that the principal may prefer to increase the bargaining power of the agent.

1.2 Relationship to the Literature

The model presented is closest to those of Khalil, Lawarrée, and Yun (2010) and Vafai (2002, 2005), which consider both bribery and extortion in vertical structures with moral hazard. Vafai finds it optimal always to deter both forms of corruption. Khalil, Lawarrée, and Yun find that the potential for extortion may lead to optimal bribery, but extortion is never tolerated. In the present paper, extortion-deterrence is always optimal under the respective assumptions of Vafai and Khalil, Lawarrée, and Yun for two different and rather special reasons, and under more general conditions the principal may prefer to tolerate extortion and bribery both.

Traditional analyses of corruption tend not to treat the provision of incentives under asymmetric information as a central issue. Shleifer and Vishny (1993) model corrupt rent-seeking behavior by public officials who sell public goods (permits or licenses, e.g.)

---

5 See Shleifer and Vishny (1993); Ades and Di Tella (1999); Choi and Thum (2004); surveys by Ades and Di Tella (1997), Aidt (2003), and Bardhan (1997); and additional references in Polinsky and Shavell (2007).
for private gain, taking as given the opportunities for such behavior and comparing the
case of many independent oligopolist officials selling complementary goods (e.g.,
multiple permits from different agencies that are perfect complements in some firm’s
production function) with the case of a cohesive corrupt entity that monopolizes the sale
of these goods.

The incentive effects of information-based corruption are only partially addressed
in the literature on collusion in hierarchies started by Tirole (1986). This literature has
largely chosen to ignore extortion, purposefully adopting information structures in which
bribery and collusion are relevant while extortion is not. The majority of this literature
studies environments in which it is optimal to deter collusion in equilibrium, although
there are several notable exceptions.

Extortion is featured in several recent agency models, but it has not been analyzed
in a way that illuminates the traditional tradeoff between efficiency and information
rent. In contrast to these, the present paper focuses on the principal’s most elemental
contracting problem – designing a schedule of transfers to the agent and supervisor,
contingent on the agent’s performance and the supervisor’s report, to maximize an
objective function subject to incentive constraints.

The remainder of the paper is organized as follows: Section 2 introduces a familiar
model with slight modifications made to highlight the effects of extortion. Section 3
presents benchmark results without corruption. Section 4 first lays out a simple, reduced-
form, bargaining model to describe bribery and extortion. It goes on to describe two
strategic-form games that provide a foundation for the reduced-form model and impose
some reasonable restrictions on the extent to which irrational threats may support

---

6 See, generally, Laffont and Tirole (1991), Kofman and Lawrèe (1993), Mookherjee and Png (1995), and
Kessler (2000). An early exception is Laffont (1990), which studies “hidden gaming” similar to extortion.
7 See for example Che (1995), Besley and McLaren (1993), Khalil and Lawrèe (2006), Kofman and
Lawrèe (1996), Laffont and N’Guessan (1999), and Straub (2009).
8 Mookherjee (1997) and Hindriks, Keen, and Muthoo (1999) study tax auditing; Polinsky and Shavell
(2001) analyze a problem of optimal law enforcement; Auriol (2006) studies procurement; and Acemoglu
extortion. Section 5 rules out a number of mechanisms, setting the stage for the exposition of optimal mechanisms in Section 6. Section 7 concludes.

2 Model

The agent \((A)\) exerts an effort \(e\) that, together with a productivity parameter \(\theta\), determines output \(x = \theta + e\). \(\theta\) takes one of two positive values, \(\theta_L\) or \(\theta_H\), with \(\theta_H - \theta_L \equiv \Delta \theta > 0\). \(\theta_L\) obtains with probability \(\xi\), and \(\theta_H\) obtains with probability \(1 - \xi\). \(\theta\) and \(e\) are the private information of \(A\) while \(x\) is publicly observable and verifiable. \(A\)'s disutility of effort is \(g(e) = e^2/2\). The output belongs to the principal \((P)\), who compensates \(A\) with a transfer \(t\).\(^9\) \(A\) maximizes his expected utility \(u = E[t - g(e)]\).

The nature of the uncertainty will be explained below.

The supervisor \((S)\) observes a signal \(\sigma \in \{L, \emptyset, H\}\) and makes a public report \(\sigma^\circ\). The correlation between \(\sigma\) and \(\theta\) is such that \(\text{Prob}(\sigma = L|\theta_L) = \text{Prob}(\sigma = H|\theta_H) = \nu\), and \(\text{Prob}(\emptyset|\theta_L) = \text{Prob}(\emptyset|\theta_H) = (1 - \nu)\). That is, \(S\) learns the true state with probability \(\nu\) and observes nothing with probability \((1 - \nu)\). \(S\) is never “wrong” in the sense that \(\text{Prob}(\sigma = H|\theta_L) = \text{Prob}(\sigma = L|\theta_H) = 0\). \(S\) may be paid a wage \(W \geq 0\), which may depend on \(\sigma^\circ\).\(^{10}\) \(P\) designs the grand mechanism \(G = (t(x, \sigma), W(x, \sigma))\) to maximize the expected net surplus \(\pi = E[\theta + e - e^2/2 - u - W]\).

We shall consider two alternative information structures: soft information and hard information, which are defined as follows: \((\sigma^\circ|\sigma) \in \{L, \emptyset, H\}\) under soft information, and \((\sigma^\circ|\sigma) \in \{\sigma, \emptyset\}\) under hard information. That is, any manipulation is possible when information is soft, but only the suppression of information is possible when information is hard.

Information is hard for \(S\). If not, \(S\) is useless.\(^{11}\) The assumption is also realistic for two reasons: First, when contemplating manipulations that are unfavorable to \(A\), \(S\) would

\(^9\) This decision is inconsequential. The same results are obtained if instead \(A\) collects the output and pays taxes to \(P\).

\(^{10}\) Without limited liability for \(S\), the expected value of any bribes or extortion payments collected by \(S\) in equilibrium could be taken by the principal \textit{ex ante}, and the cost of corruption would vanish.

\(^{11}\) The assumption that information is hard for \(S\) is standard. See for example Tirole (1986, 1992), Laffont and Tirole (1991). Baliga (1999) shows that \(S\) may still be useful when information is soft, but relies on
feel less anxiety in hiding information than in forging information. If either action
provokes a protest from A, audit failure is more easily defended than outright fraud: An
auditor who is found to have engaged in a deliberate fraud is finished, while an auditor
who reports nothing can claim to have been merely inept or lazy. Second, manipulations
that are favorable to A may only be feasible with help from A. For example, A may
possess private information that is essential in designing a sampling program that is
favorably biased.12

Information may be hard or soft for the S/A coalition (S/A). Specifically,
information is hard for S/A with some probability \( \gamma \). It simplifies the exposition to
assume that only S/A discovers whether information is hard or soft. P knows the
probability \( \gamma \) but cannot distinguish, when observing \( \hat{\sigma} = L \), between the case of \( \sigma = L \)
and that of \( \sigma = \emptyset \), say, when information is soft.13

The timing of the game is as follows:

1. A learns \( \theta \).
2. P offers \( G \) for S and A to accept or reject. If either player rejects \( G \), the game ends.
   If both players accept \( G \), the game moves to date (3).
3. A produces \( x \).
4. S may be asked to observe \( \sigma \) and make a report \( \hat{\sigma} \). In this case, the timing includes
   the following steps:
   (4.1) S and A observe \( \sigma \) and learn whether information is hard or soft.14
(4.2) $S$ and $A$ may negotiate and enter into a side mechanism $m$, to be described in Section 4.

(4.3) $m$ is carried out.

(5) $G$ is carried out.

3 Benchmarks without corruption

It is easier to understand corruption after first considering optimal mechanisms in its absence. In particular, this will enable us to see which form of corruption, bribery or extortion, is relevant in a given state.

3.1 Second-best (no supervisor)

If either $\theta$ or $e$ were publicly observable, $A$ could be paid $\frac{1}{2}$ for one unit of effort regardless of type. This scheme is impractical when both $\theta$ and $e$ are $A$’s private information. If the principal demands a verifiable output of $1 + \theta_i$ from an agent of type $\theta_i$, $i \in \{L, H\}$, and pays $\frac{1}{2}$ regardless of output, the high type ($\theta_H$) prefers to produce $1 + \theta_L$ by exerting $e = 1 - \Delta \theta$. The contract must not only compensate each type for his disutility of effort ($t_i - e_i^2/2 = u_i \geq 0, i \in \{L, H\}$) but must also satisfy the incentive compatibility constraints $u_L \geq u_H - (e_H \Delta \theta + \Delta \theta^2/2)$, which ensures that the low type prefers to exert $e_L$, and $u_H \geq u_L + (e_L \Delta \theta - \Delta \theta^2/2)$, which ensures that the high type prefers to exert $e_H$. As usual, only the participation constraint of the low type and the incentive constraint of the high type are binding at the optimum, and $P$’s problem can be expressed as follows: Maximize $\pi = \xi(\theta_L + e_L - e_L^2/2) + (1 - \xi)(\theta_H + e_H - e_H^2/2 - e_L \Delta \theta + \Delta \theta^2/2)$ with respect to $e_L$ and $e_H$. The optimal contract with no supervisor involves $e_H = 1, e_L = 1 - \frac{1 - \xi}{\xi} \Delta \theta \equiv e_L^{NS}$, and $u_H = \Delta \theta(e_L - \Delta \theta/2) > 0$.\footnote{In order for the incentive constraint of the high type to be correct as written above, we require $e_L^{NS} > \Delta \theta$. Otherwise, the high type could costlessly mimic the low type by disposing of some of his endowment $\theta_H$. We shall assume $\xi > \Delta \theta$ in order to avoid this minor difficulty. The incentive constraint of the low type and the participation constraint of the high type are slack, since $1 > e_L^{NS} > \Delta \theta > \Delta \theta/2$. The downward distortion in the effort level of the low type is a standard result in a second-best contract. At the first-best
3.2 Free supervision without corruption

It is well-known that with correlation and unlimited penalties $P$ can implement the first-best. See Riordan and Sappington (1988) and Crémer and McLean (1988). With a maximum penalty, a separation result emerges (Baron and Besanko (1984)): As maximum penalties increase, $P$ first reduces the information rent without adjusting the second-best effort. After rent has been eliminated, further increases in maximum penalties are used to restore the effort. The standard separation result is replicated here as follows:

$P$ is required to pay to $A$ at least $(x_L - \theta_L)^2/2 = e_L^2/2$ when output is low and at least $(x_H - \theta_H)^2/2 = e_H^2/2$ when output is high, regardless of $\hat{\sigma}$ (before any penalties). Since the incentive constraint of the low type is slack at the optimum, there is no reason to employ $S$ following high output. Since supervision is free, it is without loss of generality that $P$ sends $S$ with probability 1 when output is low. When $\hat{\sigma} = H$, which occurs exclusively off the equilibrium path, $P$ collects from $A$ a penalty $P_H \leq P^*$.\(^{16}\)

When $\hat{\sigma} = \emptyset$, $P$ collects from $A$ a penalty $P_\emptyset \leq P^*$. To compensate the low type for the risk of incurring $P_\emptyset$, $P$ pays to $A$ a transfer that exceeds $e_L^2/2$ by an amount $u_L$ when $\hat{\sigma} = L$.

$P$'s problem is to choose $e_L$, $e_H$, $u_L$, $u_H$, $P_\emptyset$, and $P_H$ to maximize the surplus

$$\pi = \xi(\theta_L + e_L e_L^2/2 - \nu u_L + (1 - \nu)P_\emptyset) + (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H),$$

for nonnegative $u_L$ and $u_H$, $P_\emptyset$ and $P_H$ less than $P^*$ and nonnegative, and subject to the participation (or individual rationality) constraint of the low type (IR) and incentive compatibility constraint of the high type (IC), which are as follows:

\begin{align*}
\text{(IR)} & \quad \nu u_L - (1 - \nu)P_\emptyset \geq 0 \\
\text{(IC)} & \quad u_H \geq (e_L \Delta \theta - \Delta \theta^2/2) - (1 - \nu)P_\emptyset - \nu P_H
\end{align*}

level of effort, reducing the effort by a small amount involves a second-order loss of efficiency and a first-order reduction in the information rent. See Baron and Myerson (1982) and Maskin and Riley (1984).

\(^{16}\) Apologies for the abuse of notation. When $P$ appears by itself it is the principal. A subscripted $P$ denotes a penalty and $P^*$ is the maximum penalty.
The model is solved in Appendix A, and the results are summarized as follows:

For \( P^* < \left(1 - \left( \frac{1-\xi}{\xi} \right) \Delta \theta - \frac{\Delta \theta}{2} \right) \Delta \theta \equiv P_{RE}^{ZR} \) it is optimal to pay rent to the high type and set \( P_0 = P_H = P^* \); call this regime RE (for rent-extraction). The low effort \( e_L^{RE} = 1 - \left( \frac{1-\xi}{\xi} \right) \Delta \theta = e_L^{NS} \) and the rent \( u_H = P_{RE}^{ZR} - P^* \), which is strictly decreasing in \( P^* \) and strictly positive for \( P^* < P_{RE}^{ZR} \), demonstrating the rent-extraction part of the separation result.

For \( P_{RE}^{ZR} < P^* < \left(1 - \frac{\Delta \theta}{2} \right) \Delta \theta \equiv P_{FB}^{ZR} \) the optimal regime involves \( P_0 = P_H = P^* \), \( u_H = 0 \), and \( e_L = \frac{\Delta \theta}{2} + \frac{P^*}{\Delta \theta} \); call this regime ZR (for zero-rent). The distortion in \( e_L^{ZR} \) is decreasing in \( P^* \) and vanishes at \( P^* = P_{FB}^{ZR} \), demonstrating the effort-adjustment part of the separation result.

For \( P^* > P_{FB}^{ZR} \), the optimal regime involves the first-best allocation \( (e_L = e_H = 1) \) and no rent. Maximum deterrence is no longer optimal, and either \( P_0 \) or \( P_H \) or both may be strictly less than \( P^* \) while still satisfying (IC) strictly. For comparison with optimal contracts presented in Section 6, it will be useful to consider what happens when the principal uses higher \( P^* \) to first reduce \( P_0 \) as much as (IC) will allow. With no distortion in \( e_L \) and \( P_H = P^* \), (IC) yields \( P_0 = (P_{FB}^{ZR} - \nu P^*)/(1 - \nu) \), which is strictly positive for \( P^* < \left(1 - \frac{\Delta \theta}{2} \right) \frac{\Delta \theta}{\nu} \equiv P_{FB}^{FB} \), call this regime FB\(^0\). For \( P^* > P_{FB}^{FB} \), the first-best is achieved with \( P_0 = 0 \) and \( P_H = P_{FB}^{FB} \); call this regime FB.

We can now see that collusion is the issue when \( \sigma \in \{ \emptyset, H \} \). In either case, \( A \) would be willing to offer a bribe in order to secure \( \hat{\sigma} = L \). Extortion is not relevant because \( S \) cannot threaten unilateral action that would make \( A \) worse off relative to \( \hat{\sigma} = \sigma \). If \( \sigma = \emptyset \), no unilateral manipulation is feasible because information is hard for \( S \). If \( \sigma = H \), \( \hat{\sigma} = \emptyset \) is feasible but would never make \( A \) worse off compared to \( \hat{\sigma} = H \). Extortion is the issue when \( \sigma = L \), and framing then involves \( \hat{\sigma} = \emptyset \).

4 Bargaining
The side mechanism $m$ specifies a report $\tilde{\sigma}$ and a side transfer $s$, which is paid by $A$ to $S$ if $s > 0$. Let us suppose $s$ to be the result of bargaining between $S$ and $A$, and suppose that $s$ is determined by the solution to a static Nash bargaining program (hereafter called the axiomatic bargaining solution). 18 We shall later provide some foundation for this axiomatic approach, applying work by Binmore, Rubinstein, and Wolinsky (1986) and Abreu and Gul (2000).

Consider first the case where $\sigma = \emptyset$, when bribery is the issue. Since $S$ can only report $\sigma = \emptyset$ without help from $A$, the reservation payoffs are $-P_{\emptyset}$ for $A$ and $W_{\emptyset}$ for $S$. Let $s_{\emptyset} = \operatorname{argmax}_s \{(u_L - s) - (-P_{\emptyset})\alpha\{(W_L + s) - W_{\emptyset}\}^{1-\alpha}$, where $\alpha$ denotes the relative bargaining power of $A$. If $s_{\emptyset}$ satisfies the bargaining Individual Rationality (bIR) constraints $u_L - s \geq -P_{\emptyset}$ and $W_L + s \geq W_{\emptyset}$, $m$ involves $\tilde{\sigma} = L$ and $s = s_{\emptyset}$. If $s_{\emptyset}$ violates the (bIR) constraints, no side mechanism can be agreed upon, and $\tilde{\sigma} = \emptyset$. We can see that $m$ exists and $s = s_{\emptyset} = (1 - \alpha)(u_L + P_{\emptyset}) + \alpha(W_{\emptyset} - W_L)$ if and only if $u_L + P_{\emptyset} \geq W_{\emptyset} - W_L$.

The side transfer is not assumed to be nonnegative. If $W_L > W_{\emptyset}$, both players gain by collusion, and $S$ pays $A$ if $\alpha(W_L - W_{\emptyset}) > (1 - \alpha)(u_L + P_{\emptyset})$. If $\alpha(W_L - W_{\emptyset}) = (1 - \alpha)(u_L + P_{\emptyset})$, collusion occurs without any side transfer.

Let us next consider the case where $\sigma = L$, when extortion is the issue. If $W_{\emptyset} > W_L$, framing is sequentially rational for $S$, absent a side mechanism. The reservation payoffs are therefore the same as in the case of bribery. The bargaining solution and the (bIR) constraints are as above, and $m$ exists with $\tilde{\sigma} = L$ and $s = s_{\emptyset} = (1 - \alpha)(u_L + P_{\emptyset}) + \alpha(W_{\emptyset} - W_L)$ if and only if $u_L + P_{\emptyset} \geq W_{\emptyset} - W_L$.

If $W_{\emptyset} < W_L$, framing is not sequentially rational. Let us consider the possibility that both players attach some probability $\tau$ to $S$ carrying out the threat of framing when $W_{\emptyset} \leq W_L$. 20, 21 That is, absent $m$, both players expect $\tilde{\sigma}$ to be determined by ($\sim$) the

---

17 As is standard, $m$ is assumed to be enforceable (Tirole (1986, 1992)).
19 The case where $\sigma = H$ can be ignored without loss of generality. This will be explained in Section 5.
20 The assumption of perfectly symmetric uncertainty will be relaxed later. The assumption that $\tau$ is unresponsive to the gap between $W_{\emptyset}$ and $W_L$ is too strong, but it simplifies the exposition greatly.
lottery $\langle \tau, \{\emptyset, L\} \rangle = \tau(\hat{\sigma} = \emptyset) \oplus (1 - \tau)(\hat{\sigma} = L)$. In this case, the reservation payoffs are $(1 - \tau)u_L - \tau P_\emptyset$ for $A$ and $(1 - \tau)W_L + \tau W_\emptyset$ for $S$. Let $s_L = \operatorname{argmax}_s \{(u_L - s) - [(1 - \tau)u_L - \tau P_\emptyset]\}^\alpha (\{(W_L + s) - [(1 - \tau)W_L + \tau W_\emptyset]\})^{1-\alpha}$, which can be written more compactly as $\{\tau(u_L + P_\emptyset) - s\}^\alpha \{s + \tau(W_L - W_\emptyset)\}^{1-\alpha}$. If $s_L$ satisfies the (bIR) constraints $\tau(u_L + P_\emptyset) \geq s$ and $s + \tau(W_L - W_\emptyset) \geq 0$, $m$ involves $\hat{\sigma} = L$ and $s = s_L$. If $s_L$ violates the (bIR) constraints, no side mechanism can be agreed upon, and $\hat{\sigma} \sim \langle \tau, \{\emptyset, L\} \rangle$. But we can easily see that $s_L = \tau\{(1 - \alpha)(u_L + P_\emptyset) - \alpha(W_L - W_\emptyset)\}$, which always satisfies the (bIR) constraints because $u_L + P_\emptyset \geq 0 \geq W_\emptyset - W_L$. Therefore $m$ exists whenever $W_L \geq W_\emptyset$.

We have seen that $m$ fails to exist and framing occurs if and only if $W_\emptyset - W_L > u_L + P_\emptyset$. But this strict inequality cannot hold in equilibrium because there would then exist an $\varepsilon > 0$ such that lowering $W_\emptyset$ by $\varepsilon$ would strictly reduce $P$’s cost without affecting the equilibrium outcome. Therefore, we have the following proposition:

**Proposition 1.** $u_L + P_\emptyset \geq W_\emptyset - W_L$. (Framing is never tolerated in equilibrium.)

**Proof:** Appendix B.

In Sections 5 and 6, the equilibrium bribe is denoted by $s_\emptyset^*$, and its value is given by the axiomatic bargaining solution: $s_\emptyset^* = (1 - \alpha)(u_L + P_\emptyset) + \alpha(W_\emptyset - W_L)$. The equilibrium extortion payment is $s_L^*$, and its value is also given by the axiomatic bargaining solution: If $W_\emptyset > W_L$, then $s_L^* = s_\emptyset^*$ ($= s^*$ for short). Otherwise, $s_L^* = \tau s_\emptyset^*$. The remainder of the section is not essential to the reader’s understanding of the main results.

### 4.1 Foundations of the axiomatic bargaining solution

Let us now suppose that bargaining follows a Binmore-Rubinstein-Wolinsky (1986) process with exogenous risk of breakdown: $S$ and $A$ meet at times $0, \Delta, 2\Delta, 3\Delta, \ldots$, where $\Delta$ is the interval that elapses between meetings. At each meeting, one player proposes a side transfer $s$ for the other to accept or reject. If $s$ is accepted, $A$ pays $s$ to $S$ and $S$...
reports $\bar{\sigma} = L$. Upon rejection, the process continues. Either $S$ or $A$ makes the first offer (it does not matter which), and subsequent offers alternate. The duration of the bargaining process is uncertain, but it is common knowledge that the duration is distributed exponentially with parameter $\lambda$ (although beliefs about $\lambda$ may differ), so that the probability of the process continuing to at least $(t + 1)\Delta$ if no agreement is reached at time $t\Delta$, $t = 0, 1, 2 \ldots$, is $e^{-\lambda \Delta}$.

The players have (possibly divergent) beliefs about $\lambda$ and also about what will happen if the process ends before an agreement is reached. When beliefs differ, the players “agree to disagree” in the sense that even divergent beliefs are common knowledge. Specifically, $A$’s belief is $\Pr(\lambda = \lambda_A) = 1$ and $\Pr(\tau = \tau_A) = 1$, and $S$’s belief is $\Pr(\lambda = \lambda_S) = 1$ and $\Pr(\tau = \tau_S) = 1$, where $\tau$, recall, is the probability that $S$ reports $\bar{\sigma} = \emptyset$ (framing $A$ if $\sigma = L$) if the process ends without an agreement. The reservation payoff of player $i$ is the expected payoff of the lottery $\langle \tau_i, \{\emptyset, L\} \rangle$, $i = S, A$. All this is common knowledge.

Provided each player prefers to make an offer that the other will accept, $S$’s proposal $s^*_S$ must satisfy (1) and $A$’s proposal $s^*_A$ must satisfy (2):

1. $u_L - s_S \geq (u_L - s_A)e^{-\lambda_A \Delta} + (-\tau_A P_\emptyset + (1 - \tau_A)u_L)(1 - e^{-\lambda_A \Delta})$

2. $W_L + s_A \geq (W_L + s_S)e^{-\lambda_S \Delta} + (\tau_S W_\emptyset + (1 - \tau_S)W_L)(1 - e^{-\lambda_S \Delta})$

Appealing to Binmore, Rubinstein, and Wolinsky, there is a (unique) perfect equilibrium in which $A$ always proposes $s^*_A$ and rejects any proposal strictly greater than $s^*_S$, and $S$ always proposes $s^*_S$ and rejects any proposal strictly less than $s^*_A$, where $s^*_A$ and $s^*_S$ satisfy (1) and (2) in equality, as long as $s^*_A$ and $s^*_S$ (strictly) satisfy the (bIR) constraints. That is, as long as $\tau_A(u_L + P_\emptyset) \geq (>)\tau_S(W_\emptyset - W_L)$. If the (bIR) constraints are satisfied in equality, both players are indifferent between playing this equilibrium and taking their reservation payoffs, and there is no loss of generality in assuming that the players reach an agreement.

In the limit as $\Delta$ tends to zero, $s^*_A$ and $s^*_S$ both approach the following quantity ($s^*_S$ from above and $s^*_A$ from below):

3. $s^* = \tau_A \left( \frac{\lambda_A}{\lambda_A + \lambda_S} \right) (u_L + P_\emptyset) - \tau_S \left( \frac{\lambda_S}{\lambda_A + \lambda_S} \right) (W_L - W_\emptyset)$. 

If $W_L < W_\emptyset$, framing is sequentially rational and it is natural to suppose that beliefs converge on $\tau = 1$. If $W_L > W_\emptyset$, framing is irrational and $S$ is better off if $A$ believes that $S$ himself believes that $\tau$ is rather small. The reason is that, the larger $S$ believes $\tau$ to be, the more apprehensive he is about negotiations ending without an agreement. At the same time, $S$ would like $A$ to believe that $\tau$ is large. It also can be seen that beliefs $\lambda_i < \lambda_j$ favor player $i$, as this player is more optimistic about the chances for continued negotiation if a proposal is rejected.

If we let $\tau_S = \tau_A = \tau$ and let $\alpha = \left(\frac{\lambda_S}{\lambda_A + \lambda_S}\right)$ in (3), then this equilibrium corresponds to the axiomatic bargaining solution $s = \tau\{(1 - \alpha)(u_L + P_\emptyset) - \alpha(W_L - W_\emptyset)\}$. That is, if bargaining yields an agreement $m$, this agreement is unique and equivalent to that given by the axiomatic bargaining solution.

Although it is realistic to allow for some uncertainty on the part of $S$ as to how he will behave if bargaining ends without an agreement, the assumption of perfectly symmetric uncertainty is too strong. Therefore let us consider a similar bargaining game with asymmetric information, in the spirit of the chain-store paradox (Selten (1978), Kreps and Wilson (1982), Milgrom and Roberts (1982)) as it has been applied to bargaining by Abreu and Gul (2000). $S$ privately knows himself to be either a strategic player who never frames (unless $W_L < W_\emptyset$) or a behavioral player who always proposes $\hat{s}$, rejects any proposal strictly less than $\hat{s}$, and always frames if the process ends without an agreement. It is common knowledge that $S$ is the behavioral type with probability $\beta$. This game has a (unique) perfect equilibrium in which the strategic $S$ mimics his behavioral counterpart and $A$ always proposes $\hat{s}$ and rejects any proposal strictly greater than $\hat{s}$, provided $\hat{s} \leq (\leq)\beta(u_L + P_\emptyset)$. In order for the presence of the behavioral type to yield interesting results, the behavioral type must make a reasonable demand, such that $A$ is willing to pay $\hat{s}$ with certainty rather than run the (small—$\beta$) risk of ruin.

The special case of interest is that in which $\lim_{\Delta \to 0} \hat{s} = s^*$. Naturally, this approach cannot uniquely predict the axiomatic bargaining solution, since $\hat{s}$ is exogenous. Still, the

---

22 Off the equilibrium path, $A$ infers from any proposal $s \neq \hat{s}$ that $S$ is certainly strategic and thereafter proposes $s = 0$ and rejects any proposal $s > 0$. 
case in which $\hat{s}$ tends to $s^*$ in the limit is especially instructive. Consider the agent’s participation constraint in the limiting case. Again assuming a common value for $\tau$ and letting $\alpha = \lambda_s/(\lambda_A + \lambda_S)$, we have $\tau[(1 - \alpha)(u_L + P_0) - \alpha(W_L - W_0)] \leq \beta(u_L + P_0)$. By inspection, a rather large $\tau$ can be supported only if $\beta$ is sufficiently large. The extreme case in which $S$ has all the bargaining power and can commit to carry out irrational threats is supported only if $S$ is always the behavioral type. The strategic model thus places a reasonable upper bound on $\tau$ cum $(1 - \alpha)$ according to what one is willing to consider a reasonable value for $\beta$. On the other hand, as long as one concedes the possibility of $\beta > 0$, however small, one must also concede that $s$ is likely to be strictly positive whenever $u_L + P_0 > 0$, even for $W_L > W_0$.

5 Contracting responses to corruption

Definition. A contract is said to deter corruption (or to be corruption-proof) if (1) no strictly positive side payment changes hands in equilibrium and (2) $\hat{\sigma} = \sigma$.

One approach $P$ may take in response to the threat of corruption is to set $u_L = P_0 = W_L = W_0 = 0$. By making contracts indifferent for $\hat{\sigma} = L$ and $\hat{\sigma} = \emptyset$, $P$ takes away both the stake of collusion (when $\sigma = \emptyset$) and the teeth of extortion (when $\sigma = L$). Collusion can be costlessly deterred when $\sigma = H$ by sufficiently rewarding $S$ for $\hat{\sigma} = H$. Rewarding $\hat{\sigma} = H$ does not encourage extortion, since $S$ cannot forge $\hat{\sigma} = H$ alone, and costs $P$ nothing, since $\sigma = H$ with zero probability in equilibrium. One subtlety is that $W_H$ must not be too large, else $S/A$ would prefer to report $\hat{\sigma} = H$ and compensate the agent out of $W_H$. Let us fix $W_H = u_L + P_H$ and verify ex post that this is without loss of generality. Call the set of all such contracts (with $u_L = P_0 = W_L = W_0 = 0$ and $W_H = u_L + P_H$) $CP$.

Lemma 1. Every contract in $CP$ deters corruption.
Proof: For every contract in $CP$, $s_L^* = s_\emptyset^* = 0$, and $\hat{\sigma} = \sigma$ without loss of generality.  

We have seen in Section 3 that for $P^*$ sufficiently large ($P^* \geq P^{FB}$), the first-best is achieved using only $P_H$ and setting $u_L = P_\emptyset = 0$. We can therefore state a corollary to Lemma 1 without further proof:

**Corollary.** Corruption is costlessly deterred if $P^* \geq P^{FB}$.

Lemma 1 and its corollary help to frame the discussion that follows. Interesting cases involve $P^* < P^{FB}$. When very large penalties are not available, what is the optimal contract? Is the optimal contract ever found in the complement of $CP$? Can corruption be deterred by a contract in the complement of $CP$? Let us turn to the last question first.

Suppose $P$ would like to set $P_\emptyset > 0$ while deterring collusion. If $\gamma < 1$, this requires $W_\emptyset - W_L \geq u_L + P_\emptyset > 0$. This implies $W_\emptyset > W_L$, and $S$ now has an incentive to suppress evidence when $\sigma = L$ and report instead $\hat{\sigma} = \emptyset$ in order to claim the larger reward $W_\emptyset$. If $W_\emptyset - W_L > u_L + P_\emptyset$, $S$ will report $\hat{\sigma} = \emptyset$ whenever $\sigma \in \{L, \emptyset\}$ and $A$ cannot hope to influence $\hat{\sigma}$, but this is ruled out by Proposition 1. If $W_\emptyset - W_L = u_L + P_\emptyset$, then $s^* = W_\emptyset - W_L = u_L + P_\emptyset$ and the low type’s utility, net of $s^*$, will be $-P_\emptyset$ regardless of $\sigma$. Therefore the low type’s participation constraint can be satisfied only if $P_\emptyset = 0$, and we have shown that collusion cannot be deterred in a contract with $P_\emptyset > 0$ if $\gamma < 1$. If $P_\emptyset = 0$, there is no use in setting $u_L$, $W_L$, or $W_\emptyset$ strictly positive, and the following lemma is apparent:

**Lemma 2.** If $\gamma < 1$, optimal corruption-proof contracts all belong to the set $CP$.

**Proof:** Appendix C.

For the special case of $\gamma = 1$, collusion is trivially deterred in equilibrium. If $\tau = 0$ as well, then all corruption is costlessly deterred. However, with $\tau > 0$, extortion can be prevented (by contracts in the complement of $CP$) only by setting $W_L - W_\emptyset = \left(\frac{1-\alpha}{\alpha}\right)(u_L + P_\emptyset)$, even if $\gamma = 1$. That is, if $P$ wishes to deter extortion in a contract with

---

23 The payoff of the coalition is zero, regardless of $\hat{\sigma}$. This includes the case of $\hat{\sigma} = H$, since $W_H = P_H$. 

15
If $P_\emptyset > 0$, then $P$ must reward $S$ sufficiently for reporting information favorable to $A$ so that $s^*$ is reduced to zero. The generalization of Lemma 2 to the case of $\gamma = 1$ is postponed until Section 6.

We have still to consider two (not mutually-exclusive) classes of contracts in the complement of $CP$ that may contain optima. Contracts in either class accommodate collusion in equilibrium:

(i) $u_L + P_\emptyset > W_\emptyset - W_L \geq 0$

(ii) $u_L + P_\emptyset \geq 0$ and $W_L - W_\emptyset \geq 0$ with at least one strict inequality

Define $CA^{(i)}$ and $CA^{(ii)}$ to be the sets of (collusion-accommodating) contracts satisfying conditions (i) and (ii), respectively.

Contracts in $CA^{(i)}$ reward $S$ for $\sigma = \emptyset$ if the second inequality is strict, i.e. $W_\emptyset > W_L$ (so that $s^*_L = s^*_\emptyset = s^*$), thereby raising $s^*$ but not by so much that $A$ is unwilling to pay it. Therefore we can expect $S$ to report $\hat{\sigma} = L$ in equilibrium, regardless of $\sigma$, after obtaining $s^*$ from $A$. Since the low type then obtains the same utility, net of $s^*$, regardless of $\sigma$, no reward for the low type can be put beyond the reach of a deviant high type. Therefore it seems unlikely that an optimal contract would ever involve $W_\emptyset > W_L$. If $W_\emptyset = W_L$ then surely both should equal zero, and we can state the following lemma:

**Lemma 3.** Optimal contracts satisfying the conditions (i) involve $W_\emptyset = W_L = 0$.

**Proof:** Appendix D.

Lemma 3 allows us to restrict attention to $CA^{(ii)}$, by telling us that all contracts of interest in $CA^{(i)}$ also belong to $CA^{(ii)}$. Contracts in $CA^{(ii)}$ reward $S$ for $\hat{\sigma} = L$ if the second inequality is strict (so that $W_L > W_\emptyset$ and hence $s^*_L = \tau s^*_\emptyset$), thereby lowering both $s^*_L$ and $s^*_\emptyset$. In this case, we would expect that $W_\emptyset = 0$ (proved in Appendix E). We can therefore restrict attention to the union of $CP$ and the subset of $CA^{(ii)}$ for which $W_\emptyset = 0$.

**Lemma 4.** Define $CA$ to be the subset of $CA^{(ii)}$ for which $W_\emptyset = 0$. Then, with $CP$ defined above, optimal contracts belong to $CP \cup CA$. 

Proof: Since \( u_L + P_\emptyset \geq W_\emptyset - W_L \) according to Proposition 1, and \( W_L \geq W_\emptyset \) according to Lemmas 2 and 3, it remains only to show that \( W_L \geq W_\emptyset \Rightarrow W_\emptyset = 0 \). This is proved in Appendix E.

6 Optimal mechanisms with corruption

According to Lemma 4, we can restrict our search for optima to the set \( CP \cup CA \). Moreover, regardless of the set to which the optimal contract belongs, we can assume without loss of generality that \( \hat{\sigma} = L \) whenever \( \sigma = L \) and \( \hat{\sigma} = \emptyset \) with probability \( \gamma \) (and \( \hat{\sigma} = L \) with probability \( 1 - \gamma \) when \( \sigma = \emptyset \). The reasoning is as follows: We know that in equilibrium under a \( CA \) contract collusion occurs with probability \( 1 - \gamma \) whenever \( \sigma = \emptyset \). We know also that, although extortion occurs when \( \sigma = L \), framing never occurs. Finally we know that, although no corruption occurs in equilibrium under a \( CP \) contract, it would make no difference to any player, including the principal, if it did occur, since \( u_L = P_\emptyset = W_L = W_\emptyset = 0 \).

Therefore the expected utility of the low type in equilibrium is \( v(u_L - s_L^*) + (1 - v)(1 - \gamma)(u_L - s_\emptyset^*) - \gamma P_\emptyset \), and the expected utility of the high type when he produces low output is \( (e_L \Delta \theta - \Delta \theta^2 / 2) + (1 - v)(1 - \gamma)(u_L - s_\emptyset^*) - \gamma P_\emptyset - vP_H \), where \( s_L^* = \tau s_\emptyset^* = \tau[(1 - \alpha)(u_L + P_\emptyset) - \alpha W_L] \).

P’s problem is to choose \( e_L, e_H, u_L, u_H, P_\emptyset, P_H, \) and \( W_L \) to maximize the surplus
\[
\pi = \xi\{\theta_L + e_L - e_L^2 / 2 - (1 - \gamma(1 - v))(u_L + W_L) + \gamma(1 - v)P_\emptyset\} \\
+ (1 - \xi)(\theta_H + e_H - e_H^2 / 2 - u_H)
\]
for nonnegative efforts, \( u_L, u_H, \) and \( W_L, P_\emptyset \) and \( P_H \) nonnegative and less than \( P^* \), and subject to the individual rationality constraint of the low type (IR) and incentive compatibility constraint of the high type (IC):

(IR) \( v(u_L - s_L^*) + (1 - v)(1 - \gamma)(u_L - s_\emptyset^*) - \gamma P_\emptyset \geq 0 \)

(IC) \( u_H \geq (e_L \Delta \theta - \Delta \theta^2 / 2) + (1 - v)(1 - \gamma)(u_L - s_\emptyset^*) - \gamma P_\emptyset - vP_H \)
We are now ready to generalize Lemma 2 for all $\gamma$ and $\tau$, excepting only the trivial case of $\tau = 1 - \gamma = 0$. The interesting result is that incentive payments to supervisors are never optimal (except for $W_H$, which is not paid in equilibrium), and therefore all optimal corruption-proof contracts belong to $CP$.

**Lemma 5.** $W_L = 0$ without loss of generality.

**Proof:** Appendix E.

To see this, first notice (looking at the objective function) that $P'$s total cost exceeds the direct cost of effort by $\xi[(1 - \gamma(1 - \nu))(u_L + W_L) - \gamma(1 - \nu)P_\emptyset]$ plus $(1 - \xi)u_H$. The second term is easily recognized as the standard information rent. Since (IR) binds, the first term cannot go to $A$. It goes instead to $S$. Let us call this a *corruption rent*.

Feeding in (IR) and (IC), the corruption rent paid to $S$ is

$$
(4) \quad \xi \left\{ \frac{1-f - \gamma(1-\nu)}{f} \right\} \{P_\emptyset + \nu(1 - \tau)W_L\}, \quad 24
$$

and the information rent is

$$
(5) \quad (1 - \xi) \left\{ \left( e_L - \frac{\Delta \theta}{2} \right) \Delta \theta + \left[ \frac{\alpha \nu(1-\nu)(1-\gamma)(1-\tau)}{f} \right] W_L - \left[ \frac{f(1-\nu) - \alpha(1-\nu)(1-\gamma)}{f} \right] P_\emptyset - \nu P_H \right\},
$$

where $f = \nu(1 - \tau + \alpha \tau) + \alpha(1 - \nu)(1 - \gamma)$. From this it is plain to see that increasing $W_L$ strictly raises the corruption rent whenever $\tau < 1$ (vanishing from this term when $\tau = 1$) and strictly raises the information rent whenever $\alpha \nu(1-\nu)(1-\gamma)(1-\tau) > 0$. Although for completely hard information ($\gamma = 1$), $W_L$ vanishes from the information rent, $W_L$ does not vanish from the corruption rent unless $\tau = 1$. The intuition is that a dollar’s worth of reward reduces $u_L$ by only a fraction of a dollar if $\tau < 1$. When $\tau = 1$, $W_L$ vanishes from both rents and any value for $W_L$ is optimal, hence the proviso in the Lemma.

Khalil, Lawarrée, and Yun find that it is possible to rank extortion above bribery as the greater evil, and therefore while bribery may be tolerated in equilibrium extortion never is. The reasoning is that equilibrium bribes enter the (IC) in a helpful way – imposing an extra cost on the high type who would mimic – while extortion payments only enter the (IR) – imposing a useless extra cost on the low type for which he must be

---

24 Note that $1 - f - \gamma(1 - \nu) = (1 - \alpha)(\nu \tau + (1 - \gamma)(1 - \nu)) \geq 0$. 
compensated. Although this intuition remains valid, Proposition 2 shows that the two forms of corruption are not so easily separated in general.

Proposition 2. Suppose \( \tau > 0, \alpha < 1, \) and \( \gamma < 1. \) Bribery occurs in equilibrium if and only if extortion does also.

Proof: (if) If extortion occurs in equilibrium, the optimal contract must belong to \( CA. \) But then bribery cannot be deterred since \( \gamma < 1 \) and \( W_L = W_\emptyset = 0. \) (only if) If bribery occurs in equilibrium, the optimal contract must belong to \( CA. \) But then extortion cannot be deterred since \( \tau > 0, \alpha < 1, \) and \( W_L = W_\emptyset = 0. \)

If \( \tau = 1, \) any value for \( W_L \) is optimal, and equilibrium bribes and extortion payments both can be reduced to zero by setting \( W_L = \left( \frac{1-\alpha}{\alpha} \right) (u_L + P_\emptyset) \) in an optimal contract. In this case, collusion still occurs when \( \sigma = \emptyset \) and information is soft for S/A, but the equilibrium bribe is zero because players’ gains are in exact proportion to their respective bargaining power.

If \( \tau = 0, \) extortion is deterred in the sense that \( s_L^* = 0 \) always. The result (Khalil, Lawarrée, and Yun) that extortion is never tolerated in equilibrium seems to be limited to this special case. However, the result may be resurrected if \( \tau \) is a decreasing function of \( W_L - W_\emptyset \) (\( S \) is less likely to carry out a threat, the greater the sacrifice required to do so). For example, if \( \tau \) is strictly positive for \( \epsilon > W_L - W_\emptyset > 0 \) and zero for \( W_L - W_\emptyset \geq \epsilon, \) and if \( \epsilon \) is not too large, then Lemma 5 may fail and the optimal contract will involve \( W_L = \epsilon, \ s_L^* = 0, \) and \( s_\emptyset^* > 0. \) More general functional forms for \( \tau \) admit other possibilities. There may be a corner solution with \( W_L = 0 \) and \( \tau > 0 \) (all present results hold), or there may be an interior solution with \( W_L > 0 \) and \( \tau > 0 \) (Lemma 5 fails but Proposition 2 still holds).

Proposition 3 describes optimal mechanisms, showing that bribery and extortion may coexist in equilibrium. Optimal mechanisms are divided into three corruption-accommodating regimes and three corruption-proof regimes, including the first-best. Regimes described as “zero-rent” involve zero information rent. The corruption rent is strictly positive in all corruption-accommodating regimes.
Proposition 3. (a) The optimal regime accommodates collusion, with $P_0 > 0$, if and only if
\[ \frac{1-\xi}{\xi} > \frac{(1-\alpha)(\nu\tau+(1-\nu)(1-\gamma))}{\nu(1-\nu)(\alpha\nu+(1-\alpha)(1-\tau))} \equiv \eta \] and $P^* < \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta \right) \left(\frac{\Delta \theta}{\nu}\right) \equiv P_{ZR0}^\theta$. Let us call the parameter space satisfying these conditions the tolerance region. The first existence condition is satisfied for $\nu \in (\nu, \tilde{\nu})$ as long as $\frac{1-\xi}{\xi} > \frac{1-\tau}{(1-\sqrt{\tau})^2}$ (this condition is sufficient for all $\gamma$ and necessary only if $\gamma = 0$, as shown in the proof).

(b) Corruption-accommodating (CA) regimes are of three types:

(RE\textsuperscript{*})—rent-extraction with $P_0 = P^*$. Under this regime, $u_H > 0$ and $e_L = 1 - \frac{(1-\xi)}{\xi} \Delta \theta = e_L^{NS}$. (RE\textsuperscript{*}) is optimal for $P^* < \left(1 - \frac{\Delta \theta}{2} - \frac{1-\xi}{\xi} \Delta \theta \right) \left(\frac{f \nu}{\nu(1-\tau(1-\alpha))}\right) \Delta \theta \equiv P_{RE0}^\theta$, where $f = \nu(1-\tau(1-\alpha)) + \alpha(1-\nu)(1-\gamma)$.

(ZR\textsuperscript{*})—zero-rent with $P_0 = P^*$. Under this regime, $u_H = 0$ and $e_L = \frac{\Delta \theta}{2} + \frac{(1-\tau(1-\alpha))}{f} \nu \Delta \theta$. (ZR\textsuperscript{*}) is optimal for $P_{ZR0}^\theta < P^* < \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta \right) \left(\frac{f \nu}{\nu(1-\tau(1-\alpha))}\right) \Delta \theta \equiv P_{ZR0}^\theta$.

(ZR\textsuperscript{0})—zero-rent with $0 < P_0 < P^*$. Under this regime, $u_H = 0$ and $e_L = 1 - \eta \Delta \theta$. (ZR\textsuperscript{0}) is optimal for $P_{ZR0}^{ZR0} < P^* < P_{ZR0}^{ZR0}$.

(c) Corruption-proof (CP) regimes are of three types:

(RE\textsuperscript{0})—rent-extraction with $P_0 = 0$. Under this regime, $u_H > 0$ and $e_L = 1 - \frac{(1-\xi)}{\xi} \Delta \theta = e_L^{NS}$. (RE\textsuperscript{0}) is optimal for $P^* < \left(1 - \frac{\Delta \theta}{2} - \frac{1-\xi}{\xi} \Delta \theta \right) \left(\frac{\Delta \theta}{\nu}\right) \equiv P_{RE0}^{ZR0}$. 

(ZR\textsuperscript{0})—zero-rent with $P_0 = 0$. Under this regime, $u_H = 0$ and $e_L = \frac{\Delta \theta}{2} + \frac{\nu}{\Delta \theta} P^*$. (ZR\textsuperscript{0}) is optimal for $P_{RE0}^{ZR0} < P^* < P^{FB}$.

(FB)—first-best. For $P^* > P^{FB}$ the optimal contract is the first-best of Section 3.

Proof: Appendix E.

Figure 1 shows optimal regimes in the space $(\nu, P^*)$. The separation result (Baron and Besanko (1984)) holds in the present model. As long as the information rent is strictly positive, $e_L$ remains at the second-best level. After the information rent has been eliminated, larger maximum penalties are used to restore the effort. When corruption is
accommodated, maximum deterrence is optimal when the information rent is strictly positive ($RE^*$) and at first as effort is being restored ($ZR^*$), but non-maximum deterrence is optimal for larger values of $P^*$ ($ZR^0$). In the benchmark with free supervision without corruption, maximum deterrence was always optimal, up to the point where the first-best could be achieved. The difference is that using $P_0$ is now costly because of the corruption rent, which is strictly positive whenever $P_0 > 0$.

Figure 1. Optimal regimes in the space ($\nu, P^*$)

Kofman and Lawarrée (1993) obtain a similar result in a model in which supervision is prone to error. In their model, the supervisor’s mistakes, in combination with the need to pay the supervisor to prevent collusion, make it costly to use very large penalties. It is important to understand just how similar these results are: Incentive payments that prevent collusion are appropriately viewed as corruption rents. The

---

25 The principle of maximum deterrence was introduced by Becker (1968). Bolton (1987) provides a formal analysis.
difference is that they are paid *directly* to the supervisor as specified in the contract, while in the present model corruption rents are paid *indirectly* through illicit bribes and extortion payments. The consequence is the same whether corruption rents are paid directly or indirectly: In the presence of corruption rents, as maximum penalties increase, maximum deterrence ceases to be optimal before the first-best is achieved.

If the results of the present model share much in common with those of Kofman and Lawarrée, Figure 1 illustrates two striking differences. First, the principal stops using $P_\emptyset$ altogether if $P^*$ is sufficiently large, and, for still larger $P^*$, the first-best is achieved. Remember that supervision is not prone to error, only to failure, and therefore $\sigma = H$ occurs only off the equilibrium path. Therefore $P_H$ is never collected – and $W_H$ is never paid – in equilibrium.\(^{26}\) As $P^*$ becomes large, the (costless) threat of $P_H$ becomes more effective and the principal relies less on the costly $P_\emptyset$. For $P^*$ sufficiently large, the first-best is implemented with $P_\emptyset = 0$, as we saw in Section 3. Second, corruption is not tolerated if $\nu$ is too small (as usual) but now also if $\nu$ is too large. The reason that the supervisor is ignored (unless $\delta = H$) when $\nu$ is large is due to the extortion component of the corruption rent. In equilibrium, $S$ receives a bribe with probability $\xi (1 - \nu)$ and receives an extortion payment with probability $\xi \nu$. Thus, increasing $\nu$ increases the cost of extortion and decreases the cost of bribery, and the net effect may be to increase the total corruption rent if $\tau$ is rather large and $\gamma$ is rather small. The higher are extortion payments in equilibrium, the larger is the gap between $\overline{\nu}$ and 1 (the gap vanishes if $\tau = 0$). The tolerance region is responsive not only to $\tau$, but also to $\gamma$ and $\alpha$, as summarized by Proposition 4.

*Proposition 4.* The tolerance region strictly expands as $\gamma$ increases, as $\tau$ decreases, and as $\alpha$ increases, except in the following special cases: (a) If $\gamma = 0$, the tolerance

---

\(^{26}\) Recall that in Section 5 we fixed $W_H = u_L + P_H$. With Lemma 1, we were able to see that this was without loss of generality for corruption-proof contracts. We can easily see that there is also no loss of generality for corruption-accommodating contracts: The coalition obtains a total payoff of $u_L$ if $\delta = L$, $-P_\emptyset$ if $\delta = \emptyset$, and $W_H - P_H = u_L$ if $\delta = H$. Since the coalition is indifferent between reporting $\delta = L$ and $\delta = H$ (and strictly prefers either to reporting $\delta = \emptyset$) it is without loss of generality that no collusion occurs when $\sigma = H$ and collusion leads to $\delta = L$ when $\sigma = \emptyset$.  


region is unresponsive to $\alpha$. (b) If $\alpha = 1$, the tolerance region is not responsive to either $\gamma$ or $\tau$. (c) If $\tau = 1$ and $\alpha = 0$, then the tolerance region is empty.

**Proof:** From Proposition 3, optimal contracts accommodate corruption if $1 - \xi > \eta \xi$ and $P^* < \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta \right) \left(\frac{\Delta \theta}{\nu}\right) = P^{ZR^0}_{\Delta \theta}$, where $\eta = \frac{(1-a)(\nu(1-\nu)(1-\gamma))}{\nu(1-\nu)(\alpha y+(1-a)(1-\tau))}$. Therefore, the tolerance region expands as $\eta$ decreases, and $\eta$ is increasing in $\tau$ and decreasing in $\gamma$ by inspection. It can also be seen that $\frac{\partial \eta}{\partial \alpha} = \frac{-\gamma [\nu(1-\nu)(1-\gamma)]}{2(1-\nu)(\alpha y+(1-a)(1-\tau))} \leq 0$, with equality only for $\gamma = 0$. ■

Vafai (2005) assumes that $S$ has all of the bargaining power and is able to commit to threats. This can be seen as an extreme case where $S$ is ignored unless $\delta = H$, even if information is hard for $S/A$: If $\tau = 1$ and $\alpha = 0$, then (Proposition 4.c) corruption is never tolerated. To see this, notice that the optimal contract belongs to $CP$ if $1 - \xi < \xi \eta$. But if $\tau = 1$ and $\alpha = 0$ then $\eta = \infty$. Here is good reason to think that this set of assumptions may be too extreme. By the time $S$ succeeds in establishing such a formidable reputation in negotiations (arguably no mean feat), $S$ has long since conceded all of the corruption rents associated with corruption-accommodating contracts, because $P$ has long since ceased to tolerate corruption.

Khalil, Lawarrée, and Yun (2010) find that the principal unambiguously prefers to lower $\alpha$. In their model, information is soft for $S/A$ ($\gamma = 0$). According to Proposition 4, the principal may prefer to raise $\alpha$ if $\gamma > 0$. To see this, fix parameters such that the optimal contract is $G_0 \in CP$ for some $\alpha_0$ and $G' \in CA$ for some $\alpha' > \alpha_0$. Since outcomes implemented by contracts in $CP$ are indifferent to $\alpha$, $G'$ must be preferred to $G_0$.

The remainder of the section analyzes a new tradeoff, with respect to $\alpha$, between the information rent and the corruption rent. Increasing $\alpha$ raises the information rent and reduces the corruption rent, and the overall effect on the principal’s welfare is ambiguous: Under stronger institutions (higher $\gamma$, $\nu$, and $P^*$) the principal prefers strong agents, and under weaker institutions the principal prefers strong supervisors.

Recall (from (4) and (5), now with $W_L = 0$) that the information rent is $(1 - \xi) \left\{ \left( e_L - \frac{\Delta \theta}{2} \right) \Delta \theta - \left[ \frac{f(1-\nu)-\alpha(1-\nu)(1-\gamma)}{\nu} \right] P_0 - \nu P_H \right\}$ and the corruption rent is...
\[ \xi \left\{ \frac{1-f-y(1-y)}{f} \right\} P_\theta, \] where \( f = \nu(1-\tau + \alpha \tau) + \alpha(1-\nu)(1-\gamma). \) Under the regime \( RE^* \), \( e_L \) is constant and \( P_\emptyset = P_H = P^* \), and the information and corruption rents are \( (1-\xi) \left\{ (e_L - \frac{\Delta \theta}{2}) \Delta \theta - \left[ 1 - \frac{\alpha(1-\nu)(1-y)}{f} \right] P^* \right\} \equiv \zeta_J \) and \( \xi \left\{ \frac{1-f-y(1-y)}{f} \right\} P^* \equiv \zeta_C \), respectively.

Increasing \( \tau \) is easily seen to reduce \( f \) and thereby increase both rents. Increasing \( \gamma \) can be seen to reduce both rents (strictly as long as \( 0 < \alpha < 1 \)) as follows:

\[
\frac{\partial \zeta_C}{\partial \gamma} = -\xi P^* \left\{ \nu(1-\nu)(1-\alpha)(1-\tau) \right\} f^2 \leq 0
\]

\[
\frac{\partial \zeta_I}{\partial \gamma} = -(1-\xi) P^* \left\{ \alpha \nu(1-\nu)(1-\tau + \alpha \tau) \right\} f^2 \leq 0
\]

The changes in \( \zeta_C \) and \( \zeta_J \) with respect to \( \alpha \) are as follows:

\[
\frac{\partial \zeta_C}{\partial \alpha} = -\xi P^* \left\{ (1-\gamma)(1-\nu) \right\} \left\{ \nu \tau + (1-\gamma)(1-\nu) \right\} f^2 \leq 0
\]

\[
\frac{\partial \zeta_J}{\partial \alpha} = (1-\xi) P^* \left\{ \nu(1-\nu)(1-\gamma)(1-\tau) \right\} f^2 \geq 0
\]

It can be seen that increasing \( \alpha \) reduces the corruption rent at the cost of raising the information rent. The net effect of increasing \( \alpha \) is strictly rent-reducing if and only if the following condition (7) is satisfied in a \( CA \) regime.

\[
(7) \quad \frac{1-\xi}{\xi} < \frac{(1-\gamma)(1-\nu)(\nu \tau + (1-\gamma)(1-\nu))}{\nu(1-\nu)(1-\gamma)(1-\tau)}.
\]

Condition (7) is very intuitive at a high level: Since the information rent is paid with probability \( 1 - \xi \) and the corruption rent is paid with probability \( \xi \), it is natural that reducing the corruption rent is the more urgent problem when \( \xi \) is rather large. Recall that corruption-accommodating regimes are optimal for \( \frac{1-\xi}{\xi} > \eta = \frac{(1-\alpha)(\nu \tau + (1-\gamma)(1-\nu))}{\nu(1-\nu)(\alpha \nu + (1-\alpha)(1-\tau))} \). Hence, a necessary condition for the existence of equilibria in which \( \alpha \) strictly lowers rent is that the right-hand side of (7) strictly exceed \( \eta \), which is equivalent to \( \gamma > 0 \). Before considering the general case in which both \( \gamma \) and \( \tau \) are strictly positive, some initial insight can be gleaned from the following three special cases:

---

27 It can be seen that \( \zeta_C \) is independent of \( \gamma \) if \( \tau = 1 \), and the inequality is then not strict even for \( \alpha < 1 \).

28 As long as we are not in the special case of Proposition 1.1.c (where \( \alpha = 1 - \tau = 0 \) and corruption is never tolerated), the right hand side of (7) is strictly greater than \( \eta \) if and only if \( \gamma \left( \alpha(1-\gamma)(1-\nu) + \nu(1-\alpha)(1-\tau) \right) > 0 \), but this is true if and only if \( \gamma > 0 \).
First, when \( \tau = \gamma = 0 \), then (when \( P_0 > 0 \)) the net effect of \( \alpha \) is rent-reducing only for \( \frac{1-\xi}{\xi} < \frac{1}{v} \), but it is optimal to set \( P_0 > 0 \) only for \( \frac{1-\xi}{\xi} > \frac{1}{v} \). Hence \( \alpha \) appears to always be rent-increasing. Indeed, \( \alpha \) is always rent-increasing, but only because it is never optimal to set \( P_0 > 0 \) in the region where, if you did have \( P_0 > 0 \), \( \alpha \) would be rent-reducing.

Second, the same is true if \( \tau > \gamma = 0 \): When \( P_0 > 0 \), the net effect of \( \alpha \) is rent-reducing if \( \frac{1-\xi}{\xi} < \frac{1-v(1-\tau)}{v(1-v)(1-\tau)} \), and it is optimal to set \( P_0 > 0 \) for \( \frac{1-\xi}{\xi} > \frac{1-v(1-\tau)}{v(1-v)(1-\tau)} \). Hence, again, \( \alpha \) is never rent-reducing in equilibrium.

Third and finally, suppose \( \gamma > \tau = 0 \). When \( P_0 > 0 \), the net effect of \( \alpha \) is rent-reducing if \( \frac{1-\xi}{\xi} < \frac{1}{v} - \frac{\gamma}{v} (1-v) \), and it is optimal to set \( P_0 > 0 \) for \( \frac{1-\xi}{\xi} > \frac{1}{v} - \frac{\gamma}{v} (1-v) \). But \( \frac{1}{1-\alpha(1-\gamma)} > (1-v) \) and so now the effect of \( \alpha \) depends on \( \gamma \). Since \( \frac{1}{v} - \frac{\gamma}{v} (1-v) \) is decreasing in \( \gamma \), \( \alpha \) is rent-reducing for low \( \gamma \) and rent-increasing for high \( \gamma \) as long as \( \frac{1-\xi}{\xi} < \frac{1}{v} \) — for \( \frac{1-\xi}{\xi} > \frac{1}{v} \), \( \alpha \) is always rent-increasing.

Figure 2 illustrates the relationship between the rent-reducing direction of \( \alpha \) and the other parameters of the model for the general case when both \( \tau \) and \( \gamma \) are strictly positive. The direction of the arrows indicates the direction in which \( \alpha \) should move to lower rent – heuristically, the direction in which the incentive to lower rent should be expected to exert pressure on \( \alpha \) in the optimal design of institutions. The size of the arrow indicates the magnitude of the effect. The top panel shows \( \left( \frac{(1-v)(1-\gamma)}{v} \right)^2 \) for various values of \( v \). If \( \tau \) is below (above) the line, then the right hand side of (7) is decreasing (increasing) in \( \gamma \). The bottom panel shows \( \eta \) (lower line) and the right hand side of (7) (upper line) corresponding to \( v = \frac{1}{2} \). Raising \( \alpha \) may either increase or decrease rent. Hardening information for \( S/A \) (increasing \( \gamma \)) may first strengthen and then weaken the rent-increasing effect of \( \alpha \), eventually changing the direction of the effect so

\( ^{29} \) The special case of \( \gamma = 0 \) can still be seen on the left vertical axis in the lower diagram. There, the two curves intersect at \( \frac{1-\xi}{\xi} = \frac{1-v(1-\tau)}{v(1-v)(1-\tau)} \).
that increasing $\alpha$ lowers rent, after which further hardening strengthens the rent-reducing effect of $\alpha$.\textsuperscript{30}

Figure 2. The rent-reducing direction of $\alpha$

There are essentially two institutional weaknesses through which corruption gives rise to rent: $\tau$ (through which extortion in equilibrium contributes to rent) and $1 - \gamma$ (through which collusion in equilibrium contributes to rent).\textsuperscript{31} The relationship between

\textsuperscript{30}It is even possible, for low starting levels of $\gamma$ and $\frac{1-\xi}{\xi} < \frac{1-\gamma(1-\tau)}{\nu(1-\tau)(1-\gamma)}$, to first see benefits of higher $\alpha$ attenuate and vanish as $\gamma$ increases, only to reemerge for sufficiently high $\gamma$.

\textsuperscript{31}It has been pointed out that extortion has bite even if $\tau = 0$. The lesson of Khalil, Lawarrée, and Yun is exactly that: The potential for extortion is what leads to collusion in equilibrium, even if extortion does not itself produce strictly positive
\( \alpha \) and rent – and the subtle way in which \( \gamma \) affects this relationship – can be explained by the relative magnitudes of these two weaknesses.

First observe that increasing \( \alpha \) increases \( f \), and hence decreases both rents—through a term \( \alpha \nu \tau \) (the extortion rent effect), and through a term \( \alpha(1 - \nu)(1 - \gamma) \) (the collusion rent effect). Next observe that \( \alpha \) appears as a multiplier in a penalty-reducing (hence rent-increasing) term of the information rent, \( \alpha(1 - \nu)(1 - \gamma)/f \) (the direct information rent effect). Notice that the direct information rent effect is tempered by the two rent-reducing effects that operate through \( f \).

Only the rent-reducing effects apply to the corruption rent, which is therefore decreasing in \( \alpha \). The direct information rent effect overpowers the two rent-reducing effects in the information rent, which is therefore increasing in \( \alpha \). The net effect of \( \alpha \) may be either to increase or decrease total rent, as shown in Figure 2. Moreover, as long as \( \tau > 0 \), the extortion rent effect persists as \( \gamma \) eventually drives out the collusion rent and direct information rent effects, hence for \( \gamma \) sufficiently high the net effect of \( \alpha \) is to reduce rent. However, it is also possible for \( \alpha \) to be rent-reducing for \( \gamma \) sufficiently low, because when \( (1 - \gamma) \) is large the collusion rent effect is large and the direct information rent effect, while larger for \( (1 - \gamma) \) in the numerator, is tempered by larger \( f \) in the denominator.

Increasing \( \gamma \) weakens both the collusion rent effect and the direct information rent effect. However, when \( (1 - \gamma) \) is large relative to \( \tau \), so that \( \alpha \) works more through the collusion rent effect in \( f \), increasing \( \gamma \) does relatively less to weaken the information rent effect (through which \( \alpha \) hurts) and does relatively more to weaken the collusion rent effect (through which \( \alpha \) helps), with the result that \( \alpha \) becomes less helpful or more unhelpful as \( \gamma \) increases. When instead \( (1 - \gamma) \) is small relative to \( \tau \), so that \( \alpha \) works more through the extortion rent effect in \( f \), increasing \( \gamma \) does relatively more to weaken

\footnote{side transfers. If \( \gamma = 1 \) and \( \tau > 0 \), extortion payments (and thus corruption rents) are strictly positive although collusion is infeasible.}
the information rent effect and does relatively less to weaken the collusion rent effect, and \( \alpha \) becomes less unhelpful or more helpful as \( \gamma \) increases.\(^{32}\)

Figure 3 shows the effect of \( \nu \) on the shape of the right hand side of (7) — the U in the bottom panel of Figure 2. Lower accuracy in the supervision technology (smaller \( \nu \)) tends to enlarge the region where rent is increasing in \( \alpha \). Increasing \( \nu \) can be seen to have an effect similar to that of increasing \( \gamma \): We move from inside the U where \( \alpha \) is harmful to right of the U where \( \alpha \) is helpful either by moving eastward while the U holds still or by staying put while the U rises to the left.

\[
\begin{align*}
\frac{1 - \xi}{\xi} & \quad \nu > \frac{1}{2} \\
\nu = \frac{1}{2} & \quad \nu < \frac{1}{2}
\end{align*}
\]

Figure 3. Right hand side of (7) for various \( \nu \)

Under \( ZR^* \) and \( ZR^\emptyset \), the tradeoff with respect to \( \alpha \) between informational and corruption rents no longer exists (since information rent is eliminated). However, under \( ZR^* \) there is still a tradeoff with respect to \( \alpha \) between rent and efficiency: Both corruption rent and \( e_L \) are decreasing with respect to \( \alpha \). Under \( ZR^\emptyset \), where maximum deterrence is no longer optimal, the tradeoff disappears, and \( e_L \) is increasing (and the principal’s welfare is unambiguously increasing) with respect to \( \alpha \).

\(^{32}\) Most of the foregoing results are robust to an alternative construction of hard information in which \( \gamma \) enters only (IC) and not (IR), as if forging evidence of low type is made more difficult for \( S/A \) only if \( A \)’s type is actually high. It can be shown that increasing \( \gamma \) can only make \( \alpha \) less unhelpful (when \( \alpha \) is rent-increasing for low \( \gamma \) or more helpful (when \( \alpha \) is rent-reducing for high \( \gamma \) or for \( \nu < \xi/(1 - \xi) \)). Simply put, the bottom of the U sits at \( \gamma = 0 \). Details are available from the author.
Assembling these results, it can be said that under inchoate institutions – plausibly characterized by inaccurate and easily manipulable information and lower penalties (analogous to weaker outside opportunities) – principals should prefer that supervisors have greater relative bargaining power, while under better-developed institutions principals should want to empower agents.\textsuperscript{33} Reform efforts aimed at, e.g., improving oversight of supervisors and providing appeals processes for agents may face a challenge greater than the obvious difficulty of insuring that these new measures do not themselves fall into corruption.\textsuperscript{34} Principals may indirectly benefit from the relative power of supervisors over agents when institutions are weak. Focusing first on hardening information and improving the accuracy of monitoring technology may make principals more receptive to ancillary efforts to rein in abuses by supervisors. Policies that simultaneously harden information and shift the balance of power in favor of agents may be particularly attractive.\textsuperscript{35}

7 Conclusion

The paper yields the novel insight, stemming from the tradeoff, with respect to bargaining power, between informational and corruption rents, that the principal’s welfare may actually increase with the bargaining power of the agent. This is the case when maximum penalties (analogous to outside opportunities) are high or when supervision is more accurate and yields information that is less easily manipulated. The more general lesson is that stronger institutions favor strong agents while weaker

\textsuperscript{33} Maximum penalties might be lower in the absence of institutions that facilitate the seizure of assets and the garnishing of wages. For that matter, people in developing economies have fewer assets to seize and lower wages to garnish (although these shortcomings may create demand for harsh non-monetary penalties). Raising the maximum penalty has essentially the same effect as raising the reservation utility of the agent – to allow the principal to rely more heavily on $P_H$ and less on $P_\tilde{\sigma}$, which lessens the bite of corruption since collusion is costlessly prevented when $\sigma = H$ and $\tilde{\sigma} = H$ cannot be forged. See Khalil, Lawarrée, and Yun (2010).

\textsuperscript{34} Both the UNODC and Transparency International studies cited in the introduction highlight widespread pessimism regarding governments’ effectiveness at combating corruption by such means.

\textsuperscript{35} India’s Right to Information Law might be viewed as an example of one such policy. If so, it highlights the potential to harden information by astonishingly simple means: Polgreen (2010) reports that “India’s 1.2 billion citizens have been newly empowered by the far-reaching law granting them the right to demand almost any information from the government [, which has] clearly begun to tilt the balance of power, long skewed toward bureaucrats and politicians.”
Institutions favor strong supervisors. When monitoring is less sophisticated (less accurate and yielding information that is more easily manipulated) and maximum penalties are low, larger side transfers are actually preferred and policies aimed at checking the ability of supervisors to extract such transfers are likely to be less in demand. The mandate to restrain supervisors’ rent-seeking does not arise until monitoring technology is sufficiently advanced or penalties are sufficiently large.

A concrete notion of bargaining power (both in a corruption setting and more generally) remains elusive. Studying bribery and extortion at checkpoints in the Indonesian province of Aceh, Olken and Barron (2009) find “factors that increase the bargaining power of the officer manning the checkpoint, such as whether he is brandishing a gun, and the number of officers who are visible and could provide backup if trouble arose, increase the equilibrium payment.” The difficulty in distinguishing between the determinants of bargaining power and the determinants of reservation payoffs is clear in this account.36 Theory and evidence further illuminating the determinants of bargaining power would be a worthwhile direction for future research.

Several strong assumptions were made in order to model bargaining in an informative way, and many important aspects of anticorruption policy (such as appeals processes, competition among supervisors, and efficiency wages, to name just a few) were set aside in order to focus on a few key ideas. The present framework is therefore but a step toward a more general theory of information-based corruption.

---

36 Svejnar (1986) emphasizes this issue as well: “It must be stressed that (some of) the same exogenous factors may affect bargaining power as well as the threat point (disagreement outcome) and that this simultaneous effect does not obviate the need for the concept of bargaining power […]. Suppose that a nonunion wage serves as the union’s threat point and that the legalization of a closed shop (or the right to strike) affects the nonunion wage through threat or spillover effects. It is quite likely that this change in the legal environment also affects the union’s bargaining power, i.e. its ability to realize a gain over and above the threat point. It is for instance quite possible that even if the nonunion wage rises, the union–nonunion wage differential increases as well, ceteris paribus. Accounting properly for movements in the threat point is hence important, but it may not be sufficient for identifying changes in the division of the subject of bargaining that occur above the threat point.”
References


Appendix

Appendix A. Free incorruptible supervision

The principal solves the following constrained maximization problem:

$$\max_{e,u,P} \pi = \xi(\theta_L + e_L - e_L^2/2 - \nu u_L + (1 - \nu)P_\emptyset)$$
$$+ (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H) s.t.$$ (IR) $$\nu u_L - (1 - \nu)P_\emptyset \geq 0$$ (IC) $$u_H \geq (e_L\Delta\theta - \Delta\theta^2/2) - (1 - \nu)P_\emptyset - \nu P_H$$

and $$P_\emptyset \leq P^*; P_H \leq P^*$$. The Lagrangian for P’s problem is the following:

$$\mathcal{L} = \xi(\theta_L + e_L - e_L^2/2 - \nu u_L + (1 - \nu)P_\emptyset) + (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H)$$
$$+ \lambda_1\{\nu u_L - (1 - \nu)P_\emptyset\} + \lambda_2\{u_H - (e_L\Delta\theta - \Delta\theta^2/2) + (1 - \nu)P_\emptyset + \nu P_H\}$$
$$+ \lambda_3\{P^* - P_\emptyset\} + \lambda_4\{P^* - P_H\}$$

The Kuhn-Tucker conditions for maximization are the following (plus the constraints and their complementary slackness conditions):

(A.1) $$\frac{\partial \mathcal{L}}{\partial e_L} = \xi(1 - e_L) - \lambda_2\Delta\theta \leq 0; e_L \frac{\partial \mathcal{L}}{\partial e_L} = 0$$

(A.2) $$\frac{\partial \mathcal{L}}{\partial e_H} = (1 - \xi)(1 - e_H) \leq 0; e_H \frac{\partial \mathcal{L}}{\partial e_H} = 0$$

(A.3) $$\frac{\partial \mathcal{L}}{\partial u_L} = -\xi\nu + \lambda_1\nu \leq 0; u_L \frac{\partial \mathcal{L}}{\partial u_L} = 0$$

(A.4) $$\frac{\partial \mathcal{L}}{\partial u_H} = -(1 - \xi) + \lambda_2 \leq 0; u_H \frac{\partial \mathcal{L}}{\partial u_H} = 0$$

(A.5) $$\frac{\partial \mathcal{L}}{\partial P_\emptyset} = \xi(1 - \nu) - \lambda_1(1 - \nu) + \lambda_2(1 - \nu) - \lambda_3 \leq 0; P_\emptyset \frac{\partial \mathcal{L}}{\partial P_\emptyset} = 0$$

(A.6) $$\frac{\partial \mathcal{L}}{\partial P_H} = \lambda_2\nu - \lambda_4 \leq 0; P_H \frac{\partial \mathcal{L}}{\partial P_H} = 0$$

The following statements characterize optimal contracts:

(a) $$e_H = 1$$. Immediate from (A.2).

(b) $$e_L = 1 - \left(\frac{\lambda_2}{\xi}\right)\Delta\theta$$. The statement follows from (A.1) since $$\lambda_2 \leq (1 - \xi)$$ (from (A.4)) and $$\xi > \Delta\theta$$ is assumed.
(c) The optimal contract is the first-best for \( P^* > \left(1 - \frac{\Delta \theta}{2}\right) \Delta \theta \equiv P_{ZR}^{FB} \). Proof: \( (IC) \) is satisfied with \( u_H = 0 \) for \( e_L = e_H = 1 \) and \( P_0 = P_H = P_{ZR}^{FB} \).

(d) \( \lambda_2 > 0 \) for \( P^* < P_{ZR}^{FB} \). Proof: If \( \lambda_2 = 0 \), then \( u_H = 0 \) (from (A.4)) and \( e_L = 1 \) (from (b)). But \( \left(1 - \frac{\Delta \theta}{2}\right) \Delta \theta - (1 - \nu)P_0 - \nu P_H > 0 \) for both \( P_0 \) and \( P_H \) less than \( P^* < P_{ZR}^{FB} \), which violates \( (IC) \).

(e) \( P_0 = P_H = P^* \) for \( P^* < P_{ZR}^{FB} \). Proof: \( \lambda_2 > 0 \) from (d). \( \lambda_2 > 0 \) implies \( \lambda_3 > 0 \) from (A.3). Therefore \( \lambda_2 > 0 \) implies \( \lambda_4 > 0 \) from (A.6). \( \lambda_2 > 0 \) from (d).

(f) The optimal contract for \( P_{ZR}^{FB} > P^* > P_{RE}^{ZR} \equiv \left(1 - \left(1 - \frac{\xi}{\xi}\right) \Delta \theta - \frac{\Delta \theta}{2}\right) \Delta \theta \) involves \( u_H = 0 \) (call this regime \( ZR \), for zero-rent), and the optimal contract for \( P^* < P_{RE}^{ZR} \) involves \( u_H > 0 \) (call this regime \( RE \), for rent-extraction). Proof: \( (d) \) implies that \( (IC) \) binds, and \( (e) \) implies \( P_0 = P_H = P^* \). From (A.4) we have \( \lambda_2 \leq (1 - \xi) \) and \( u_H > 0 \) \( \Rightarrow \lambda_2 = (1 - \xi) \Rightarrow u_H = P_{RE}^{ZR} - P^* \) (from \( (IC) \) and \( (b) \)), and \( u_H = 0 \) implies \( P^* = \left(1 - \left(\frac{\lambda_2}{\xi}\right) \Delta \theta - \frac{\Delta \theta}{2}\right) \Delta \theta \), or \( \lambda_2 = \left(\frac{\xi}{\lambda_0}\right) \left(1 - \frac{\Delta \theta}{2} - \frac{P^*}{\Delta \theta}\right) \), which is strictly positive for \( P^* < P_{ZR}^{FB} \) and strictly less than \( (1 - \xi) \) for \( P^* > P_{RE}^{ZR} \).

(g) \( e_{RR}^L = 1 - \left(\frac{1 - \xi}{\xi}\right) \Delta \theta = e_{NS}^L \), and \( e_{RR}^L = \frac{\Delta \theta}{2} + \frac{P^*}{\Delta \theta} \). Immediate from (a), (b), and (f). \( \blacksquare \)

Appendix B. Optimal contracts with \( W_0 - W_L \geq u_L + P_0 \)

Consider \( P^* \)’s problem when framing occurs:

\[
\max_{e,u,P,W} \pi = \xi \left(\theta_L + e_L - e_L^2 / 2 - (W_0 - P_0)\right) + (1 - \xi) \left(\theta_H + e_H - e_H^2 / 2 - u_H\right) \quad s.t.
\]

\( (IR) \quad -P_0 \geq 0 \)

\( (IC) \quad u_H \geq (e_L \Delta \theta - \Delta \theta^2 / 2) - (1 - \nu)P_0 - \nu P_H \)

\( (F) \quad W_0 - W_L \geq u_L + P_0 \)

\quad and \( P_0 \leq P^* ; P_H \leq P^* \).

\( (F) \) must be satisfied for framing to occur in equilibrium, hence collusion is deterred.

\( (IR) \) requires \( P_0 = 0 \) and \( P^* \)’s problem can be rewritten:
\[
\max_{e,u,p,w} \pi = \xi(\theta_L + e_L - e_L^2/2 - W_0) + (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H) \quad \text{s.t.}
\]

\begin{align*}
\text{(IC)} & \quad u_H \geq (e_L \Delta \theta - \Delta \theta^2/2) - \nu P_H \\
\text{(F)} & \quad W_0 - W_L \geq u_L \\
\text{and} & \quad P_H \leq P^*.
\end{align*}

The Lagrangian for \( P \)'s problem is the following:
\[
\mathcal{L} = \xi(\theta_L + e_L - e_L^2/2 - W_0) + (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H) + \lambda_1 (u_H - (e_L \Delta \theta - \Delta \theta^2/2) + \nu P_H) + \lambda_2 (W_0 - W_L - u_L) + \lambda_3 (P^* - P_H)
\]

The Kuhn-Tucker conditions for maximization include the following:

\begin{align*}
\text{(B.1)} & \quad \frac{\partial \mathcal{L}}{\partial u_L} = -\lambda_2 \leq 0; \quad u_L \frac{\partial \mathcal{L}}{\partial u_L} = 0 \\
\text{(B.2)} & \quad \frac{\partial \mathcal{L}}{\partial W_L} = -\lambda_2 \leq 0; \quad W_L \frac{\partial \mathcal{L}}{\partial W_L} = 0 \\
\text{(B.3)} & \quad \frac{\partial \mathcal{L}}{\partial W_0} = -\xi + \lambda_2 \leq 0; \quad W_0 \frac{\partial \mathcal{L}}{\partial W_0} = 0 \\
\text{(B.4)} & \quad \frac{\partial \mathcal{L}}{\partial \lambda_2} = W_0 - W_L - u_L \geq 0; \quad \lambda_2 \frac{\partial \mathcal{L}}{\partial \lambda_2} = 0
\end{align*}

Proof of Proposition 1:

\begin{enumerate}
\item[(a)] \( u_L = 0. \) Proof: \( u_L > 0 \Rightarrow \lambda_2 = 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial W_0} < 0 \Rightarrow W_0 = 0. \) \( \Leftarrow \) (B.4)
\item[(b)] \( W_L = 0. \) Proof: \( W_L > 0 \Rightarrow \lambda_2 = 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial W_0} < 0 \Rightarrow W_0 = 0. \) \( \Leftarrow \) (B.4)
\item[(c)] \( W_0 = 0. \) Proof: \( W_0 > 0 \Rightarrow \lambda_2 = \xi \Rightarrow \frac{\partial \mathcal{L}}{\partial \lambda_2} = 0 \Rightarrow u_L + W_L > 0. \) \( \Leftarrow \) (a) + (b)
\end{enumerate}
Appendix C. Optimal corruption-proof contract

Consider $P$’s problem when collusion is to be deterred:

$$\max_{e,u,P,W} \pi = \xi \left( \theta_L + e_L - e_L^2/2 - \nu(u_L + W_L) - (1 - \nu)(W_\theta - P_\theta) \right) + (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H) \text{ s.t.}$$

(IR) $\nu(u_L - s_L^*) - (1 - \nu)P_\theta \geq 0$, where $s_L^* = (1 - \alpha)(u_L + P_\theta) + \alpha(W_\theta - W_L)$

(IC) $u_H \geq (e_L \Delta \theta - \Delta \theta^2/2) - (1 - \nu)P_\theta - \nu P_H$

(CP) $W_\theta - W_L \geq u_L + P_\theta$

and $P_\theta \leq P^*; P_H \leq P^*$.

(CP) cannot be strict without leading to framing, which is ruled out by Proposition 1. Therefore (CP) implies $s_L^* = u_L + P_\theta$, hence (IR) requires $P_\theta = 0$.

With $P_\theta = 0$ and $u_L = W_\theta - W_L$, $P$’s problem can be rewritten:

$$\max_{e,u,P,W} \Pi = \xi \left( \theta_L + e_L - e_L^2/2 - W_\theta \right) + (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H) \text{ s.t.}$$

(IC) $u_H \geq (e_L \Delta \theta - \Delta \theta^2/2) - (1 - \nu)P_\theta$

(CP) $W_\theta - W_L \geq u_L + P_\theta$

and $P_\theta \leq P^*; P_H \leq P^*$.

The Lagrangian for $P$’s problem is the following:

$$\mathcal{L} = \xi \left( \theta_L + e_L - e_L^2/2 - W_\theta \right) + (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H) + \lambda_1 \{u_H - (e_L \Delta \theta - \Delta \theta^2/2) + \nu P_H \} + \lambda_2 \{W_\theta - W_L \} + \lambda_3 \{P^* - P_H \}$$

The Kuhn-Tucker conditions for maximization include the following:

(C.1) $\frac{\partial \mathcal{L}}{\partial W_\theta} = -\xi + \lambda_2 \leq 0; \ W_\theta \frac{\partial \mathcal{L}}{\partial W_\theta} = 0$

(C.2) $\frac{\partial \mathcal{L}}{\partial W_L} = -\lambda_2 \leq 0; \ W_L \frac{\partial \mathcal{L}}{\partial W_L} = 0$

(C.3) $\frac{\partial \mathcal{L}}{\partial \lambda_2} = W_\theta - W_L \geq 0; \ \lambda_2 \frac{\partial \mathcal{L}}{\partial \lambda_2} = 0$

Proof of Lemma 2:

(a) $W_L = 0$. Proof: $W_L > 0 \Rightarrow \lambda_2 = 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial W_\theta} < 0 \Rightarrow W_\theta = 0$. $\Leftarrow (C.3)$

(b) $W_\theta = 0$. Proof: $W_\theta > 0 \Rightarrow \lambda_2 = \xi \Rightarrow \frac{\partial \mathcal{L}}{\partial \lambda_2} = 0 \Rightarrow W_L > 0$. $\Rightarrow (a)$
We have shown above that \( P_\emptyset = 0 \), and we have now shown that both \( W_L \) and \( W_\emptyset \), and hence \( u_L \), are also zero. Therefore we have shown that every optimal contract satisfying \( W_\emptyset - W_L \geq u_L + P_\emptyset \) involves \( u_L = P_\emptyset = W_L = W_\emptyset = 0 \) (and of course \( W_H = P_H \) to deter collusion off the equilibrium path).

Appendix D. Optimal contracts in \( CA^{(i)} \)

Consider \( P \)’s problem of selecting the best contract from among those in \( CA^{(i)} \):

\[
\max_{e,u,P,W} \pi = \xi (\theta_L + e_L - e_L^2/2 - (1 - \gamma(1 - v))(u_L + W_L) - \gamma(1 - v)(W_\emptyset - P_\emptyset)) \\
+ (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H) \quad \text{s.t.} \\
\begin{align*}
\text{(IR)} & \quad \nu(u_L - s^*) + (1 - v)\big((1 - \gamma)(u_L - s^*) - \gamma P_\emptyset\big) \geq 0 \\
\text{(IC)} & \quad u_H \geq (e_L\Delta\theta - \Delta\theta^2/2 + (1 - v)\big((1 - \gamma)(u_L - s^*) - \gamma P_\emptyset\big) - \nu P_H \\
\text{(CA^{(i)})} & \quad W_\emptyset \geq W_L \quad \text{and} \quad P_\emptyset \leq P^*; \quad P_H \leq P^*; \quad s^* = (1 - \alpha)(u_L + P_\emptyset) + \alpha(W_\emptyset - W_L).
\end{align*}
\]

The Lagrangian for \( P \)’s problem is the following:

\[
\mathcal{L} = \xi \big(\theta_L + e_L - e_L^2/2 - (1 - \gamma(1 - v))(u_L + W_L) - \gamma(1 - v)(W_\emptyset - P_\emptyset)\big) \\
+ (1 - \xi)(\theta_H + e_H - e_H^2/2 - u_H) \\
+ \lambda_1 \big[[1 - \gamma(1 - v)]\big[\alpha(u_L - W_\emptyset + W_L) - (1 - \alpha)P_\emptyset\big] - \gamma(1 - v)P_\emptyset\big] \\
+ \lambda_2 \big\{u_H - (e_L\Delta\theta - \Delta\theta^2/2) \big[1 - (1 - \gamma)[\alpha(u_L - W_\emptyset + W_L) - (1 - \alpha)P_\emptyset] - \gamma P_\emptyset\big] + \nu P_H\big\} \\
+ \lambda_3 \{W_\emptyset - W_L\} + \lambda_4 \{P^* - P_\emptyset\} + \lambda_5 \{P^* - P_H\}
\]

The Kuhn-Tucker conditions for maximization include the following:

\[
\begin{align*}
\text{(D.1)} & \quad \frac{\partial \mathcal{L}}{\partial u_L} = -(\xi - \lambda_1 \alpha)(1 - \gamma(1 - v)) - \lambda_2 \alpha(1 - v)(1 - \gamma) \leq 0; \quad u_L \frac{\partial \mathcal{L}}{\partial u_L} = 0 \\
\text{(D.2)} & \quad \frac{\partial \mathcal{L}}{\partial W_\emptyset} = -\xi \gamma(1 - v) - \alpha(\lambda_1 - \lambda_2)(1 - \gamma(1 - v)) - \lambda_2 \alpha \nu + \lambda_3 \leq 0; \\
\end{align*}
\]

\[
W_\emptyset \frac{\partial \mathcal{L}}{\partial W_\emptyset} = 0
\]
(D.3) \frac{\partial \mathcal{L}}{\partial W_L} = -\xi (1 - \gamma (1 - \nu)) + \alpha (\lambda_1 - \lambda_2) (1 - \gamma (1 - \nu)) + \lambda_2 \alpha \nu - \lambda_3 \leq 0; \\
W_L \frac{\partial \mathcal{L}}{\partial W_L} = 0 \\
(D.4) \frac{\partial \mathcal{L}}{\partial \lambda_3} = W_\emptyset - W_L \geq 0; \quad \lambda_3 \frac{\partial \mathcal{L}}{\partial \lambda_3} = 0

Proof of Lemma 3:

(a) $W_L = 0$. Proof: $W_L > 0 \Rightarrow \lambda_3 = -\xi (1 - \gamma (1 - \nu)) + \alpha (\lambda_1 - \lambda_2) (1 - \gamma (1 - \nu)) + \lambda_2 \alpha \nu \Rightarrow \frac{\partial \mathcal{L}}{\partial W_\emptyset} = -\xi < 0 \Rightarrow W_\emptyset = 0$. \(\Leftrightarrow\) (D.4)

(b) $W_\emptyset = 0$. Proof: $W_\emptyset > 0 \Rightarrow \lambda_3 = \xi \gamma (1 - \nu) + \lambda_1 \alpha (1 - \gamma (1 - \nu)) - \lambda_2 \alpha (1 - \nu) (1 - \gamma)$. But if $u_L > 0$ then, using (D.1) to get $\alpha (\lambda_1 - \lambda_2) (1 - \gamma (1 - \nu)) + \lambda_2 \alpha \nu = \xi$, we have $\lambda_3 = \xi + \xi \gamma (1 - \nu) > 0$, $\Rightarrow \frac{\partial \mathcal{L}}{\partial \lambda_3} = 0, \Rightarrow W_L > 0$. \(\Leftrightarrow\) (a) Finally, if $u_L = 0$ then $W_\emptyset > 0$ results in framing, which was ruled out by Proposition 1.

Appendix E. Optimal contracts with corruption

Consider $P$’s problem of selecting the best contract from among those in $CP \cup CA$:

$$
\max_{e, u, \bar{P}, \bar{W}} \pi = \xi \{\theta_L + e_L - e_L^2 / 2 - [1 - \gamma (1 - \nu)](u_L + W_L) - \gamma (1 - \nu)(W_\emptyset - P_\emptyset)\} \\
+ (1 - \xi)(\theta_H + e_H - e_H^2 / 2 - u_H) \quad s.t.
$$

(IR) $u_L \{\nu - \tau \nu (1 - \alpha) + \alpha (1 - \nu) (1 - \gamma)\}$

$$
- P_\emptyset \{1 - \nu + \tau \nu (1 - \alpha) - \alpha (1 - \nu) (1 - \gamma)\} \\
+ (W_L - W_\emptyset) \{\alpha \nu \tau + \alpha (1 - \nu) (1 - \gamma)\} \geq 0;
$$

(IC) $u_H \geq (e_L \Delta \theta - \Delta \theta^2 / 2) + u_L \alpha (1 - \nu) (1 - \gamma) - P_\emptyset (1 - \nu)[1 - \alpha (1 - \gamma)] - \nu P_H + \\
(W_L - W_\emptyset)[\alpha (1 - \nu) (1 - \gamma)]$;

$$
W_L \geq W_\emptyset; \quad P_\emptyset \leq P^*; \quad P_H \leq P^*.$$

The Lagrangian for $P$’s problem is as follows:

$$
\mathcal{L} = \xi \bigg\{\theta_L + e_L - e_L^2 / 2 - (1 - \gamma (1 - \nu))(u_L + W_L) - \gamma (1 - \nu)(W_\emptyset - P_\emptyset)\bigg\} \\
+ (1 - \xi)(\theta_H + e_H - e_H^2 / 2 - u_H)
$$
\[ +\lambda_1\{u_L\{v - \tau v(1 - \alpha) + \alpha(1 - v)(1 - \gamma)\} \]
\[ - P_\emptyset\{1 - v + \tau v(1 - \alpha) - \alpha(1 - v)(1 - \gamma)\} \]
\[ + (W_L - W_\emptyset)\{\alpha v + \alpha(1 - v)(1 - \gamma)\}\]
\[ + \lambda_2\{u_H - (e_L\Delta \theta - \Delta \theta^2/2) - u_L\alpha(1 - v)(1 - \gamma) + P_\emptyset(1 - v)[1 - \alpha(1 - \gamma)] + vP_H \]
\[ - (W_L - W_\emptyset)\alpha(1 - v)(1 - \gamma)\}\]
\[ + \lambda_3\{W_L - W_\emptyset\} + \lambda_4\{P^* - P_\emptyset\} + \lambda_5\{P^* - P_H\}\]

The Kuhn-Tucker conditions for maximization are the following (plus the constraints and their complementary slackness conditions):

\[ \text{(E.1)} \quad \frac{\partial L}{\partial e_L} = \xi(1 - e_L) - \lambda_2\Delta \theta \leq 0; \quad e_L \frac{\partial L}{\partial e_L} = 0 \]

\[ \text{(E.2)} \quad \frac{\partial L}{\partial e_H} = (1 - \xi)(1 - e_H) \leq 0; \quad e_H \frac{\partial L}{\partial e_H} = 0 \]

\[ \text{(E.3)} \quad \frac{\partial L}{\partial u_L} = -\xi(1 - \gamma(1 - v)) + \lambda_1 v(1 - \tau(1 - \alpha)) + \lambda_1 - \lambda_2\alpha(1 - v)(1 - \gamma) \leq 0; \quad u_L \frac{\partial L}{\partial u_L} = 0 \]

\[ \text{(E.4)} \quad \frac{\partial L}{\partial u_H} = -\xi(1 - \gamma(1 - v)) + \lambda_1 \alpha(1 - v)(1 - \gamma) - \lambda_1 v(1 - \alpha) - 4 \leq 0; \quad P_\emptyset \frac{\partial L}{\partial P_\emptyset} = 0 \]

\[ \text{(E.5)} \quad \frac{\partial L}{\partial P_\emptyset} = \xi \gamma(1 - v) - (\lambda_1 - \lambda_2)(1 - v)(1 - \alpha(1 - \gamma)) - \lambda_1 \tau v(1 - \alpha) - 4 \leq 0; \quad P_\emptyset \frac{\partial L}{\partial P_\emptyset} = 0 \]

\[ \text{(E.6)} \quad \frac{\partial L}{\partial P_H} = \lambda_2 v - \lambda_3 \leq 0; \quad P_H \frac{\partial L}{\partial P_H} = 0 \]

\[ \text{(E.7)} \quad \frac{\partial L}{\partial W_\emptyset} = -\xi \gamma(1 - v) - (\lambda_1 - \lambda_2)[\alpha(1 - v)(1 - \gamma)] - \lambda_1 \tau v - \lambda_3 \leq 0; \quad W_\emptyset \frac{\partial L}{\partial W_\emptyset} = 0 \]

\[ \text{(E.8)} \quad \frac{\partial L}{\partial W_L} = -\xi(1 - \gamma(1 - v)) + (\lambda_1 - \lambda_2)[\alpha(1 - v)(1 - \gamma)] + \lambda_1 \tau v + \lambda_3 \leq 0; \quad W_L \frac{\partial L}{\partial W_L} = 0 \]

(a) (Lemma 4) \( W_\emptyset = 0 \). \text{Proof:} \quad W_\emptyset > 0 \Rightarrow \lambda_3 = -\xi \gamma(1 - v) - \alpha(\lambda_1 - \lambda_2)(1 - v)(1 - \gamma) - \lambda_1 \alpha \tau v \Rightarrow \frac{\partial L}{\partial W_L} = -\xi < 0 \Rightarrow W_L = 0. \quad \Rightarrow (W_L \geq W_\emptyset).
(b) (Lemma 5) (1) Suppose $\tau < 1$. Then $W_L = 0$. (2) Suppose $\tau = 1$. Then $W_L = 0$
without loss of generality. \textit{Proof:} If $P_\emptyset = 0$ then setting $W_L > 0$ is ruled out by Lemma
2. But $P_\emptyset > 0 \Rightarrow u_L > 0 \Rightarrow \lambda_1 = [\xi(1 - \gamma(1 - \nu)) + \lambda_2\alpha(1 - \nu)(1 - \gamma)]/f$, where
$f = \nu(1 - \tau + \alpha \tau) + \alpha(1 - \nu)(1 - \gamma) \in [0,1]$. Feeding this into (E.8) yields
$\frac{\partial L}{\partial W_L} = \lambda_3 - \frac{\nu(1 - \tau)}{f} [\xi(1 - \gamma(1 - \nu)) + \lambda_2\alpha(1 - \nu)(1 - \gamma)]$. But $W_L > 0 \Rightarrow \lambda_3 = 0$, and (1) is
proved. If $\tau = 1$, then $\frac{\partial L}{\partial W_L} = 0$ for all $W_L \geq 0$, and (2) is proved.

(c) $e_H = 1$. Immediate from (F.2).

(d) $e_L = 1 - \left(\frac{\lambda_2}{\xi}\right)\Delta \theta$. The statement follows from (E.1) since $\lambda_2 \leq (1 - \xi)$ (from (E.4)) and $\xi > \Delta \theta$ is assumed.

(e) The optimal contract is the first-best for $P^* > \left(1 - \frac{\Delta \theta}{2}\right)\frac{\Delta \theta}{\nu} \equiv P^{FB}$. \textit{Proof:} (IC)
is satisfied with $u_H = 0$ for $e_L = e_H = 1$ and $P_H = P^{FB}$.

(f) $\lambda_2 > 0$ for $P^* < P^{FB}$. \textit{Proof:} Suppose instead that $\lambda_2 = 0$, so that $u_H = 0$
(from (E.4)) and $e_L = 1$ (from (d)). Feeding these values into (IC), it can be seen that
$P_\emptyset > 0$ is required in order to satisfy (IC), and this implies (using (E.5)) $\lambda_4 =
\xi(1 - \nu) - \lambda_1(1 - f)$. Then (IR) requires $u_L \geq P_\emptyset(1 - f)/f > 0$, which implies
(using (E.3)) $\lambda_1 = \xi(1 - \gamma(1 - \nu))/f$. Feeding this back into (E.5), we obtain for $\lambda_4$ the
strictly negative amount $\xi(1 - \nu) - \xi(1 - \gamma(1 - \nu))(1 - f)/f$, which is a
contradiction, and the statement is proved. As a corollary, (since (E.6) requires $\lambda_5 > 0$ if $\lambda_2 > 0$), $P_H = P^*$ for $P^* < P^{FB}$.

The following three statements (g) – (i) partition the parameter space $P^* < P^{FB}$ into
two parts: $CP$ ($P_\emptyset = 0$) and $CA$ ($P_\emptyset > 0$).

(g) $P_\emptyset = 0$ if $\frac{1 - \xi}{\xi} < \frac{(1 - \alpha)(\nu \tau + (1 - \nu)(1 - \gamma))}{\nu(1 - \nu)(\alpha \gamma + (1 - \alpha)(1 - \nu))} \equiv \eta$. \textit{Proof:} $P_\emptyset > 0 \Rightarrow \frac{\partial L}{\partial P_\emptyset} + \lambda_4 =
\xi(1 - \nu) - \lambda_1(1 - f) + \lambda_2(1 - \nu)(1 - \alpha(1 - \gamma)) \geq 0$ and $u_L > 0 \Rightarrow \frac{\partial L}{\partial u_L} =
-\xi(1 - \gamma(1 - \nu)) + \lambda_1f - \lambda_2\alpha(1 - \nu)(1 - \gamma) = 0 \Rightarrow \lambda_1 =
\[\xi \left(1 - \gamma (1 - \nu)\right) + \lambda_2 \alpha (1 - \nu)(1 - \gamma)\] / f. Feeding this back into \(\frac{\partial L}{\partial P_0} + \lambda_4 \Rightarrow \lambda_2 \geq \xi \eta\). But \(\lambda_2 \leq (1 - \xi)\) (from (F.4)) and the statement is proved.

(h) \(P_0 = 0\) if \(P^* > \left(1 - \frac{\Delta \theta}{2} - \frac{\Delta \theta}{\xi} \Delta \theta\right) \left(\frac{\Delta \theta}{v}\right) \equiv P_{ZR0}^{ZR0}\). Proof: \(P_0 > 0 \Rightarrow \lambda_2 \geq \xi \eta\) (from (g)). Together, (IC) and (IR), along with the fact that \(P_H = P^*\) (from (f)), imply \(u_H = (e_L \Delta \theta - \Delta \theta^2/2) - P_0 \nu(1 - \nu) \left[1 - \alpha (1 - \gamma) - \tau(1 - \alpha)\right]/f - v P^* \geq 0 \Rightarrow (e_L \Delta \theta - \Delta \theta^2/2) \geq v P^* \Rightarrow \left(1 - \frac{\Delta \theta}{2} - \frac{\lambda_2}{\xi} \Delta \theta\right) \Delta \theta \geq v P^*\). But \(\lambda_2 \geq \xi \eta \Rightarrow \left(1 - \frac{\Delta \theta}{2} - \frac{\lambda_2}{\xi} \Delta \theta\right) < \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta\right)\) and so we have a contradiction if \(P^* > \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta\right) \left(\frac{\Delta \theta}{v}\right)\) and the statement is proved.

(i) \(P_0 > 0\) if both \(\frac{1 - \xi}{\xi} > \eta\) and \(P^* < P_{ZR0}^{ZR0}\). Proof: Suppose \(P_0 = 0\). (E.3) and (E.5) \(\Rightarrow \lambda_2 \leq \xi \eta\). If \(u_H > 0\) then \(\lambda_2 = 1 - \xi\) and we are done. If \(u_H = 0\) then (IC) requires \(v P^* \geq v P_H \geq \left(e_L - \frac{\Delta \theta}{2}\right) \Delta \theta\). But with (d) this is equivalent to \(P^* \geq \left(1 - \frac{\Delta \theta}{2} - \frac{\lambda_2}{\xi} \Delta \theta\right) \left(\frac{\Delta \theta}{v}\right) \geq \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta\right) \left(\frac{\Delta \theta}{v}\right) = P_{ZR0}^{ZR0}\) and we are done.

(j) Optimal contracts belonging to \(CA\) exist for \(\nu \in (\nu, \bar{\nu})\), where \(\nu > 0\) and \(\bar{\nu} < 1\), as long as \(\frac{1 - \xi}{\xi} > \frac{1 - \tau}{(1 - \sqrt{T})^2}\). Proof: We know that for \(P^* < P_{ZR0}^{ZR0}\), optimal contracts belong to \(CA\) whenever \(\frac{1 - \xi}{\xi} > \eta\). It is obvious that \(\frac{\partial \nu}{\partial \gamma} \leq 0\). It will therefore suffice to derive the condition for which \(\frac{1 - \xi}{\xi} > \eta\) for \(\gamma = 0\). \(\frac{1 - \xi}{\xi} > \frac{1 - \nu + \nu \tau}{\nu(1 - \nu)(1 - \tau)}\) fails obviously for \(\nu = 0\) and \(\nu = 1\). It is straightforward to show that the right hand side of the inequality is the smallest for \(\nu = \frac{1 - \sqrt{T}}{1 - \tau}\), and at this value of \(\nu\) we have \(\frac{1 - \xi}{\xi} > \frac{1 - \tau}{(1 - \sqrt{T})^2}\).

(k) The region in which optimal contracts belong to \(CP\) and are not first-best (when \(P^* < P^{FB}\) and either \(1 - \xi < \xi \eta\) or \(P^* > P_{ZR0}^{ZR0}\)) is divided into two parts: For \(P^* < \left(1 - \frac{\Delta \theta}{2} - \frac{1 - \xi}{\xi} \Delta \theta\right) \left(\frac{\Delta \theta}{v}\right) \equiv P_{RE0}^{ZR0}\), \(u_H > 0\), and for \(P^* > P_{RE0}^{ZR0}, u_H = 0\). Proof: From (E.4), \(u_H > 0 \Rightarrow \lambda_2 = (1 - \xi) \Rightarrow u_H = v P_{RE0}^{ZR0} - v P^*\) (from (IC) and (d)), and \(u_H = 0 \Rightarrow\)
\[ \nu P^* = \left(1 - \frac{\Delta \theta}{2} - \frac{\lambda_2}{\xi} \Delta \theta \right) \Delta \theta = \lambda_2 = \left(1 - \frac{\Delta \theta}{2} - \frac{\nu P^*}{\Delta \theta} \right), \] which is strictly positive for \( P^* < P^{FB} \) and strictly less than \((1 - \xi)\) for \( P^* > P^\text{ZR}_0 \).

(l) The region in which optimal contracts belong to \( C_A \) (when \( 1 - \xi > \xi \eta \) and \( P^* < P^\text{ZR}_0 \)) is divided into three parts: For \( P^* < \left(1 - \frac{\Delta \theta}{2} - \frac{1-\xi}{\xi} \Delta \theta \right) \left(\frac{f}{\nu(1-(1-\alpha))}\right) \Delta \theta =: P^\text{RE}_*, u_H > 0 \) and \( P_\emptyset = P^* \); for \( P^\text{RE}_* < P^* < \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta \right) \left(\frac{f}{\nu(1-(1-\alpha))}\right) \Delta \theta =: P^\text{ZR}_0, \) \( u_H = 0 \) and \( P_\emptyset = P^* \); and for \( P^\text{ZR}_0 < P^* < P^\text{ZR}_0 \), \( u_H = 0 \) and \( 0 < P_\emptyset < P^* \). Proof: (IR) requires \( u_L = P_\emptyset(1-f)/f \) and (IC) requires \( u_H = \left(1 - \frac{\Delta \theta}{2} - \frac{\lambda_2}{\xi} \Delta \theta \right) \Delta \theta - P_\emptyset \left(\frac{\nu(1-v)(1-\nu)(1-\alpha)}{f}\right) - \nu P^* \). From (E.3) and (E.5), \( \lambda_2 > \xi \eta \Rightarrow \lambda_4 = 0 \Rightarrow P_\emptyset = P^* \) and \( \lambda_2 = \xi \eta \Rightarrow \lambda_4 = 0 \). From (E.4), \( u_H > 0 \Rightarrow \lambda_2 = (1 - \xi) \Rightarrow P_\emptyset = P^* \Rightarrow u_H = \left(1 - \frac{\Delta \theta}{2} - \frac{1-\xi}{\xi} \Delta \theta \right) \Delta \theta - P^*/\left(\frac{f}{\nu(1-(1-\alpha))}\right), \) which is strictly positive for \( P^* < P^\text{RE}_* \), and \( u_H = 0 \Rightarrow \left(1 - \frac{\Delta \theta}{2} - \frac{\lambda_2}{\xi} \Delta \theta \right) \Delta \theta = P_\emptyset \left(\frac{\nu(1-v)(1-\nu)(1-\alpha)}{f}\right) + \nu P^* \). By inspection, \( \lambda_2 \) decreases as \( P_\emptyset \) increases, and for \( P_\emptyset \leq P^* < P^\text{ZR}_0 \), we have \( \lambda_2 > \xi \eta \Rightarrow P_\emptyset = P^* \). Finally for \( P^\text{RE}_* < P^* < P^\text{ZR}_0 \), we have \( \lambda_2 = \xi \eta \) and \( P_\emptyset = \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta \right) \Delta \theta - \nu P^*/\left(\frac{\nu(1-v)(1-\nu)(1-\alpha)}{f}\right), \) which is strictly positive for \( P^* < \left(1 - \frac{\Delta \theta}{2} - \eta \Delta \theta \right) \left(\frac{\nu}{\nu(1-v)}\right) = P^\text{ZR}_0. \)