Controlling Collusion and Extortion: The Twin Faces of Corruption

Ajit Mishra\textsuperscript{y} \hspace{1cm} Dilip Mookherjee\textsuperscript{z}

Abstract: Corruption has two faces: collusion and extortion. The former refers to under-reporting of offenses by inspectors in exchange for bribes. The latter refers to bribes extracted against the threat of over-reporting. Both distort penalties that offenders expect to pay, relative to what they are supposed to pay, and can seriously damage incentives to commit offenses. Existing theoretical literature on corruption control has recognized the tension between these two forms, while taking the structure of legally mandated fines for offenses as given. It has argued that extortion poses serious problems for high-powered incentive schemes for inspectors designed to combat problems of collusion. We argue that corruption control policies should be enlarged to include choice of fines, as these can help greatly in addressing both problems. We demonstrate a variety of contexts where problems of dilution of deterrence incentives owing to both collusion and extortion can be resolved by adjusting mandated fines, even though judicial systems may be weak and ineffective, and civil service norms may restrict use of high-powered bonus schemes for inspectors.

\textsuperscript{y}Department of Economics, University of Bath, BA2 7AY, UK. Email: a.mishra@bath.ac.uk
\textsuperscript{z}Department of Economics, Boston University, 270 Bay State Road, MA 02215. Email: dilipm@bu.edu

\textsuperscript{*}This version: 25 April 2013.
1 Introduction

The theoretical literature on law enforcement has identified extortion possibilities as a key obstacle to successful control of corruption (Banerjee 1994, Mookherjee 1997, Hindriks et al. 1999, Polinsky and Shavell 2001). Providing auditors or inspectors with high powered incentives is necessary to curb collusive under-reporting of offenses they discover. Indeed, such under-reporting can be completely eliminated by privatization of inspections, where all fines paid by offenders are paid as bonus rewards to inspectors. However, a key drawback of privatized enforcement is that they give rise to a different problem, wherein inspectors are motivated to fabricate or over-report offenses in order to extort bribes. The historical experience with tax farming in medieval times is replete with problems of extortion where tax collectors harassed and extracted bribes from citizens against the threat of over-reporting their incomes and citing them falsely for tax evasion. With a weak judicial process, appeals made by citizens against over-reporting of crimes by inspectors cannot be satisfactorily addressed, either because appellate authorities lack the capacity to discover the truth, or to impose sanctions on inspectors for over-reporting. The anticipation of extortion can motivate citizens to under-report their incomes or commit large offenses, as better behavior is unlikely to be rewarded if they will be falsely implicated by extortionary inspectors. So the welfare consequences under privatization could be just as grievous, or even worse than collusive under-reporting generated by low-powered incentive schemes for inspectors.

In this paper we address the twin problems of collusion and extortion in law enforcement. We show that it is possible to achieve second-best compliance levels, compliance levels which would be achieved in the absence of any form of corruption, by enlarging the set of policy instruments if revenue considerations are ignored.\footnote{Revenue problems can also be taken care of if there are no limited liability constraints for inspectors.} Specifically we argue that adjusting fines for tax evasion or pollution can be an effective means of dealing with both collusion and extortion. This possibility has been ignored or overlooked by existing literature. If choice of the fine structure can be coordinated with the design of incentives for inspectors, this allows two sets of policy instruments to deal with the twin problems of collusion and extortion. Specifically, under a wide variety of circumstances, appropriate choice of the fines for offenders (tax evaders or polluters) can effectively ‘solve’ both problems, even with very weak judicial appeals processes.
It is important to clarify what exactly we mean by ‘solve’, and the circumstances under which these results apply. By ‘solve’ we mean achieve consequences for utilitarian welfare that are the same as in a world where problems of collusion or extortion do not exist. In our ‘solution’ both collusive under-reporting and extortionary bribes will arise. But fines for offenses will be calibrated in such a way that the expected effective penalties for offenses incurred by firms will end up exactly the same as in a world without corruption, so that the resulting levels of offenses will be the same. In some cases, consequences for net revenues will also be the same, provided expected bribes earned by corrupt inspectors can be mopped up through fixed fees, assuming these do not violate inspectors’ wealth constraints. We abstract from considerations of uncertainty and risk aversion of either inspectors or firms, though we do respect wealth constraints for firms that limit fines that can be imposed on them. We do not claim universality of this solution, but the main implication of our analysis is that legal fines for offenses ought to be an important tool in combating corruption.

We consider a setting of pollution where firms have heterogenous preferences for polluting and can choose how much to pollute (denoted by $a$). Suppose the government wishes to implement a given pattern of pollution behaviour, where appropriate deterrence is efficiently accomplished by some schedule of penalties $f^*(a)$ borne by firms as a function of the level of reported pollution. Under-reporting by the inspector is sought to be prevented by the use of bonus schemes, where the inspector receives $rf^*(a)$ with the bonus rate $r \leq 1$. An example of such high-powered incentive scheme which eliminates under-reporting is the case of privatization, $r = 1$. Over-reporting is sought to be discouraged by an institutional setting where firms can appeal over-reporting by the inspector by incurring a cost, denoted by $L$. This cost is reimbursable in case of a successful appeal. Appeals succeed in discovering and then punishing over-reporting with some positive probability $x$. Parameter $x$ captures the quality of the judicial system.\(^2\) The game between inspectors and firms will result in inspectors extorting firms up to the point where the latter are indifferent between appealing and not. Privatization can be combined with lowering the actual pollution penalties below $f^*(a)$ in such a way that the sum of penalties and expected bribes paid by firms will exactly equal $f^*(a)$ following choice of pollution $a$. The reduction in penalties collected directly from the firms will be made up by lowering the salaries of inspectors by the same amount.

The same principle can be applied when such high-powered incentive schemes cannot be used for whatever reasons, such as civil service norms or problems of equity and morale within

\(^2\)We allow for $x$ to be as low as possible subject to the condition that inspectors would prefer not to over-report whenever they expect the firm to appeal with certainty. See the discussion regarding assumption A1 in Section 2.2.
bureaucracies. In such contexts both collusion and extortion can potentially arise. Collusion is associated with under-reporting by inspectors, while extortion is associated with the opposite phenomenon of over-reporting. Both cannot of course happen at the same time. Collusion will typically arise on the equilibrium path owing to low-powered incentives for inspectors, while extortion will not. Nevertheless extortion threats increase the bargaining power of inspectors, which distort ex ante pollution incentives of firms. We show that under similar assumptions regarding the appeals process as in the case of high-powered incentives, both problems of collusion and extortion problem can be resolved through appropriate adjustments of pollution fines. Extortion raises the level of effective penalties incurred by firms. However, due to the presence of wealth constraint limiting fines to lie below certain wealth level, not all pollution levels face exactly same degree of extortion threat. This leads to lowering of marginal penalties at high levels of pollution, thereby encouraging firms with high preferences for pollution to increase their levels of pollution. This can be addressed by lowering mandated fines suitably, so that the effective penalties end up just as they would be in a world without extortion. Conversely, collusion lowers both levels and marginal rates of penalties effectively borne by firms. Here also, presence of limited liability or wealth constraints lead to firms facing different expected costs from collusion. This can be ‘cured’ by raising marginal fine rates suitably and adjusting their level. Desired patterns of marginal deterrence can thus be achieved without violating limits on liabilities of firms.

This approach can be extended in suitable ways even when appeal costs are not constant but differ with the extent of over-reporting, or when firms have heterogenous and privately known appeal costs. However the ‘solution’ to the extortion problem that we are exploring in this paper may entail awarding subsidies for low levels of pollution, as a way of inducing firms to avoid large levels of pollution. When there are limits on liability of inspectors, extortion problems cannot be resolved costlessly by the method of lowering pollution fines, since the loss of government revenues cannot be recovered by lowering inspector salaries. Desired patterns of marginal deterrence can be achieved in the presence of extortion, but at a cost in terms of lower net government revenues. Nevertheless, the essential point remains that selecting the penalties for pollution remains a valuable tool for combating problems of corruption.

Our approach is in stark contrast to the common approach underlying the existing literature on corruption and enforcement. Both Hindriks et al. (1999) and Polinsky-Shavell (2001) study optimal enforcement policies in different settings. Unlike these papers, we do not attempt

While there is a substantial literature on enforcement problems in the presence of collusion (i.e. Mookherjee and Png,1995), the literature dealing with both forms of corruption is quite small. Apart from the papers mentioned in the text, a recent paper by Dufwenberg and Spagnolo (2012) also studies both forms of corruption but in a different context. There is also a related literature on harassment, see Marjit et al. (2000).
to come up with any optimal anti-corruption policy which seeks to deter corruption. Rather we take anti-corruption policy to be given and ask whether we can implement the second-best compliance levels by appropriate calibration of the mandated fines and by suitably adjusting fixed salaries of the inspectors. Our paper is also related to the recent agency literature on contracting with corrupt supervisors. Following Tirole (1986), the literature focuses on the problem of collusion and assumes away extortion. This is primarily because agent related information is assumed to be hard, something which cannot be manipulated. Baliga (1999) and Faure-Grimaud et al. (2003) extend the collusion model to allow for extortion in a procurement setting with heterogenous agents. But collusion prevention does not lead to extortion because in their model settings the agent always the option of leaving the contract after information exchanges have taken place but before any production is undertaken.

More recently, Khalil et al. (2010) examines the trade-off between collusion and extortion in a moral-hazard context. They consider a model where the hard-or-soft nature of information depends on the context: information is hard when the inspector is trying to extort but it is soft when both collude. In their model too, preventing collusion leads to extortion. But, while outcomes with collusion can be optimal in some cases, an outcome with extortion is never optimal. It is always dominated by another outcome where extortion is eliminated. Our result on the other hand suggests that extortion does not matter in the sense that we can implement the corruption-free outcome even with extortion. In our model, information is always soft but extortion is limited by the existence of an appeals system. While the court system plays an important role, the key to our result is the ability of the regulator to choose the mandated fine over a wider domain and extract extortion rents from the inspector. We discuss these issues later in the paper.

We set out the basic enforcement model in section 2 by outlining the second-best outcome. This refers to the compliance levels by different types of firms sought to be achieved in the absence of corruption. We introduce the institutional structure and the anti-corruption policy environment aimed at deterring collusion and extortion. We show that when extortion is the only threat, the second-best outcome can be achieved where the firms pay exactly the same amount of fines (expected) and there is no revenue implications for the regulator. In section 3, we consider the case where both collusion and extortion are possible threats. Section 4 considers various extensions and discussions of our main result. We conclude in section 5 with a few brief remarks.

\[\text{\textsuperscript{4}}\text{De Chiara and Livio (2012) consider a model where information is 'purely soft' but there is no outside agency (courts) trying to limit extortion. However, their information structure is quite different, the principal can condition transfers on outcomes.}\]
2 The Model

Our basic model follows Mookherjee and Png (1994) which focused on problems of securing marginal deterrence for pollution, where assets of firms limit penalties that can be imposed for pollution. That paper abstracted from problems of corruption. Call this the second best problem. We extend this to incorporate the possibility of both types of corruption: collusion and extortion. Our main focus is on showing circumstances where choosing pollution fines will fully resolve both problems in the sense that the second-best outcome can be implemented in the presence of collusion and extortion possibilities.

2.1 The Second-Best Problem: Benchmark with No Corruption

Firms choose pollution levels denoted by $a$ where $a \geq 0$. Firms derive heterogeneous benefits from pollution depending on their types, represented by $t \in [0, T]$. Firm types are unobserved by inspectors, and are distributed according positive and continuous density $z(t)$. The benefit to type $t$ from choosing pollution level $a$ is $tB(a)$. The social cost associated with pollution level $a$ is given by $h(a)$. Both benefits and costs are differentiable and strictly increasing, $B'(a) > 0, h'(a) > 0$. In addition, benefits are bounded above, $\lim_{a \to \infty} B(a) = \overline{B}$ and $\overline{B}$ is finite. Firms are inspected with probability $\mu$ and a firm detected with pollution level $a$ pays penalty $f(a)$.

The probability $\mu$ of inspection is determined by the scale of the resources devoted by the government to enforcement. To a first approximation it is proportional to the ratio of inspectors to firms. Inspectors incur time costs but no unobservable effort costs in inspecting, as in Laffont-Tirole (1993), so as to abstract from the moral hazard considerations emphasized by Mookherjee and Png (1995). Let $c$ denote the marginal social cost of raising the monitoring probability $\mu$ and we assume this is exogenously given (determined by the costs of employing inspectors and providing them necessary resources for inspection).

Also let $\lambda$ denote the marginal social value of government revenues. This is essentially the marginal deadweight loss of tax revenues, and we shall take it to be a parameter. Net government revenues will equal the difference between fines collected and costs associated with inspections.

As in Mookherjee-Png (1994), the two major restrictions on enforcement here are that (i) fines that can be imposed are limited by the offender’s wealth $W$, which is taken as given, and (ii) the inspection probability $\mu$ cannot be varied with the level of the offense. Hence marginal deterrence requires graduating fines to the severity of the offense. (ii) requires $f(a)$
to be an increasing function, while (i) requires it to be bounded above by $W$. We add to this model a concern for government revenues, which rules out the possibility of providing marginal deterrence costlessly by lowering the fine function sufficiently, without running into the limited liability constraint (i).

With honest inspectors, the true pollution will be reported by the inspector, and a type $t$ firm will choose $a(t)$ to maximize $tB(a) - \mu f(a)$.

The **second-best problem** is to select $\mu$, $f(a)$ and $a(t)$ to maximize utilitarian welfare

$$SW = \int_0^T \{tB(a(t) - h(a(t))\} z(t)dt - \lambda[c\mu - \int_0^T f(a(t))z(t)dt]$$

subject to the constraints

$$0 \leq \mu \leq 1$$

$$f(a) \leq W,$$

and the incentive constraint requiring firms to respond optimally by selecting offenses according to $a(t)$ which maximizes

$$tB(a) - \mu f(a)$$

Let $a^*(t)$ denote the second-best action schedule, and $f^*(a)$ the corresponding second-best fine function. Mookherjee-Png (1994) provides a detailed characterization of the second-best; here we note the following features.

(i) Every implementable action schedule $a(t)$ is non-decreasing in $t$, which follows from the incentive constraint (4).

(ii) The corresponding fine function $f(a)$ (which implements a non-decreasing $a(t)$) is non-decreasing in $a$. If it were decreasing over some range, firms would be incentivized to select actions at the upper end-point of this range rather than intermediate levels. The same outcomes would result by ‘ironing’ the fine function to make it flat over the entire range.

(iii) The second-best policy may involve an enforcement threshold level $a_0 \geq 0$ such that actions below this level are associated with rewards, i.e., negative fines. This will be the case if the limited liability constraint limits fines that can be imposed on large offenses, so deterrence of such offenses can be achieved only by graduating penalties sufficiently for lower levels of offenses. If the weight on government revenues is small enough relative to social harms from large offenses, and $W$ is small enough, it will make sense to deter large offenses by rewarding small offenses.

---

5Mookherjee and Png (1994) instead imposed a nonnegativity constraint on the fine function.
2.2 Collusion and Extortion

Now we allow for the possibility of corruption among inspectors. Inspectors may be inclined to under-report offenses they discover, motivated by bribes paid by firms to induce them to do so. This is the problem of *collusion* or *under-reporting*. Deterring under-reporting will require providing the inspector with carrots (in the form of bonuses for fines collected) and sticks (penalties imposed if such under-reporting is subsequently discovered). Inspectors can (or credibly threaten to) over-report offenses motivated by higher bonuses or bribes paid by firms to prevent them from doing so. This is the problem of *extortion* or *over-reporting*. We discuss how this restricts the enforcement policy.

Inspectors compensation consists of two parts, a fixed wage $w$ and bonuses or penalties associated with variable measures of inspector performance. Bonuses take the form of rewarding the inspector at a fixed rate $r$ for every dollar of fines collected, $r \leq 1$. Penalties can also be imposed following discovery of corruption. This is explained further below. The inspector has an exogenously given outside option payoff if he were to not work for the government.

The sequence of events is as follows. The regulator announces the enforcement policy $\mu, f$ in addition to inspector compensation policy. Firms choose pollution levels $a(t)$. Each firm is inspected with probability $\mu$. The true pollution level is detected costlessly by the inspector. However, the inspector is corruptible and its reported action $a'$ can be different from the true action. We describe the interaction between the inspector and the firm later. Let $e(a \mid f)$ be the effective payment/penalty associated with action $a$; given regulator’s stipulated penalty $f(a)$. Note that with a non-corrupt inspector $e(a \mid f) = f(a)$. With corrupt inspectors, firm’s maximization problem is slightly different now. Type $t$ firm will choose $a_t$ which is the solution to

$$\max_a \mu tB(a) - \mu e(a)$$

Comparing this with Eqn(4), it is clear that second-best actions $a_t^*$ can be implemented by any fine function $f(a)$ if $\partial e(a \mid f)/\partial a = \partial f^*(a)/\partial a, \forall a$. Compared to the second-best, the *enforcement problem* now involves the modified incentive compatibility constraint for the firm (5), and the constraint that ensures that inspectors are willing to work in the government.

The Game Form: The interaction between the inspector (I) and the firm (F) is captured by a game form with two components. This is quite similar to game forms used in Hindriks et al. (1999) and Polinsky and Shavell (2001).

The first stage of the game is the cooperative phase where the inspector and the firm
bargain over the report \( a' \) and payment \( b \) from the firm to the inspector. We assume that both the inspector and the firm are risk neutral and bribe determination follows the symmetric Nash bargaining solution. If they agree, inspector reports \( a' \) and receives a bribe \( b \) from the firm. If they fail to agree, both \( I \) and \( F \) choose their strategies noncooperatively with their disagreement payoffs being determined by the following non-cooperative game \( N \).

In the non-cooperative game, first \( I \) reports \( a'' \geq a' \). Then \( F \) decides whether to accept it or appeal. If accepted, the game ends and \( F \) pays \( f(a'') \) as penalty and \( I \) receives \( rf(a'') \). For the firm there is a fixed cost associated with filing an appeal. This cost denoted by \( L \) is constant and same for everyone, irrespective of the type and action of the firm.\(^7\) The inspector knows the exact value of \( L \). If the firm appeals the true action is discovered with probability \( x > 0 \), the legal cost \( L \) is reimbursed, as are the excess fines paid, so the firm ends up paying the correct fine \( f(a) \). The inspector on the other hand, refunds the excess rewards and over that pays a penalty \( m(a, a'') \); where \( m(a, a'') = 0 \) if \( a = a'' \), and \( m(a, a'') > 0 \) if \( a < a'' \). With probability \( (1-x) \), the true action is not discovered and \( F \) loses the appeal, so \( F \) pays \( f(a') \) and \( I \) receives \( rf(a') \).\(^8\)

The final stage of the game is a possible media leak which forces the government to take punitive action against exposed collusion. Since \( a' \) is the reported action following agreement in the first stage, in the absence of a media leak, the inspector receives a bonus of \( rf(a') \). Let \( q \) be the exogenous probability of the media leak. Following the leak, the firm pays \( k(a, a') \geq f(a) \), i.e., an amount that exceeds the fine that it would have paid in the absence of under-reporting. The inspector receives commission \( rf(a) \) whenever \( a = a' \) and pays a fine \( j(a, a') \) otherwise.\(^9\)

We shall assume that \( k(a, a') = f(a), j(a, a') = 0 \) if \( a = a' \), and \( k(a, a') > f(a), j(a, a') > 0 \) for all \( a > a' \). We shall assume that the bonus rates and the penalties for collusion are exogenously given.

**Penalties and Limited Liability:** While inspectors are not subject to any limited liability constraint, we nevertheless assume that inspector’s fines cannot exceed some arbitrarily given limit \( J \), \( j(a, a') \leq J \). This rules out the possibility of controlling collusion via arbitrarily large penalties following exposure in a media leak. Similarly we assume that the penalty following

\(^6\)Throughout we take \( a', a'' \) to be reports while \( a \) is the actual pollution level. \( a' \) refers to the report \( f \) both agree, where as \( a'' \) refers to the report following disagreement.

\(^7\)We relax this assumption later and discuss various cases where \( L \) is a variable cost. Additionally, we can also allow for disagreement games where \( I \) asks for a bribe by proposing a bribe and an action. An earlier version considered this but overall qualitative results are not affected.

\(^8\)Note that the appeal system is therefore assumed to be effective in the following sense. Upon discovery of over-reporting, it is is able to recover the reward and reinstate the true penalties. Hence the inspector will never pay the firm to make an over-report and collect a high reward, which the firm can avoid by appealing. With an ineffective legal system, over-reporting can be a form of collusion too.

\(^9\)We also assume that the entire bribe amount can be recovered from the inspector, following a leak.
A successful appeal is non-negligible so that while over-reporting I does not wish to induce an appeal. The following assumption ensures this fact. Consider the disagreement game where I reports $a^{\parallel}$. Following an appeal, I gets $(1 - x)rf(a^{\parallel}) + x(-m(a,a^{\parallel}))$. For a given court system and its effectiveness $x$, the inspector will not induce an appeal if following condition is always satisfied.\footnote{Suppose we set $m(a,a^{\parallel}) = m$ for any $a^{\parallel} > a$. Then a sufficient condition is $(1 - x)rW < xm$.}

\[(1 - x)rf(a^{\parallel}) + x(-m(a,a^{\parallel})) < rf(a), \forall \, a, a^{\parallel} > a. \quad (A1)\]

As mentioned earlier, we specialize to the case where collusion penalties following a media leak $k(a,a')$, and $j(a,a')$ depend linearly on the extent of under-report.

\[k(a,a') = k[f(a) - f(a')] \quad \text{and} \quad j(a,a') = j[f(a) - f(a')], \quad \forall \, a' < a, k, j > 1 \quad (A2)\]

Firm penalties are also subject to the following two limited liability constraints. These constraints play an important role in shaping the effective penalties for certain firms. Consider the noncooperative extortion game described earlier. A firm having chosen $a$ is reported by the inspector as $a^{\parallel}(a)$. Following an appeal, in the unsuccessful state, F pays $f(a^{\parallel})$ towards fine and $L$ towards court costs, the total being bounded above by $W$.

\[f(a^{\parallel}(a)) + L \leq W, \forall a \leq a_T \quad (LL1)\]

This implies that the extent to which different firms can get extorted could be different with firms with higher $a$ being extorted to a lesser extent. From (A1) it is clear that I chooses $a^{\parallel}$ so as not to induce an appeal. F does not appeal if the expected payment from an appeal is bigger:

\[xf(a) + (1 - x) \min\{f(a^{\parallel}) + L, W\} > f(a^{\parallel}) \quad (6)\]

The inspector will over-report the firm to fullest extent possible; that is the firm will be indifferent between appealing and not appealing.\footnote{Condition (6) is equivalent to $f(a^{\parallel}) < \min\{f(a) + \frac{1-x}{x}L, xf(a) + (1-x)W\}$.} Define $a_1$ by the property that

\[f(a_1) = W - \frac{L}{x} \quad (D1)\]
provided this is non-negative. If \( W < \frac{L}{x} \), \( a_1 \) is not defined. Hence,

\[
f(a'/) = f(a) + \frac{1-x}{x} L, \quad \forall a \leq a_1
\]

\[
= xf(a) + (1-x)W, \quad \text{otherwise.}
\]

If \( W < \frac{L}{x} \), the latter case applies for all \( a \). The range \( a > a_1 \) is always non-empty, which becomes the set of all \( a \)'s in the case where \( W < \frac{L}{x} \). The second limited liability condition limits the collusion penalties for the firm following a media leak. Consider a firm with \( a \) being reported as \( a'/ < a \) following collusive agreement with the inspector.

\[
k[f(a) - f(a')] \leq W \quad \text{(LL2)}
\]

When this condition is binding, firms with larger under-reporting face smaller marginal penalties for collusion.\(^{12}\) Similar to the previous case, define \( a_2 \) such that

\[
k[f(a_2) - f(0)] = W - f(0). \quad \text{(D2)}
\]

All firms with \( a \geq a_2 \), if under-reported as \( a = 0 \), will face the same penalty following a media leak.

**Payoffs and Solution:** Let \( d_I, d_F \) denote the disagreement payoffs to the inspector and the firm, \( \pi_I, \pi_F \) denote the corresponding payoffs from agreement. Using (A1) & (LL1) it is possible to solve for \( d_I \) and \( d_F \). In the absence of any appeal, firm pays \( f(a'/) \) and the inspector receives \( rf(a'/) \). From the view point of the inspector and the firm, over-reporting clearly involves a deadweight loss for any \( r < 1 \). So over-reporting will not occur because the joint payoff from reporting the true action \( a \) will always exceed the joint payoff from disagreement for \( r \leq 1 \). Hence, there will always be an agreement between \( F \) and \( I \).\(^{13}\) However, agreement does not necessarily mean under-reporting. Given our assumption that media leak is possible with reports \( a'/ < a \), we can have two cases. If under-reporting is not profitable because of the media leak and subsequent penalties, agreed report cannot be an under-report, \( a'/ \geq a \). The joint payoffs from agreement will be given by \( \pi_F + \pi_I = (r - 1) f(a'/) \). Since \( r \leq 1 \), this is maximized at \( a'/ = a \). If under-reporting is profitable, we have \( a'/ < a \) and the linearity of penalty functions (A2) imply that under-reporting is maximal, \( a'/ = 0 \). The following Lemma confirms this.

\(^{12}\)Both these conditions imply that firms with high \( a \) are likely to benefit more. Hindriks at al. (1999) note this 'regressive' feature in the context of tax enforcement.

\(^{13}\)This rules out "framing" in our model. Framing refers to the case where there is actual over-reporting. See Polinsky and Shavell (2000).
Lemma 1 Consider the bribe game described above. Suppose assumptions A1-A2 hold. Inspector’s reported action $a'$ will be given by the following

$$a' = 0 \text{ if } 1 - r > q(k + j)$$

$$= a \text{ otherwise.}$$

The detailed arguments are provided in Appendix A. Here we simply note that the condition determining the profitability of under-reporting is quite intuitive. It is a reflection of the standard carrot and stick policy. The condition can be easily derived when both the limited liability constraints are non-binding. In such a case the sum of disagreement payoffs equals $W - (1 - r)f(a'/i)$. With collusive under-reporting, $a' < a$, the sum of agreement payoffs will be $W - (1 - r)f(a') - q(k + j)[f(a) - f(a')]$. Hence, using (7), surplus from an agreement is

$$[(1 - r) - q(k + j)][f(a) - f(a')] + (1 - r)\frac{1 - x}{x}L. \quad (8)$$

If $1 - r > q(k + j)$, given that $f(a)$ is increasing, this surplus is maximized at $a' = 0$. On the other hand, for $1 - r \leq q(k + j)$, collusive under-reporting is not profitable and the surplus from an agreement is maximized at $a' = a$. In what follows we shall be discussing these two cases.

2.3 Extortion Only: $1 - r \leq q(k + j)$

Since under-reporting is subject to external monitoring such as media leaks or other forms of audits, collusive under-reporting will not be profitable for higher values of $r, q, k$, and $j$. Whenever $1 - r \leq q(k + j)$, only extortion can take place. Note that this case includes privatized enforcement, where $r = 1$. It is well-known from Laffont-Tirole (1993) and others that if no lower bound is imposed on the fixed payment $w$ to inspectors, the second-best can be achieved, in the absence of extortion, by a high-powered incentive scheme involving $r = 1$ which removes all incentives for inspectors to under-report, and then selecting $w$ low enough so as to meet the inspector’s participation constraint with equality. Rents earned by inspectors are thereby taxed away up-front, if necessary with fixed payments $w$ that are negative, representing bids posted by inspectors for the right to collect.\footnote{If there was a positive lower bound to $w$, such high-powered incentives would represent high revenue losses for the government, as fines paid by the firms will be used to pay bonuses.} We argue that second-best action schedule can be implemented even when extortion possibilities are present.

As shown above, over-reporting will not actually occur, as firms will be willing to pay
bribes to induce the inspector to not over-report. These extortion-induced bribes can distort
incentives of firms to commit offenses in the first place. To see how the enforcement problem
is affected, consider the second best action schedule and fine $f^*(a)$. We examine whether $f^*$
can implement the same action schedule in the presence of extortion.

Recall that in this case $a' = a$. It is clear that effective penalties will go up only to the
extent extortion affects the disagreement payoffs. If effective penalties go up uniformly by the
same amount so that \textit{marginal pollution costs are unchanged} for all actions we can implement
$a^*(t)$ with same fine function $f^*$, even with extortion. From (LL2) it is clear that firms with
$a \leq a_1$ face same extent of over-reporting threat, but it is different for firms with $a > a_1$. Hence
extortion bribes and effective penalties are likely to be different. In such a case, for small
values of $t$ and corresponding low levels of $a_t \leq a_1$, incentives will be unaffected but beyond
$a_1$, marginal incentives will fall and firms will choose higher pollution levels. The following
proposition confirms this.

**Proposition 1** Suppose appeals processes are weak enough in the sense that $\frac{L}{x} > W - f^*(a_T)$,
second-best outcome cannot be implemented by second-best fine function $f^*$ even when under-
reporting is deterred completely, $1 - r \leq q(k + j)$.

\textbf{Proof.} Using Nash Bargaining it is easy to verify that bribe $b = \frac{1 + r}{2} \frac{1 - x}{x} L$, if $a \leq a_1$ and
$b = \frac{1 + r}{2} (1 - x) [W - f(a)]$ otherwise. The details are provided in Appendix B. The expected
pollution costs anticipated by the firm (sum of fines and bribes) associated with pollution $a$ is
now

\[
e(a | f) = f(a) + \beta \frac{L}{x}, \text{ for } a \leq a_1, \quad \beta = \frac{(1 + r)(1 - x)}{2}, \text{ for } a > a_1
\]  

(9)

Pollution costs are raised uniformly by $\frac{\partial L}{x}$, until the wealth constraint is hit. Since $\frac{L}{x} > W - f^*(a_T)$ implies that $a_1 < a_T$, where $f^*(a_1) = W - \frac{L}{x}$, fines for all offenses below $a_1$ will rise
uniformly by a constant amount. \textit{Marginal pollution costs are unchanged until $a_1$, but fall to
$(1 - \beta)f^{'\prime}(a)$ thereafter, thus destroying marginal deterrence for high levels of pollution. The
second-best action schedule $a_t^*$ can then no longer be implemented. ■}

This result underlies the argument of Banerjee (1994), Mookherjee (1997) and Hindriks et
al. (1999) that extortion poses a drawback to the use of high-powered incentive schemes for
inspectors, a particular manifestation of multi-task problems associated with use of incentive
schemes highlighted by Holmstrom and Milgrom (1991). They argue accordingly that low-powered incentives ought to be used to address the problem of extortion. But this will in turn create problems of collusive under-reporting. Collusion therefore has to be tolerated in order to avoid excessive extortion. We argue in this paper that this trade-off need not be so hopeless. With two problems that need to be controlled, additional instruments need to be utilized. One such instrument is the fine function for pollution.

The problem with extortion is that it raises the expected penalties associated with pollution. Hence one way to address this problem is to lower the mandated fine function in such a way that the resulting schedule of expected penalties generate same level incentives as in the second-best situation. In the situation depicted above, the government can choose a fine function \( f(a) \) so that \( e(a \mid f) = f^*(a) \), for every \( a \leq a_T \). Above \( a_T \) the fine function \( f \) will be constructed so that it is strictly increasing and tending to \( W \) as \( a \to \infty \).

\[ {\text{Figure 1}} \]

**Proposition 2** Suppose \( 1 - r \leq q(k + j) \). There exists a fine function \( f \) which together with high-powered incentives enables implementation of the second-best outcome.

**Proof.** From the bribe amounts that the firm has to pay to avoid being over-reported (see Appendix B), it is clear that second-best deterrence is ensured if

\[
f(a) + \beta \frac{L}{x} = f^*(a), \quad \beta = \frac{(1 + r)(1 - x)}{2}
\]

for all \( a \leq a_1 \), and

\[
(1 - \beta) f(a) + \beta W = f^*(a)
\]

for all \( a > a_1 \).

This can be achieved by the following function:

\[
f(a) = f^*(a) - \beta \frac{L}{x}
\]

for \( a \leq a_1^* \) where \( a_1^* \) is defined by the condition that \( f(a_1^*) = W - \frac{L}{x} \), i.e., that \( f^*(a_1^*) = W - \frac{L}{x}[1 - \beta] \). For \( a > a_1^* \):

\[
f(a) = (1 - \beta)^{-1} f^*(a) + C
\]

\footnote{In a related context Khalil et al. (2010) argue that it might be optimal to tolerate some collusion but extortion has to be prevented. We discuss this later.}
where $C$ is chosen to ensure that $(1 - \beta)^{-1} f^*(a_1^*) + C = W - \frac{L}{x}$, i.e., so

$$C = W - \frac{L}{x} - (1 - \beta)^{-1} f^*(a_1^*)$$

and over the range $a > a_1^*$ we have

$$f(a) = [1 - \beta]^{-1}[f^*(a) - f^*(a_1^*)] + W - \frac{L}{x}.$$ 

It can be checked that $f(a) \leq W, \forall a$, so this is a feasible fine function. Moreover $e(a | f) = f^*(a)$, for all $a$. ■

The lowering of the fines paid by firms results in a loss of government revenues by $[f^*(a) - f(a)]$ per firm. Note that reward payments also fall, hence net revenue fall by $(1 - r)[f^*(a) - f(a)]$. However, the inspector receives extortion bribes to the tune of $[f^*(a) - f(a)]$ with a net gain of $(1 - r)[f^*(a) - f(a)]$. Hence salaries can be reduced by the same amount without violating the inspector’s participation constraint. When collusion is sought to be resolved via high-powered incentives $r \geq 1 - q(k + j)$, government revenues are unaffected by addressing extortion through lowering of fines in the way described above. Extortion does not disappear — indeed it is rampant — but it has no consequences for deterrence of pollution. Neither does it have consequences for rents illegally appropriated by inspectors, as these are extracted up-front in the form of lower salaries. Note that second-best deterrence can be implemented by other fine functions as well, but these are going to have different revenue implications. For example, a fine function $g(a) = f^*(a) - \frac{L}{x} - \epsilon, \epsilon > 0$, can also implement $a^*$ by ensuring the same degree of marginal deterrence but effective fines will not equal $f^*$. This will have revenue implications for the government even in the absence of any limited liability constraints for the inspector.

**Corollary 2** In the case of high-powered incentives, $r \geq 1 - q(k + j)$, the second-best deterrence level can be implemented with no revenue implications provided wealth constraints for inspectors do not bind, and appeal costs $L$ are constant for all firms and known to all.

**3 Collusion & Extortion:** $1 - r > q(k + j)$

Now consider the case where collusive under-reporting arises because of low-powered inspector incentives and weak oversight by the media or other watchdog agencies. There are several reasons, such as civil service norms and equity concerns, which makes high-powered incentive schemes difficult to employ. Hence we are more likely to see the coexistence of collusion and
extortion. Collusion takes place in equilibrium, while over-reporting will not but the threat of over-reporting will affect bribes, allowing inspectors to extract more compared to a situation where over-reporting is not possible.

As shown earlier in Lemma 1, if $1 - r > q(k + j)$, under-reporting will take place and it will be maximal, $a' = 0$. Even though all firms will be reported as $a' = 0$, bribes paid by the firms will be different depending the levels of pollution $a$. Payoffs from agreement as well as disagreement would reflect the two types of limited liability constraints discussed earlier. In one case, extortion is limited because total payments cannot exceed $W$. In the other case, the ability to impose fine following collusive under-reporting and its subsequent discovery is limited because $k.[f(a) - f(0)]$ can not exceed $W - f(0)$. It is clear that bribes will be affected by whether these constraints are binding or not. We shall consider the case where $a_2 < a_1 \leq a_T$, the other case with $a_1 < a_2 \leq a_T$ can be analyzed in a similar manner.

Using symmetric Nash bargaining solution, we can work out the bribe $b$ and the effective penalty function $e$ for any given fine function $f$. To see how collusion and extortion determine the effective penalty, consider first $a \leq a_2$. The details of the bribe determination for this as well as other case are in Appendix C. The effective fine for action $a$, given mandated fine function $f$, will be given by

$$e(a \mid f) = f(0) + \frac{1}{2}(1 + r + qk + qj)f(a) - \frac{1}{2}(1 + r + qk + qj)f(0) + \frac{1 - x}{2x}L(1 + r) \quad (10)$$

Since gain from collusion is positive, $1 - r > q(k + j)$, it can be seen that $1 + r + q(k + j) < 2$. This implies that collusion results in a dilution of marginal penalties, $\partial e(a \mid f)/\partial a < \partial f(a)/\partial a$. Dilution can be both in levels and the rate at which penalties increase with the level of pollution. In the absence of extortion (say $x = 1$ or $L = 0$), it can be shown that $e(a \mid f) < f(a)$. Hence firms pay less both in absolute as well as marginal terms for each level of pollution. However, since extortion possibilities raise the bribe paid by the firm, overall effective penalty can be higher.

Irrespective of whether actual penalties go up or not, what matters for firm pollution incentives is the marginal penalty associated with an increase in the pollution level. If the second-best fine function $f^*$ is mandated, collusion implies marginal expected penalties are less than what they were in the second-best context. Consequently pollution levels will rise.

To remedy this dilution, marginal fine rates need to be adjusted upwards for every level up to pollution levels of $a_T$ so that same level of deterrence can be achieved.16 This adjustment

---

16Basu et al. (1992) use a similar argument in a single-action and hierarchical setting, to ensure deterrence.
can be achieved by increasing the slope of second-best fine function. But this will affect the limited liability constraints and we need to ensure that these constraints are respected. For the mandated \( f^* \), let \( f(a) \) denote the scaled-up fine function. We need to ensure that effective marginal penalties remain unchanged, \( \partial e(a \mid f)/\partial a = \partial f^*(a)/\partial a \) for all \( a \leq a_T \). Second, we need to ensure that the new function also satisfies \( f(a) \leq W \). We achieve both these by adjusting downwards the intercept of the new fine function. The following proposition shows that such a fine function exists and it can implement the second-best schedule \( a^* \).

**Proposition 3** Let \( f^* \) implement the second-best schedule \( a^* \) in the absence of corruption. Let \( q(k + j) < 1 - r \) so collusion cannot be prevented, and extortion can also happen, with all firms having the same appeals cost \( L > 0 \). There exists a fine function \( f(a) \) such that the second-best action schedule can be implemented using this fine function, in the presence of collusion and extortion.

**Proof.** Define

\[
\alpha = \frac{2}{(1 + r + qj + qk)}, \quad \alpha' = \frac{2}{(1 + r + qj)}, \quad \alpha'' = \frac{2}{(1 + r)x + qj} \quad 1 < \alpha, \alpha', \alpha'' < 2
\]

Since \( qk > 0 \) and \( x < 1 \), it can be seen that \( \alpha < \alpha' < \alpha'' \). Denoting \( f(0) = f_0 \), define \( a_1, a_2 \) for the new function in a similar manner, \( f(a_1) = W - \frac{L}{x} \), and \( k.f(a_2) - k.f_0 = W - f_0 \). We have implicitly assumed maximal under-reporting for the new fine function. Given (A1-A2) and \( 1 - r < q(k + j) \), it can be shown that this will always be true. Consider the fine function \( f(a) \):

\[
f(a) = \alpha f^*(a) + W(1 - \alpha'') + (\alpha' - \alpha)f^*(a_2) + (\alpha'' - \alpha')f^*(a_1), \forall a \leq a_2
\]

\[
= \alpha' f^*(a) + W(1 - \alpha'') + (\alpha'' - \alpha')f^*(a_1), \forall a_2 < a \leq a_1
\]

\[
= \alpha'' f^*(a) + W(1 - \alpha''), \forall a > a_1
\]

By construction, this fine function satisfies the wealth constraint \( f(a) \leq W \). Note that \( f^*(a) \leq W \) implies \( f(a) = \alpha'' f^*(a) + W(1 - \alpha'') = W - \alpha''(W - f^*(a)) \leq W \). As shown in Appendix C, the effective fine corresponding to \( f \) will be given by

\[
e(a \mid f) = f_0 + \frac{1}{\alpha}f(a) - \frac{1}{\alpha}f_0 + \frac{\beta L}{x}, \forall a \leq a_2
\]

\[
= \frac{1}{\alpha'}f(a) + \frac{1}{q}[W - f_0] - \frac{1}{\alpha'}f_0 + \frac{\beta L}{x}, \forall a_2 < a \leq a_1
\]

\[
= \frac{1}{\alpha''}f(a) + (\beta + \frac{q}{2})W - \frac{1}{2}[(1 + r) - qj]f_0, \forall a > a_1
\]
It is clear that marginal expected (effective) penalties resulting from \(e(a \mid f)\) will be the same as in the second-best, at every action level \(\partial e(a \mid f)/\partial a = \partial f^*(a)/\partial a\). Given the absence of wealth constraints, this will implement the second-best actions \(a^*\).

We continue to assume absence of wealth constraints for inspectors, so financial costs incurred by the government in lowering pollution fines can be recouped from inspectors up-front in the form of lower salaries. In this situation, collusion does occur and inspectors earn bribes. Inspector fixed salaries can be adjusted to ensure that rents to the inspectors are minimized. However, unlike the case in section 2, we are unable to claim that this policy will not have any revenue implications. Of course financial cost considerations represent one rationale for unwillingness of governments to select high-powered incentive schemes for inspectors or lower fines for firms. Our purpose in this section is to establish the result that *the second-best can still be implemented with judicious choice of fines, which overcome both collusion and extortion problems, under the conditions described in this and previous sections.* We obtain these results for the case where collusion penalties following exposure after a media leak are proportional to the fines underpaid (and hence linear in the bribes). Extension to more general set of collusion penalties will have to be considered in future work.

**4 Extensions**

In what follows we consider various weakenings of the pristine conditions assumed earlier. In particular, we assumed that legals costs \((L)\) are constant and known to the inspector as well as the planner. We relax this assumption and consider various cases where legals costs are variable and possibly, private information of the firm. To focus on this issue, we ignore the limited liability constraints (LL1 & LL2) and consider the first case with extortion only (Section 2.3).

**4.1 Variable Legal Costs**

Above we assumed that the legal cost \(L\) is same irrespective of what \(a\) and \(a^{il}\) are. This may be unrealistic in some situations. Appeal costs might depend on the level of \(a\) and the extent of extortion. Consider the case where appeal costs \(L\) depend on the extent of excess fine imposed on the firm \(f(a^{il}) - f(a) = S\). Let \(L = L(f(a^{il}) - f(a))\). The firm will appeal iff

\[
\frac{1-x}{x} L(f(a^{il}) - f(a)) < f(a^{il}) - f(a)
\]

(12)
A priori, it is not clear whether the legal cost of appeal should be increasing or decreasing in the extra fines resulting from extortion. In either case, it is reasonable to assume that \( L(0) > 0 \). Additionally, suppose \( \frac{1-x}{x} L'/ \leq 1 \). There is a unique fixed point \( S^* \) of the function \( f \). Then the firm with true pollution \( a \) will appeal if and only if \( f(a^*) - f(a) > S^* \).

Consequently the inspector will be able to over-report the offense up to the point where \( f(a^*) = f(a) + S^* \) without inciting the firm to appeal. Hence the extortion bribe will be equal to \( L(S^*) \), irrespective of the level of pollution. This reduces to the case of a constant appeal cost, with \( L = L(S^*) \). Hence the analysis in Sections 2 & 3 continues to apply in this case.\(^{17}\)

4.2 Some Honest Inspectors

Similar issues arise when a fraction of inspectors are honest and do not extort. This does not distort patterns of marginal deterrence either, under the other conditions assumed so far. With penalties chosen according to \( f(a) \) which lower the second-best penalties uniformly by the equilibrium extortion bribe \( b \) charged by corrupt inspectors, which is a constant and independent of \( a \), second-best marginal deterrence is again ensured for all firms, irrespective of what kind of inspector they are assigned to. Firms with honest inspectors will pay \( g(a) = f(a) - b \) and those with corrupt inspectors will pay \( f(a) \). So those with honest inspector will pay less but by a constant amount \( b \).

The problem this might give rise to, is with regard to the determination of inspector salaries since the government will not know which inspectors are honest. These are familiar issues from the work of Besley and McLaren (1993). If it offers a salary which makes the corrupt inspectors indifferent between working or not for the government, the honest inspectors will not work for the government. But this has no welfare consequences. In other words, the government may as well pay such a low salary that induces only the corrupt inspectors to work for the government, and still achieve the second-best.\(^ {18}\)

4.3 Privately Known Legal Costs

Firms may vary with regard to the legal costs they will incur to file an appeal, owing to underlying differences in legal expertise and connections. In the bargaining over extortionary

\(^{17}\)Matters are more complicated if legal costs depend on the actual pollution levels themselves, rather than just the size of the fines involved. Suppose \( L \) depends on \( a^*/a \), the extent to which pollution was over-reported. Now firm will appeal if \( \frac{1-x}{x} L(a^*/a) < f(a^*) - f(a) \). In this case, extortion bribe will typically depend on \( a \). Whether second-best actions can be implemented requires further investigations.

\(^{18}\)An ancillary problem arises if inspectors vary in their level of honestness, a problem stressed by Besley and McLaren (1993): auctioning the right to inspect would result in appointment of corrupt inspectors as the more corrupt would want to bid more.
bribes, this will hamper the ability of inspectors to extract bribes. Suppose each firm faces a fixed legal cost \( L \) which can be different from the cost of other firms. Suppose it is distributed according some distribution function \( \theta(L) \). Assume that \( \theta \) is independent of \( t \), i.e., legal costs are independent of preferences for pollution. The firm knows the realization of \( L \), but the inspector does not.\(^{19}\) Given this informational asymmetry, we modify the game form used earlier to determine the interaction between the firm and the inspector (Section 2.2).

The inspector will now be limited in its ability to discriminate between firms of varying legal costs. Upon discovering a firm which has polluted \( a \), it can threaten to over-report the pollution to \( f(a^\prime) = f(a) + b \) and not do so if the firm pays a bribe \( b \). Whereupon the firm will appeal if \( L \) is smaller than \( b \) (with probability \( \theta(b) \)) and not otherwise. In the former case the appeals process will impose a fine \( m(f(a^\prime) - f(a)) = m(b) \) on the inspector with probability \( x \), which the inspector will trade off against the benefit of the bribe if either the firm does not appeal, or the appeal does not succeed. The key point to note that the optimal bribe \( b^* \) for the inspector which will depend on the distribution \( \theta \), and the effectiveness and sanctions of the appeals process (\( x \) and \( m(.) \)), but will be independent of the actual pollution level \( a \).

Therefore the second-best actions can still be implemented as above, lowering the fine function up to \( a_T \) uniformly by the equilibrium extortion bribe \( b^* \). Those that do not appeal will face an expected penalty equal to \( g(a) + b^* = f(a) \), those that do will incur an expected penalty of

\[
e(a \mid f) = (1 - x)[f(a) + b^* + L] + xf(a) = f(a) + (1 - x)(b^* + L).
\]

Hence the pattern of marginal deterrence is unaffected. Given the assumed absence of wealth effects, the second-best actions \( a_t \) will continue to be chosen.

Of course in this situation some deadweight losses may arise owing to extortion — associated with appeals that happen in equilibrium. So it is possible that the level of welfare is lower as a result, depending on the deadweight resource costs of appeals that are filed. But the primary impacts of extortion on levels of pollution or expected penalties paid by firms can be avoided. Of course, this depends partly on the assumption that legal costs and preferences for pollution are independently distributed, which ensures that the equilibrium bribe is independent of the pollution level. However, it is not clear whether we have any compelling reasons to believe that legal costs and pollution preferences should be correlated. This issue needs further exploration.

\(^{19}\)We assume that the government and the inspector have the same information about the legal costs incurred by any given firm. Clearly if this were not true in the cases considered above where all firms have the same legal cost \( L \), which is known by the inspector but not the government, the latter would not know exactly by how much to lower the pollution fines to deal with the extortion problem.
4.4 Collusion-Extortion: trade-off?

As mentioned earlier, the literature has focussed on the trade-off between collusion and extortion. Efforts to prevent one invariably lead to the other form of corruption. In some cases it is possible to prevent both collusion and extortion but the resulting outcome can be highly inefficient (i.e. Hindriks et al. 1999, Khalil et al. 2010). Hence one is tempted to ask: if it is not possible or desirable to eliminate both and if one has to live with one form of corruption, which one is least harmful? Khalil et al. (2010) posed this question and points out that it is better to tolerate collusion than extortion.\(^\text{20}\)

Although we do not explicitly study optimal enforcement policy and our model is quite different, it is possible derive implications on this issue.

Consider the privatized enforcement case, i.e. \(r = 1\), which is a special case of extortion-only or collusion-free case studied in section 2.3. We have shown that the presence of extortion has no effect on welfare because the second-best compliance levels can be achieved with no revenue loss for the government. In this sense extortion does not matter. Now consider a collusion-only or extortion-free outcome. In our model context, it is not possible to guarantee extortion-free outcome because the inspector can still over-report provided the report is not going to be challenged by the firm. We can amend our model slightly and assume that over-reports are also subject to media leaks with small but positive probability. We assume this probability to be very small so that rest of the analysis is unaffected. Let \(r = 0\), we have an extortion-free outcome. This will be a special case of the case with collusion and extortion studied in section 3. With some further simplifications, \(j = 0\), we can solve for the effective penalty associated with \(a\) for the second-best fine \(f^*\).

\[
e(a \mid f^*) = \frac{(1 + qk)}{2} f^*(a), \forall a \leq a_2
= \frac{f^*(a)}{2} + \frac{qW}{2}, a > a_2
\]

Since \(1 + qk < 1\), this clearly amounts to a dilution. Like before, we can implement the second-best action schedule \(a^*\), but this will result in revenue loss to the government. Suppose we follow the construction used in proposition 1 and try to find a fine function \(f\) such that \(e(a \mid f) = f^*\) for all \(a\). Such a function exists, but it will violate \(f \leq W\).\(^\text{21}\) However, it is possible to implement \(a^*\) by restoring the marginal incentives. This seems to suggest that collusion matters, where

\(^{20}\)Polinsky and Shavell (2000) arrive at seemingly opposite conclusion, in a different setting, where it is not optimal to deter extortion when framing cannot be prevented.

\(^{21}\)Such a function is given by \(f(a) = \frac{2}{1 + qk} f^*(a), \forall a \leq a_2\) and \(f(a) = 2 f^*(a) + \frac{W}{k} - 2 f^*(a_2)\), \(\forall a > a_2\). Note that \(k f(a_2) = W\).
as extortion did not matter. The intuition for this contrasting implications lies in the different ways collusion and extortion distort effective penalties. Extortion raises effective penalty and we need to lower it, hence there is no problem of violating the firm’s wealth constraint. Lower fines leads to correspondingly lower revenues but the government is able to claw it back from the inspector. Collusion lowers effective penalty, hence we need to raise the original penalty but this cannot be achieved as fines are bounded above. As a result, effective penalties will fall short of the second-best levels.

5 Concluding Comments

All of the preceding arguments have utilized the lack of any wealth constraint of inspectors. Clearly this is unrealistic. The key argument is that inspectors may earn bribes owing to collusion or extortion, but judicious choice of the fine function imposed on firms ensures that bribes depend on pollution in a way that firms end up with the right marginal disincentives for pollution. To avoid losses of net revenues for the government, controlling extortion requires the fixed salaries of inspectors be adjusted downward by the expected bribes, divided by the probability of an effective audit. These salaries could well end up being negative: inspectors will have to bid for the right to inspect. There may well be limits to how much inspectors can bid, owing to wealth constraints. Alternatively, there may norms or regulations concerning minimum salaries. In that case controlling extortion by lowering pollution fines will result in revenue losses for the government. The same reason may restrict the use of high-powered incentives for collectors.

If these wealth constraints are binding, the government will no longer be able to implement second-best actions without incurring revenue losses. The preceding arguments show that it can implement second-best actions, at a financial cost. This revenue loss has to be weighed off against the welfare costs of lowering pollution incentives. We suspect this is the key problem underlying problems of corruption, which deserves more attention in future research.

The other assumption which played a role in the analysis is risk-neutrality of firms, which enabled us to abstract from wealth effects (apart from limited liability constraints on choice of fines). In situations where this does not hold, firm’s choice of pollution will depend on absolute levels of fines for different pollution levels. Consequently, the circumstances where we can implement the second-best pollution levels would be limited.

References


**Appendix**

A. Proof of Lemma:

(i) Consider the non-cooperative game N first. Firm will never appeal if \(a'' \leq a\). Since \(r \geq 0\), we can consider only \(a'' > a\). From (A1) it is clear that I chooses \(a''\) so as not to induce an appeal. F does not appeal if the expected payment from an appeal is bigger:

\[
x f(a) + (1 - x) \min \{f(a'') + L, W\} > f(a'')
\]

or equivalently if

\[
f(a'') < \min \{f(a) + \frac{1 - x}{x} L, xf(a) + (1 - x)W\}
\]

It can be seen that I chooses \(a''\) such that

\[
a'' = f^{-1}(\min \{f(a) + L(1 - x)/x, xf(a) + (1 - x)W\}).
\]

Using this, the disagreement payoffs are given by

\[
d_I = rf(a'') \text{ and } d_F = W - f(a''),
\]

(ii) Consider the first stage. Following agreement, suppose \(a' \geq a\). Then payoffs to both F and I from agreement \((b, a')\) will be \(\pi_F = W - f(a') - b\) and \(\pi_I = r f(a') + b\). Joint payoffs from agreement exceed the sum of disagreement payoffs iff

\[
W + (r - 1)f(a') \geq W + (r - 1)f(a'')
\]
This inequality always holds for any \( r \leq 1 \) and \( a' < a'' \). Surplus from agreement is given by

\[
(\pi_I + \pi_F) - (d_I + d_F) = (r - 1)[f(a') - f(a'')]
\]

Since \( a' \geq a \), the above expression is maximized at \( a' = a \).

(iii) Now consider \( a' < a \). Since this case involves possible leak with probability \( q \) and subsequent fines for bribery, limited liability considerations come in to play. Using (A2), payoffs to \( F \) and \( I \) will be given by

\[
\pi_I = rf(a') + b - q\{j[f(a) - f(a') + b]\}
\]

\[
\pi_F = W - f(a') - b - q\min\{k[f(a) - f(a')], W - f(a')\}
\]

The joint payoff from agreement is simply \( W - (1 - r)f(a') - q\{\min\{k[f(a) - f(a')], W - f(a')\} + j[f(a) - f(a')]\}. To see when the joint payoff from agreement is maximized, consider first the case where limited liability constraint is not binding. Clearly, maximizing \( \pi_I + \pi_F \) is equivalent to minimizing \( \{(1 - r) - q(k + j)\}f(a') \) for a suitable choice of \( a' \). If \( 1 - r > q(k + j) \), the joint payoff is maximized at \( a' = 0 \).

The optimality of maximal under-reporting holds when the limited liability condition is binding. Consider an action \( a \) and a report \( \tilde{a} \) such that the limited liability constraint is binding. Since \( k > 0 \), the constraint is likely to bind for bigger under-reports rather than for smaller under-reports. Clearly, if the optimal under report is \( a' \leq \tilde{a} \), then \( a' = 0 \). If under reporting is profitable with a higher total penalty \( (k + j) \), it will of course be profitable with \((1 + j)\).

When \( 1 - r \geq q(k + j) \), under-reporting is not optimal and joint payoff is increasing in \( a' \).

But for \( a' \geq a \), joint payoff is maximized at \( a' = a \).

B. Bargaining over Extortionary Bribes: \( 1 - r \leq q(k + j) \)

As discussed in the text earlier, if \( r < 1 \) over-reporting by \( I \) generates less in extra rewards for \( I \) than the extra fines paid by \( F \), which is inefficient for the coalition. They can avoid this deadweight loss by \( I \) ceasing to over-report, in exchange for a bribe \( b \) paid by \( F \).

If \( a \leq a_1 \), Nash bargaining equates the gains of each party: \( b - r\frac{L(1-x)}{x} = \frac{L(1-x)}{x} - b \), yielding \( b^* = \frac{(1+r)L(1-x)}{2x} \) and an expected penalty to \( F \) of

\[
e(a|f) = f(a) + \frac{(1 + r) L(1 - x)}{2x}. \tag{17}
\]

If \( a > a_1 \), the bribe solves \( rf(a) - r\{xf(a) + (1 - x)W\} + b = xf(a) + (1 - x)W - f(a) - b \),
yielding \( b^* = \frac{1}{2}(1 - x)(1 + r)[W - f(a)] \) and an expected penalty

\[
e(a|f) = [1 - \frac{1}{2}(1 - x)(1 + r)]f(a) + \frac{1}{2}(1 - x)(1 + r)W
\]  

(18)

C. Bargaining over Collusive Bribes: \( 1 - r > q(k + j) \)

We can rewrite the joint payoff from agreement as

\[
W - (1 - r)f_0 - q\min\{k:f(a) - f_0), W - f_0\} - qj[f(a) - f_0]
\]  

(19)

Joint payoff from disagreement will be

\[
W - (1 - r)f(a//), \text{ where } f(a//) = \min\{f(a) + \frac{1-x}{x}L, xf(a) + (1 - x)W\}.
\]

Consider the case where neither limited liability condition is not binding, \( a \leq a_2 \).

\[
\pi_F - d_F = \{W - f_0 - (1 - q)b - qk[f(a) - f_0]\} - \{W - f(a) - \frac{1-x}{x}L\} \]  

(20)

\[
= (1 - qk)[f(a) - f_0] - (1 - q)b + \frac{1-x}{x}L
\]  

(21)

\[
\pi_I - d_I = \{rf_0 + (1 - q)b - qj[f(a) - f_0]\} - r\{f(a) + \frac{1-x}{x}L\}
\]  

(22)

Bribe \( b \) is given by the solution to \((1 - qk)[f(a) - f_0] - (1 - q)b + \frac{1-x}{x}L = (1 - q)b - r\frac{1-x}{x}L - (r + qj)[f(a) - f_0];\)

\[
b = \frac{1}{2(1 - q)}\{(1 - qk + r + qj)[f(a) - f_0] + \frac{1-x}{x}L(1 + r)\}
\]  

(23)

Since expected penalty is given by \( e(a) = f_0 + (1 - q)b + qk[f(a) - f_0], \) plugging the value of \( b \) we get

\[
e(a | f) = f_0 + \frac{1}{2}(1 + r + qk + qj)f(a) - \frac{1}{2}(1 + r + qk + qj)f_0 + \frac{1-x}{2x}L(1 + r)
\]  

(24)

Likewise, we can solve for bribe \( b \) and expected penalty \( e \) when only one limited liability constraint is binding, \( a_2 < a \leq a_1 \).
\[
\pi_F - d_F = \{W - f_0 - (1 - q)b - q(W - f_0)\} - \{W - f(a) - \frac{1-x}{x}L\} \tag{25}
\]
\[
= f(a) + \frac{1-x}{x}L - f_0 - (1 - q)b - q(W - f_0) \tag{26}
\]
\[
\pi_I - d_I = (1 - q)b - r \frac{1-x}{x}L - (r + qj)[f(a) - f_0]
\]

We can solve for bribe,
\[
b = \frac{1}{2(1-q)}\{(1+r+qj)[f(a) - f_0] + \frac{1-x}{x}L(1+r)\} \tag{27}
\]
and expected penalty will be given by
\[
e(a \mid f) = f_0 + \frac{1}{2}(1 + r + qj)f(a) + \frac{1}{2}q[W - f_0] - \frac{1}{2}(1 + r + qj)f_0 + \frac{1-x}{2x}L(1+r). \tag{28}
\]

Finally, consider \( a > a_1 \). Proceeding as before, we can show that
\[
e(a \mid f) = f_0 + \frac{1}{2}[(1+r)x+qj]f(a) + \frac{1}{2}[(1+r)(1-x) - q]W - \frac{1}{2}[(1+r-q)(1+j)]f_0 + q(W - f_0). \tag{29}
\]
Figure 1