# Contents

List of Figures ........................................ iv

Introduction ........................................ 1

1 Ex-ante and Ex-Post Corruption .................. 3
   1.1 Introduction .................................... 3
   1.2 The Model ..................................... 6
      1.2.1 Ex-post corruption ...................... 8
      1.2.2 Ex-ante corruption .................... 12
      1.2.3 Ex-ante and ex-post corruption ....... 17
   1.3 Asymmetric information ...................... 19
      1.3.1 Ex-ante corruption .................... 19
      1.3.2 Pooling and separating equilibria .... 21
      1.3.3 Examples ................................ 24
   1.4 Conclusion .................................... 25

Bibliography .......................................... 27

2 Self Reporting Reduces Corruption in Law Enforcement .... 29
   2.1 Introduction .................................... 29
   2.2 The Model ..................................... 32
      2.2.1 Structure ................................ 32
      2.2.2 No Self Reporting ..................... 35
      2.2.3 Self Reporting ......................... 38
## Contents

2.2.4 Comparative Statics ................................................. 41

2.3 Discussion and Limits to the Theory .................................. 43
  2.3.1 Weakening of enforcement effort .................................... 43
  2.3.2 Adverse selection of officers ...................................... 44
  2.3.3 Failure to self report ............................................. 44
  2.3.4 Unconstrained officers ............................................ 45
  2.3.5 More complex anti-corruption policies .............................. 45

2.4 Conclusions .................................................................. 46

Bibliography .................................................................. 47

Appendix: Proof of Proposition 2.1 ........................................ 51
Appendix: Proof of Proposition 2.2 ........................................ 52

3 Conditional Delegation and Optimal Supervision ......................... 53
  3.1 Introduction ................................................................ 53
  3.2 Benchmarks .............................................................. 55
    3.2.1 Collusion-free Supervision (CF) ................................. 57
    3.2.2 Collusion-proof Supervision (CP) ............................... 59
  3.3 Conditional Supervision–CS ........................................... 60
    3.3.1 An intuitive Explanation ...................................... 60
    3.3.2 A Formal Model ................................................. 62
  3.4 Sequential Contracting–SC ............................................ 67
    3.4.1 What happens when the efficient F chooses the fast contract with probability 1? ................................. 70
    3.4.2 On the attainability of welfare-improving Sequential Contracting ............................................ 71
  3.5 Conclusions .............................................................. 71

Bibliography .................................................................. 73

Appendix: Proof of proposition 3.2 ........................................ 76

4 Collusion and Partial Delegation ........................................... 77
  4.1 Introduction .............................................................. 77
List of Figures

1-1 Decision Loci ................................................................. 22
3-1 Timing of the Game ....................................................... 57
3-2 Timing of the Game ....................................................... 63
4-1 Organizational Responses .............................................. 83
Introduction

This Ph.D. thesis is a collection of four papers, each one corresponding to a chapter. The background challenge, common to all chapters, is the design of anti-corruption policies in a principal/supervisor/agent framework, where the supervisor and the agent can collude at the expense of the principal. The first two works tackle the issue of corruption from the perspective of law enforcement. The third and fourth chapters consider the problem of collusion from a mechanism design point of view. Accordingly, the focus is shifted from incentives to organizational responses, where the choice between centralized and decentralized organization of economic activity is relevant.

More precisely, the first chapter studies the optimal compensation policy for a corruptible inspector, in charge with monitoring evasion from taxpayer. Namely, I discuss how the optimal compensation policy varies according to the timing of collusion, which is allowed to occur either before or after inspection takes place. Results show that increasing the inspector’s bonus rate is a better policy than increasing the penalty rate, when corruption occurs after inspection. The contrary is true when the collusive agreement is established before the inspection. Implications for privatization of law enforcement and economic development are also analyzed.

The second chapter analyses the impact of self reporting on law enforcement when officers are corruptible. The threat of corruption highlights two additional advantages to the use of self reporting beyond those identified elsewhere. First, by allowing individuals to self report their unlawful acts, the government is able to increase welfare by eliminating rents to its officers. Second, the introduction of self reporting further benefits those governments, which in its absence would tolerate corruption, by allowing them to fully eliminate corruption.

The third chapter considers a simple modification of Laffont and Tirole’s (1991) standard
mechanism in hierarchical centralized structures, where two agents (a firm and an agency that acts as supervisor) can collude at the expense of the principal. The firm can choose between two competing contracts: a fast contract, which is free from supervision, or a grand contract, that is subject to it. This mechanism eliminates the agency costs of supervision by providing firms with the possibility of avoiding supervision altogether in the first stage. Thus, the model yields results that are superior to the standard hierarchical model. When firms are risk averse, this mechanism also provides an insurance coverage to productive agents. As a consequence, it would be worthwhile even abstracting from collusion.

Finally, the fourth chapter studies the optimal design of an organization within a principal-supervisor-agent setting, with an agent informed on the cost of his economic activity and a supervisor better informed than the principal about agent’s type. This work shows that collusion is not harmful if the principal contracts with both parties and collusion is allowed only after the acceptance of the principal’s contract. Moreover, supervision is valuable regardless of the intensity of asymmetric information inside the coalition. These results are robust to alternative information structures and hold for a quite generic specification of agent’s type.
Chapter 1

Ex-ante and Ex-Post Corruption

1.1 Introduction

When it comes to tax collection, there is pervasive evidence of several forms of dishonesty and malpractice, not only limited to developing countries but also to much higher profile cases\(^1\). Collection of tax revenues typically implies several forms of illicit behavior: taxpayers may try to evade their legal liabilities, while tax inspectors may abuse their authority soliciting bribe or threatening the taxpayer to report a taxable income higher than the true one. The economic literature has already well documented that, whenever law enforcement is delegated, opportunities for collusion between taxpayer and inspector may arise: most of the articles, however, allow collusion to happen only after inspection has already occurred. The main object in this article is to incorporate the possibility for the collusive contract to be established both before and after the inspector has already exerted effort: in what follows, the first type of collusion is denoted as *ex-ante corruption* (or preemptive collusion), while the former as *ex-post corruption*. As long as timing of corruption is endogenously determined, the relation between compensation policies and corruption may be different with respect to the standard economic wisdom. Conceptually, the main novelty is that inspector can avoid the monitoring effort by choosing to collude ex-ante; therefore she decides whether or not to accept a preemptive bribe taking into consideration the equilibrium level of effort she is going to exert if ex-ante

\(^1\)Young et al. (2001) analyse corruption in the contest of American IRS.
collusion is not taking place. Ceteris paribus, higher levels of effort imply increased profitability of preemptive collusion, i.e. profitability of ex-ante corruption is increasing in the out-of-equilibrium optimal level of effort that the inspector would have exerted by monitoring. As a consequence of this result, any kind of compensation policy aimed to increase the inspector’s monitoring effort may also increase the likelihood of ex-ante corruption. This is of course an undesired side effect which arises directly from allowing endogenous timing of corruption. The other side of the story concerns taxpayers: evaders benefits from preemptive collusion are due to lower levels of monitoring; as a consequence evasion is generally larger in case of ex-ante collusion. But taxpayer is also harmed by ex-ante collusion since it eliminates the monitoring lottery which may be favorable to him with some probability, i.e. due to imperfect monitoring technology, inspector may not discover the concealed income even after inspection is carried out. Even if ex-ante and ex-post corruption have been already documented in detail\(^2\), only few contributions combine the two types of corruption in one model, studying their interaction. Among them Guriev’s (2003) work analyzes the interactions between ex-post, ex-ante corruption and red tape: in his paper the bureaucrat may extort bribes from the agent in exchange for reducing the amount of red tape (ex-ante corruption). Moreover, the bureaucrat may take bribes to conceal the information produced through red tape. The former kind of corruption tends to reduce red tape, while the later is increasing it. In any case, authors show that the equilibrium level of red type is above the social optimum. Bac and Bag (2000) also analyses ex-ante corruption in relation to law enforcement costs and legal presumptions.

Apart from the ones cited above, several other articles study the optimal incentive policy for a corruptible law enforcer but none of them allow for timing of corruption to be endogenously determined. Becker and Stigler (1974) focus on controlling bribery and consider paying commission to enforcers. Mookherjee and Png (1995) consider bribery and the use of both sanctions and rewards as means of control. Bowels and Garoupa (1997) discuss bribery control through sanctions. Hindricks, Keen and Muthoo (1999) analyze bribery and extortion in the contest of tax evasion, considering commissions and penalties as methods of control. Polinsky

\(^2\)For example, serbian custom officials are known to accept bribe from travelers without checking their belongings: see "Investigation: Serbia Losing Customs Corruption Battle" by John Simpson (2005). William et all in their survey of corruption in Eastern Europe discuss the problem of ex-ante as well as ex-post corruption. See also surveys by Tirole (1992) Bardhan (1997).
and Shavell (2000) also examine both bribery and extortion allowing for both penalties and rewards to be used against corruption. Acconcia, D’Amato and Martina (2003) restrict their attention on tax evasion and analyze the interaction between evasion, corruption, monitoring and incentive schemes. Hasker and Okten (2005) study the impact of intermediaries on corruption and show that traditional methods of fighting corruption, i.e. penalty and rotation, may not be appropriate when interaction between clients, public official and intermediary agents is considered. More important for the purpose of this article is Mookherjee’s (1997) contribution: he considers bribery and extortion in the context of tax evasion, allowing the possibility to use both sanctions and reward as means of control. The model presented here is the natural extension of this paper when *ex-ante* collusion is incorporated: in particular it addresses the problem of designing incentives mechanisms for public bureaucrats considering the effect of these incentive schemes on corruption, taxpayer compliance and tax revenue.

This paper also discuss the issue of privatization of law enforcement. Indeed, it has been argued that when it comes to reform bureaucracies the side effects of incentive reforms may be so costly to neutralize the positive ones\(^\text{3}\). Given these concrete difficulties in designing effective incentives mechanisms for public bureaucrats, it has been questioned if there is any scope for the existence of a Civil Service in charge with monitoring evasion from taxpayers. In other words, given the growing disillusionment with government bureaucracies, it has been questioned whether should be more reasonable to reform or dismantle these bureaucracies. In favor of this last option, Becker and Stigler (1974) advocated privatizing law enforcement to motivate the inspector. More recently Yang (2005) analyzed the effectiveness of the so called "hiring integrity" strategy: this approach encompasses hiring private firms to monitor potentially corrupt activity\(^\text{4}\). Yang found that countries implementing such inspection programs experienced large increases in import duty collections and moreover hired integrity appears to have been cost-effective. Nonetheless it remains conceptually unclear whether or not privatization is the optimal policy to adopt. More specifically, arguments against privatization of law enforcement

\(^3\)At the contrary, Mookherjee (1998) argues that governments truly committed to reform may succeed in reforming tax administration if they make use of instruments aimed at altering the institutional attributes of public bureaucracies.

\(^4\)Dozens of developing countries have adopted this strategy to fight corruption in customs services. The approach consists in hiring private firms to conduct preshipment inspection of imports. Inspection firms’ reputation plays a crucial role in guaranteeing honesty and reliability when competition among the private monitors generates incentives for integrity.
can be found in Mookherjee (1997) and Mookherjee and Png (1995) - in their setting privatization could be interpreted as a reward rate of 100%. In both articles, dismantling Civil Service need not be the optimal strategy. The most immediate draw-back is harassment of honest citizens and extortion. On the one hand, privatization gives inspector an incentive to resist the evasion of taxes; on the other hand, it also gives them an incentive to over-state taxes. As long as privatization entails an increased stake for denouncing evaders, it also enhances their ability to extort by making credible the threat of over-reporting taxpayer’s income. This paper shows that it exists another crucial aspect to be considered in adopting policy devices which entail increasing inspector reward rate. Indeed, a different and more subtle reason why privatization may not solve the problem of corruption lies in the timing of collusion itself: if the taxpayer can bribe the inspector before she actually inspects, further increases in reward rate may not affect the likelihood of corruption and, at the same time, it may produce counter effect in terms of both taxpayer compliance and tax revenue. Under certain conditions, therefore, the revenue authority may be interested in holding both "carrots" (bonuses) and "sticks" (penalties), as policy instruments: more precisely, under ex-ante corruption, increasing the bonus rate is never a better policy than increasing the penalty rate. Therefore in this case no privatization at all is warranted. Reversed results applied if ex-post corruption is considered.

Finally, when penalties are poorly designed (the extreme case being privatization of law enforcement), economic development, i.e. higher levels of income, increases the likelihood of ex-ante corruption. In this case, the use of penalties is required to prevent corruption.

Section 1.2 presents the outline of the model. Section 1.3 analyses ex-ante and ex-post collusion under symmetric information. Section 1.4 introduces asymmetric information. Section 1.5 concludes.

1.2 The Model

The model entails a risk-neutral taxpayer (whom is referred to as “he”) with true income $y \in \mathbb{R}^+$. The income is distributed with a cumulative distribution function $G(.)$, which is common knowledge. The taxpayer can conceal income by an amount $e$ by self reporting an income equal to $y - e$ and paying a tax of $t(y - e)$; $t$ is assumed to be constant. Under the
law, evasion is subject to a fine $f$ and the revenue authority employs an inspector (whom is referred to as “she”) to monitor taxpayers and enforce the regulation. To obtain evidence of evasion with probability $p \in [0, 1]$, the inspector must exert effort $E(p)$ which is unobservable by the tax administration on a routine basis, unless special audits are arranged; $E(p)$ is assumed to be strictly increasing, convex, differentiable and $E(0) = 0$. In this section, I assume that the inspector observes the real taxpayer’s income. Accordingly, inspector’s job may be viewed as consisting in finding evidence of evasion, while the amount of evasion is already common information. This setting may apply to those situations in which taxpayer cannot help signaling his own true income\(^5\). The timing of the game is as follows: before having actually exert the effort, $E(p)$, tax collector must decide whether to initiate a preemptive collusion with the taxpayer or whether to start investigating. In case tax collector decides to collude ex-ante, she will indicate the size of the bribe, $b$, that taxpayer must paid in exchange for not being monitored. The side contract is assumed to be enforceable. If taxpayer accepts the ex-ante agreement, information about the bribe and the taxpayer’s true income leaks anyway to the regulator with an exogenous probability $l$; this may happen through an internal audit or external vigilance agency. The first regulator’s policy consists of two instruments: penalty for bribe-taking $g$ and penalty for bribe-giving $c$. These components are the ‘sticks’: penalty can be imposed on the tax collector in the form of a fine that has a constant pecuniary (present value) of $c$ on the amount of underreport income\(^6\). Symmetrically, the tax payer is penalized at a fixed proportional rate, $g$.

If the taxpayer rejects the inspector’s collusive offer, she proceeds to investigate him. With some probability $p$ evidence of evasion is discovered and tax collector has discretion regarding the level of evasion, $d$, that she reports to the regulator. $d$ is further assumed not to exceed $e$, the true level of income concealed by the taxpayer\(^7\). Given that inspector reports an evasion

\(^5\)For example, living expenses and consumption behaviors may be used by the inspector to infer the real taxpayer’s income.

\(^6\)The penalty may be thought as a transfer to an undesired location, a refusal of promotion in the future, or the extreme punishment of being fired.

\(^7\)Having this assumption in place implies that extortion or overassessment are not allowed in the model: the result wouldn’t change substantially if those issues were to be considered. To see this point, suppose that a taxpayer can file an appeal when the inspector reports $d > e$. The overall cost of appealing is $A$, which includes monetary and psychic costs. The probability of having a successful appeal is $a$: in this case, the taxpayer is refunded the excess taxes and fines paid $(t + f)(d - e)$ as well as some fraction, $k$, of the costs incurred in appealing the assessment. The tax collector is made to pay back her commissions plus an additional penalty for
of $d$, the taxpayer must pay additional taxes of $td$ and an additional penalty at a constant rate $f$ on the amount of evasion disclosed by the inspector; that is, overall, taxpayer has to pay back a total of $(t + f)d$. At this stage, the inspector may ask for a bribe (ex-post corruption) for under-reporting taxpayer’s evasion. Again, with some exogenous probability $l$, the bribery is discovered; same penalties apply as for the ex-ante collusion case. Apart from penalties, which represent the first policy instrument, the second regulator’s policy corresponds to the ‘carrot’: the tax collector can retain a certain fraction, $r$, of the additional revenues generated. On top of these collection-based bonuses, the tax collector is paid a fixed salary $W$, where $U$ is her outside option. Having this schedule in place, it is possible to analyze different compensation mechanisms. One extreme case entails no incentive pay at all ($c = r = 0$), that is, the compensation mechanism reduces to a form of fixed salary. This solution may be rare: usually, corrupt and inefficient behaviors are penalized along the lines of an implicit code of conduct. When this is the case, sticks represent the incentive. The use of promotions based on performance is also quite usual: collection-based bonuses indirectly have the same effect. Finally, the other extreme case is privatization of tax collection where all revenues are retained by tax collector ($r = 1$); in this case $W$ is typically negative, representing a transfer from the tax collector to the government. This transfer could take the form of a permit for the right to collect taxes.

### 1.2.1 Ex-post corruption

To begin with, suppose that only ex-post corruption is possible. If inspection is successful, tax collector discovers evidence of taxpayer’s evasion, $e$. In this case, the expected benefit to the taxpayer from not being denounced is\(^8\)

\[(t + f)e - l(t + g)e,
\]

overassessment at a rate $x$, i.e. $(x + r)(t + f)(d - e)$. In the event of a unsuccessful appeal, the taxpayer receives no refund, while the tax collector is entitled to keep the entire commission, $r(t + f)d$. Having this schedule in place, the possibility of overassessment exercises no effect on monitoring or tax evasion incentives: indeed, extortion ends up representing a lump sum taxation. We address the reader to Mookherjee (1997) for a formal proof.

\(^8\)It is easy to notice that either the tax collector will report the entire evasion ($d = e$) or nothing ($d = 0$).
while the cost to the tax collector of not reporting the evasion equals the sum of expected penalties for bribery and forgone commissions,

\[ lce + r(t + f)e. \]

Accordingly, the condition for corruption to occur is

\[ (1 - r)(t + f) > l(c + t + g), \]

which ensures that the collective gains to tax payer and tax collector are positive.

Notice that (1.1) never holds when tax collection is privatized, i.e. \( r = 1 \), while it is more likely to be satisfied when there is no incentive pay, i.e. \( r = c = 0 \).

When this condition is satisfied, the supervisor determines the ex-post bribe by keeping in mind that the taxpayer will reject his collusive offer if the bribe that he demands is too large. Thus, the ex-post bribe \( b_p \) and the expected cost of giving a bribe \( l(t + g)e \) must be less than the fine \((t + f)e\). That is, \((t + f)e - l(t + g)e > b_p\). However, the supervisor will not offer a bribe unless it is profitable to him. Thus, \( b_p \) minus the penalty for taking a bribe \( lce \) must be greater than \( r(t + f)e \). That is, \( b_p > lce + r(t + f)e \). For simplicity the supervisor is allowed to extract all the surplus from taxpayer, making him indifferent between paying the bribe or let the evasion to be disclosed\(^9\). In this case the ex-post bribe level would be

\[ b_p = e[t + f - l(t + g)]. \]  

\(^9\)Our main results are not affected by this assumption regarding the bargain process.

**Corrupt regime**

Suppose the regime is corrupt, i.e. (1.1) holds. Tax collector’s expected payoff equals

\[ W + pe[t + f - l(t + g + c)] - E(p), \]  

and the taxpayer’s expected payoff is given by
\[ y - t(y - e) - pe(t + f). \]  

(1.4)

In the second stage of the game the two will simultaneously select their respective strategies: the tax collector will select the inspection effort \((p)\), while the taxpayer will decide the level of evasion \((e)\). From now on, \((e_p, p_p)\) denotes the equilibrium values in the ex-post corrupt regime. The unique Nash equilibrium of this game depends on the parameter values. If the amount of evasion is “interior”, equilibrium monitoring intensity and evasion solve

\[
\begin{align*}
  p_p &= \frac{t}{(t + f)}, \\
  e_p &= \frac{E'(p_p)}{(t + f) - l(c + t + g)}. 
\end{align*}
\]  

(1.5)

(1.6)

In the corner solution taxpayer discloses nothing at all \((e_p = y)\) and the equilibrium monitoring intensity \(p\) solves \(E'(p_p) = y[(t + f) - l(c + t + g)].\) For the sake of simplicity, in what follows I assume that \(p_p \geq \frac{t}{(t + f)}\) when \(e_p = y.\) This assumption doesn’t affect our main results; it simply allows us to rule out corner solutions which may complicate the exposition. Notice that when (1.5) holds, the taxpayer is actually indifferent between concealing all his income and not evading at all: therefore he randomizes accordingly. The equilibrium level of expected evasion solves (1.6)\(^{10}\). From equation (1.5) and (1.6), notice that a small increase in bonus rate, i.e. positive incentives, has no effect on the optimal level of tax evasion\(^{11}\). On the contrary, increased use of penalties, i.e. negative incentive, increases tax evasion. This suggests that the "carrot" is more effective than the "stick".

In order to analyze this policy implication properly, tax revenue and social welfare must be computed. The expected government’s net revenues equal the difference between expected tax revenues and the tax collectors’s wage

---

\(^{10}\) These mixed strategies could be easily "purified" by introducing heterogeneity among taxpayers.

\(^{11}\) This result is due to our assumption on the bargain process between inspector and taxpayer; as long as inspector detains all the bargain power, both carrot and stick hasn’t direct impact on the equilibrium bribe. In turn this implies that carrots (bonuses) and stick (penalty for bribe taking) have limited influence on the equilibrium evasion and monitoring effort.
\[ NR = t(y - e_p) - W + p_p l(c + t + g)e_p. \] (1.7)

The social welfare is calculated aggregating the (shadow) value of net government revenues and tax collector’s and taxpayer’s utility. Given these assumptions, social welfare equals

\[ SW = y + (\lambda - 1)(ty - W) - (\lambda - 1)[t - p_p l(c + t + g)] e_p - E(p_p), \] (1.8)

where \( \lambda > 1 \) denotes the shadow value of net government revenues. To maximize welfare and revenues, tax collector’s salary is set at the smallest possible value (that is, \( W = E(p_p) + U \)). Inserting this into equations (1.7) and (1.8) and substituting (1.5) and (1.6), yields reduced-form expressions for revenues and welfare:

\[ NR = t(y - e_p) - E(p_p) - U + p_p l(c + t + g)e_p, \]

\[ SW = y + (\lambda - 1)(ty - U) - (\lambda - 1)t \frac{E'(\frac{t}{(t+f)})}{(t+f)} - \lambda E\left(\frac{t}{(t+f)}\right). \]

Somehow surprisingly, local increases in incentive components have neither effect on net revenues nor on welfare. However, while an increase in the bonus rate will cause no effect on tax evasion, increases in penalties increase evasion. Therefore, if policy maker’s target is to jointly defeat corruption and evasion, increasing the bonus rate is a better policy than increasing the penalty rate for corruption. This is reasonable if additional negative side-effects of evasion were to be considered. These may include distortion in productive decisions and a general deterioration of the central government’s ability to control and regulate the economy.

The following proposition summarize the results.

**Proposition 1.1** When the regime is ex-post corrupted, small changes in the regulator’s policy have the following effects:
<table>
<thead>
<tr>
<th>Effect on</th>
<th>Monitoring (p)</th>
<th>Evasion (e)</th>
<th>Likelihood of corruption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small increase in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Penalty on inspector, c</em></td>
<td>None</td>
<td>Increase</td>
<td>Decrease</td>
</tr>
<tr>
<td><em>Reward to inspector, r</em></td>
<td>None</td>
<td>None</td>
<td>Decrease</td>
</tr>
<tr>
<td><em>Penalty on taxpayer, g</em></td>
<td>None</td>
<td>Increase</td>
<td>Decrease</td>
</tr>
</tbody>
</table>

### Clean regime

Consider now a clean regime, i.e. (1.1) does not hold. In this case, the expected payoff for inspector is $W + p[r(t + f)e] - E(p)$ and $y - t(y - e) - p[t + f]e$ for taxpayer. From now on, $(e_c, p_c)$ denotes the equilibrium values in the clean regime. If the amount of evasion is interior, the optimal level of effort by inspector, $p_c$, and the optimal level of evasion $e_c$ are given by

\[
\begin{align*}
  p_c &= \frac{t}{(t + f)}, \quad (1.9) \\
  e_c &= \frac{E(p_c)}{r(t + f)}. \quad (1.10)
\end{align*}
\]

Otherwise, if evasion is maximal ($e_c = y$) the equilibrium monitoring intensity solves $E'(p_c) = yr(t + f)$. As before, for the sake of exposition, we rule out corner solutions by assuming that $p_c \geq \frac{r}{(t+f)}$ when $e_c = y$.

#### 1.2.2 Ex-ante corruption

In the following section only ex-ante collusion is allowed\(^\text{12}\). In this case, the timing of the game can be represented as follow:

- Tax collector decides whether to initiate an ex-ante collusion with the taxpayer or whether to start inspecting. If the tax collector decides to engage a preemptive agreement, she makes a take-it-or-leave-it offer to the taxpayer for not monitoring him. This contract is assumed enforceable.

---

\(^{12}\) This is equivalent to assume that ex-post corruption is not profitable, i.e. condition (1.1) doesn’t hold.
- Taxpayer decides whether to accept ex-ante collusion or not.

- Taxpayer decides how much to evade.

- If ex-ante collusion has not taken place, the inspector selects the monitoring effort, inspection is carried out and its results are delivered to the principal; fines eventually apply and payoffs are realized.

- If the ex-ante agreement has been established, inspection doesn’t take place, while the principal receives notification from inspector, which certifies that the taxpayer has not conceal any income. Bribery is discovered with exogenous probability, fines and penalties eventually apply and payoffs are realized.

In case the preemptive agreement between inspector and taxpayer fails to be established, the game’s setting is identical to the one presented in the previous section, i.e. clean regime outputs apply. On the contrary, if the taxpayer decides to accept inspector’s offer the structure of the game is different. In order to solve the model, the concept of subgame perfect Nash is adopted. That is, first the individual’s optimal choice of evasion $e$ is computed, contingent on him accepting the preemptive agreement. Next, the necessary and sufficient conditions for an ex-ante collusion are analyzed.

In the last stage of the game, taxpayer must decide how much to evade; in the ex-ante corrupt regime the expected payoff of the tax collector will be

$$W + b_a - lce,$$

and the expected payoff of the taxpayer is given by

$$y - t(y - e) - b_a - l(t + g)e.$$

For a given level of the ex-ante bribe, which is exchanged in the first stage of the game, the equilibrium level of evasion is given by $e_a = y$, if the following condition holds:

$$\frac{\partial U}{\partial e} = t - l(t + g) \geq 0. \quad \tag{1.11}$$
Otherwise, the agent self-reports his true income and $e_a = 0$. Differently from equation (1.6), where ex-post corruption was considered, it is evident that neither a small increase in positive incentives, $r$ (a higher bonus rate) nor $c$ (a higher penalty for bribe taking) causes tax evasion to vary. On the other hand, penalty for bribe giving, $g$, has negative impact on evasion; if $g$ is raised sufficiently, the outcome may be to switch the system to a corner equilibrium at which taxpayers disclose all their income. In this sense, the level of tax evasion is decreasing in penalty for bribe giving, $g$. The equilibrium strategies in the ex-ante corrupted regime are denoted by $(e_a, p_a)$, which represent respectively the level of evasion and the effort devoted to monitoring, where $p_a = 0$.

In the first stage of the game, what determines whether there will be ex-ante corruption or not? The expected benefit to the taxpayer from not being inspected equals

$$t(e_a - e_c) - l(t + g)e_a + p_c[(t + f)e_c].$$

(1.12)

Note that taxpayer's benefit depends on the out-of-equilibrium level of monitoring effort, $p_c$, which is exerted by inspector in case preemptive collusion is not taking place. Therefore, $p_c$ represents the optimal level of monitoring that solves equation (1.9). Up to this point, the only difference between ex-ante and ex-post benefit from corruption consists in the presence of two extra terms, $p_c$ and $t(e_a - e_c)$; indeed, when the preemptive agreement is proposed to the taxpayer, he considers his probability of getting caught during the inspection, $p_c$, in case he refuses to pay the ex-ante bribe. Moreover, he keeps in mind the difference in the equilibrium level of evasion between the two regimes, the clean one, $e_c$, and the ex-ante corrupt one, $e_a$. These elements were not present in the previous section, where only ex-post corruption was allowed to be established. In that setting collusion happens only after monitor activity has already occurred and evidence of evasion has been collected.

The gain to the tax collector of colluding ex-ante equals the saved effort minus the sum of expected penalties for bribery and forgone commissions,

$$E(p_c) - lce_a - p_c[r(t + f)e_c].$$

(1.13)
Confronting (1.3) with (1.13), it is easy to notice that two new elements enter in the former expression. First, the inspector’s gain includes the effort that he avoids by not monitoring, \( E(p_c) \). Second, the forgone commissions are obtained only with a certain probability, \( p_c \), contingent on the collection of the hard evidences required to obtain legal persecution.

Again, the expected gains from corruption are captured entirely by the inspector, who holds all the bargain power, so that the bribe level would be

\[
b_a = t(e_a - e_c) - l(t + g)e_a + p_c[(t + f)e_c]. \tag{1.14}
\]

Corruption occurs only if the collective gains to the pair are positive,

\[
p_c(1 - r)(t + f)e_c + E(p_c) + t(e_a - e_c) \geq l(t + g + c)e_a. \tag{1.15}
\]

By substituting (1.9), (1.10) and \( e_a = y \) into the above expression, I obtain

\[
E(p_c) - \frac{E'(p_c)}{(t + f)}t + ty \geq l(t + g + c)y,
\]

and the solution for the equilibrium levels of \( p_c \) solve equation (1.9).

In general, the impact of increasing incentive pay on corruption is ambiguous; it is easy to notice that increase in the value of \( r \) has no effect on the likelihood of corruption. This result contrasts sharply with the one obtained in the previous section; by introducing ex-ante collusion the down-side of increasing in the level of bonus rate is uncovered. This effect was not present in case taxpayer and inspector can only collude after monitor activity has already been carried out.

On the other hand, I obtain more familiar results when considering increases in the values of \( c \) and \( l \), which make inequalities less likely to hold and in this sense reduce the likelihood of corruption.

As before, in order to further analyze this regime, tax revenue and social welfare must be computed. The expected government’s net revenues equal

\[
NR = ty - e_a - W + l(c + t + g)e_a.
\]
By substituting $e_a = y$, the expected value of the government’s net revenues is

$$NR = l(c + t + g)y - W,$$  \hspace{1cm} (1.16)

while social welfare is given by

$$SW = y + (\lambda - 1)l(c + t + g)y - (\lambda - 1)W.$$  \hspace{1cm} (1.17)

To maximize welfare and revenues, tax collector’s salary is set at the smallest possible value (that is, $W = U$). Inserting this into equations (1.16) and (1.17), yields reduced-form expressions for revenues and welfare:

$$NR = l(t + g)y - U + b_a,$$
$$SW = y + (\lambda - 1)l(c + t + g)y - (\lambda - 1)U.$$

Local increases in incentive component $c$ determines social welfare to increase. If $c$ is raised sufficiently, the outcome may be to switch the system from a corrupt regime to a clean one. Indeed, while an increase in penalties will cause the likelihood of corruption to decline, increases in the bonus rate, $r$, have no effect on the likelihood of corruption nor on welfare. Hence, increasing the bonus rate is never better policy than increasing the penalty rate for corruption. Moreover, in this model, as long as the penalties can be freely varied, the optimal compensation policy entails increasing it enough to eliminate corruption entirely. The results are encapsulated in the following proposition

**Proposition 1.2** When the regime is ex-ante corrupted and condition (1.1) doesn’t hold, small changes in the regulator’s policy have the following effects:

<table>
<thead>
<tr>
<th>Effect on</th>
<th>Monitoring( (p) )</th>
<th>Evasion( (e) )</th>
<th>Likelihood of corruption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small increase in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penalty on inspector, (c)</td>
<td>None</td>
<td>None</td>
<td>Decrease</td>
</tr>
<tr>
<td>Reward to inspector, (r)</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Penalty on taxpayer, (g)</td>
<td>None</td>
<td>Decrease</td>
<td>Decrease</td>
</tr>
</tbody>
</table>
1.2.3 Ex-ante and ex-post corruption

In this section both ex-ante and ex-post corruption are considered, studying the interaction between these two forms of collusion. This is equivalent to assuming that condition (1.1) holds, i.e. ex-post corruption is preferred to the clean regime. In the first stage of the game, agents will decide whether to collude ex-ante or not; they take into consideration that in case they fail to establish a preemptive agreement, they will collude ex-post after inspection has been carried out and contingent on hard evidences being discovered by the inspector during his monitor activities. The expected benefit to the taxpayer from not being inspected equals

$$t(e_a - e_p) - l(t + g)e_a + p_pl(t + g)e_p. \quad (1.18)$$

Taxpayer’s benefit depends on the out-of-equilibrium level of monitoring and evasion effort under the ex-post corrupt regime, i.e. $p_p$ and $e_p$. When the preemptive agreement is proposed to the taxpayer, he considers his probability of getting caught during the inspection, $p_p$, in case he refuses to pay the ex-ante bribe. A second aspect is related with the difference in the equilibrium level of evasion between the two regimes, the ex-post corrupt one, $e_p$, and the ex-ante corrupt one, $e_a$.

The gain to the tax collector of colluding ex-ante equals the saved effort minus the sum of expected penalties for bribery and forgone commissions,

$$E(p_p) - le_a + p_pl e_p. \quad (1.19)$$

Confronting (1.3) with (1.19), two new elements are introduced. First, the inspector’s gain includes the effort that he avoids by not monitoring, $E(p_p)$. Second, the ex-post penalty for bribe-taking is now avoided.

Ex-ante corruption occurs only if the collective gains to the pair are positive:

$$E(p_p) + t(e_a - e_p) + p_pl(t + g + c)e_p \geq l(t + g + c)e_a. \quad (1.20)$$

By substituting (1.5) and (1.6) into the former expression I obtain
\[ E(p_p) + t e_a \geq l (t + g + c) e_a + e_p \left( t \left( 1 - \frac{l (t + g + c)}{(t + f)} \right) \right). \]

Defining \( G_a \) as the coalition gain under the ex-ante corrupt regime,

\[ G_a = E(p_p) + t e_a - l (t + g + c) e_a - e_p \left( t \left( 1 - \frac{l (t + g + c)}{(t + f)} \right) \right). \]

Given that condition (1.1) must hold, I derived the above equation with respect to \( c \), obtaining

\[
\frac{\partial G_a}{\partial c} = \left( e_a - e_p \frac{t}{(t + f)} \right) + \frac{\partial e_p}{\partial c} \left| t \left( 1 - \frac{l (t + g + c)}{(t + f)} \right) \right| \geq 0.
\]

It follows that increases in \( c \) decrease the likelihood of ex-ante corruption. Note that \( r \) has no impact at all on the likelihood of corruption.

Hence, increasing the bonus rate is never a better policy than increasing the penalty rate for corruption. Moreover, in this model, as long as the penalties can be freely varied, the optimal compensation policy entails increasing it enough to eliminate corruption entirely\(^\text{13}\). A summary of the results is proposed in the following proposition.

**Proposition 1.3** When the regime is ex-ante corrupt and condition (1.1) holds, small changes in the regulator’s policy have the following effects:

<table>
<thead>
<tr>
<th>Effect on</th>
<th>Monitoring ((p))</th>
<th>Evasion ((e))</th>
<th>Likelihood of corruption</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Small increase in</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penalty on inspector, (c)</td>
<td>None</td>
<td>None</td>
<td>Decrease</td>
</tr>
<tr>
<td>Reward to inspector, (r)</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Penalty on taxpayer, (g)</td>
<td>None</td>
<td>Decrease</td>
<td>Decrease</td>
</tr>
</tbody>
</table>

\(^\text{13}\)Unlike this paper, the fact that some countries may prefer to let corruption widespread, has been shown in several economic articles that consider second best policy involving a certain level of corruption. For example, Acemoglu and Verdier (2000) analyze the case in which preventing all corruption is too costly and second best intervention may involve a certain fraction of bureaucrats accepting bribes. Besley and McLaren (1993) identify conditions under which it may be optimal to allow corruption among tax collectors. Rose-Ackerman (1978) and Tirole (1986) are also examples of articles that do not assume corruption to be *per se* a bad phenomenon.
1.3 Asymmetric information

1.3.1 Ex-ante corruption

In this section I propose a slight modification of the previous model: the inspector can no longer observe the real taxpayer’s income, \( y \in \mathbb{R}^+ \). Anyway, when the inspector is matched with a taxpayer, she learns that taxpayer’s income could be either "high" (\( y_h \)) with probability \( \theta \) or "low" (\( y_l \)) with probability \( (1 - \theta) \). All the other assumptions remain unchanged. Notice that the analysis of ex-post corruption is identical to the one presented in the previous section. On the contrary, ex-ante corruption is affected by the introduction of asymmetric information between the taxpayer and the inspector: this is due to the fact that she ignores which type of taxpayer she is going to be dealing with during the ex-ante agreement.

Suppose that condition (1.1) holds. In this case ex-post corruption is preferred to the clean regime. It is possible to show that it exists a separating equilibrium where the high income type accepts the ex-ante agreement, while the low income type refuses it. It is useful to check the conditions such that this separating equilibrium exists. As in the previous sections, all the bargain power is assumed to be in inspector’s hands. The taxpayer, characterized by a level of income \( y_h \), is not deviating from accepting to pay the ex-ante bribe if the following condition holds:

\[
y_h - t(y_h - e_a) - b_a - l(t + g)e_a \geq 0
\]

where the RHS (LHS) indicates the rich type utility under the ex-post corrupt regime (ex-ante corrupt regime). By substituting \( e_a = y_h \), the above expression can be rewritten in the following fashion:

\[
y_h - t(y_h - e_p) - p_pe_p [t + f - l(t + g)] - p_pe_p (t + g),
\]

Recall from equation (1.11) that the condition for evasion to occur is
\[ \frac{\partial U}{\partial e} = t - l(t + g) \geq 0. \]

Therefore it is easy to check that the equilibrium bribe is increasing in the level of income.

The other side of the story concerns the poor type; in order for the low income type not to accept the ex-ante bribe, the following condition must hold:

\[
y_t - t(y_t - e_a) - b_a - l(t + g)e_a \leq y_t - t(y_t - e_p) - p_p e_p (t + f),
\]

where the RHS (LHS) indicates the poor type utility under the ex-post corrupt regime (ex-ante corrupt regime). By substituting \( e_a = y_t \), the following expression is obtained:

\[
b_a \geq y_t [t - l(t + g)] - te_p + p_p e_p (t + f). \tag{1.22}
\]

Since \( \bar{b}_a \geq b_a \), it always exists a bribe that sustains the separating equilibrium.

If condition (1.1) doesn’t hold, the alternative to ex-ante corruption is the clean regime. In this case the rich type is not deviating from accepting to pay the ex-ante bribe if the following condition holds:

\[
\bar{b}_a \leq y_t [t - l(t + g)] - te_c + p_c e_c (t + f).
\]

In order for the low income type not to accept the ex-ante bribe the following condition must hold:

\[
b_a \geq y_t [t - l(t + g)] - te_c + p_c e_c (t + f).
\]
1.3.2 Pooling and separating equilibria

Ex-post corrupt regime

Suppose that (1.1) holds. If the inspector opts for the separating equilibrium, he asks the expensive bribe and only the rich citizen will pay it. Inspector expected payoff is given by

\[
U_{\text{separating}} = \theta \{ W + y_h \left[ t - l(c + t + g) \right] - te_p + p_p e_p \left( t + f \right) \} + (1 - \theta) \{ W + p_p e_p \left[ t + f - l(c + t + g) \right] - E(p_p) \}. \tag{1.23}
\]

First, the higher the frequency of rich type in the population, the larger the profitability of the "separating" strategy. Also, the higher the income of the rich type, the higher the profitability of the separating equilibrium.

In the pooling equilibrium, the inspector asks for the cheap bribe and both types of citizen are willing to pay it.

Inspector’s expected payoffs equals

\[
U_{\text{pooling}} = W + y_l \left[ t - l(c + t + g) \right] - te_p + p_p e_p \left( t + f \right). \tag{1.24}
\]

Given that \( \theta \) and \( y_h \) do not enter in the last equation, it is straightforward to note that the higher the frequency of rich type in the population, the larger the profitability of the "separating" strategy with respect to the pooling one. Also, the higher the income of the rich type, the higher the profitability of the separating equilibrium with respect to the pooling one.

While the low income type is indifferent between the two types of equilibria, the same doesn’t apply for the high income type. Indeed, his utility is always higher under the pooling equilibrium, where he is made to pay a lower bribe with respect to the separating equilibrium. As a consequence the pooling equilibrium would be "regressive" with respect to the separating one.

Having this schedule in place, the two equilibria are compared. Note that the inspector will select between a separating or pooling equilibrium by comparing the two levels of welfare (1.23) and (1.24). It is evident that the choice depends critically on the level of the parameters in
the model, particularly on $\theta$, $y_h$ and $y_l$. In order to compare the two equilibria, consider two decision loci, representing values of $\theta$ and $(y_h - y_l)$ for which the inspector would be indifferent between any of the two equilibria. This can be done by comparing (1.23) and (1.24): the separating equilibrium and the pooling one yield the same level of inspector’s utility if and only if

$$G = (\theta y_h - y_l) [t - l(c + t + g)] + (1 - \theta) [-pe_pl(c + t + g) - E(p_p) + te_p] = 0. \quad (1.25)$$

Therefore,

$$\theta = \frac{te_p - y_l [t - l(c + t + g)] - pe_pl(c + t + g) - E(p_p)}{te_p - y_h [t - l(c + t + g)] - pe_pl(c + t + g) - E(p_p)}.$$

The curve in Figure 1-1 represents the values of $\theta$ and $y_h$ for which the inspector would be indifferent between any of the two equilibria. This curve divides the space in two areas: the area on the right indicates the values of $\theta$ and $y_h$ such that the inspector prefers the separating equilibrium to the pooling one. The contrary applies for the area on the left.

![Figure 1-1: Decision Loci](image)

When the pooling equilibrium is selected, condition (1.20) changes accordingly,
\[ E(p_p) + t(y_l - e_p) + p_p l (t + g + c) e_p \geq l (t + g + c) y_l. \] (1.26)

In case the separating equilibrium is chosen, condition (1.20) is replaced by

\[ E(p_p) + t(y_h - e_p) + p_p l (t + g + c) e_p \geq l (t + g + c) y_h. \] (1.27)

**Clean regime**

If condition (1.1) doesn’t hold, inspector expected payoffs under the separating equilibrium is given by

\[
U_{\text{separating}} = \theta \{ W + y_h [t - l (c + t + g)] - t e_c + p_c e_c (t + f) \} + \\
+ (1 - \theta) \{ W + p_c [r (t + f) e_c] - E(p_p) \}.
\]

In the pooling equilibrium inspector’s expected payoffs is given by

\[
U_{\text{pooling}} = W + y_l [t - l (c + t + g)] - t e_c + p_c e_c (t + f).
\]

The separating equilibrium and the pooling one yield the same level of inspector’s utility if and only if

\[
\theta = \frac{t e_c - y_l [t - l (c + t + g)] - p_c (t + f) e_c - E(p) + p_c [r (t + f) e_c]}{t e_c - y_h [t - l (c + t + g)] - p_c (t + f) e_c - E(p) + p_c [r (t + f) e_c]}.
\]

Same discussion applies here as the one presented in the previous paragraph where the ex-post corrupt regime was considered.

When the pooling equilibrium is selected, condition (1.15) changes in the following fashion:

\[
E(p_c) - \frac{E'(p_c)}{t + f} t + ty_l \geq l (t + g + c) y_l. \] (1.28)

In case the separating equilibrium is chosen, condition (1.15) is given by
\[ E(p_c) - \frac{E'(p_c)}{(t+f)} t + ty_h \geq l(t + g + c) y_h. \] (1.29)

### 1.3.3 Examples

It may be useful at this stage of the exposition to examine some examples, in order to stress the policy implications of the results presented in this paper. Indeed, whenever ex-ante corruption is introduced in the analysis, interesting consequences arise.

**Privatization of law enforcement**

Consider an ex-ante corrupt regime. Moreover suppose that (1.25) is negative, in which case the inspectors prefer to ask for a low bribe by implicitly selecting the pooling equilibrium. Given this setting, all citizens will accept the ex-ante agreement, paying a low bribe $b_a$. Suppose that the government decides to shiftily fight corruption by privatizing law enforcement ($r = 1$). In this case, the equilibrium may shift from a pooling to a separating one. This is driven by the fact that having poor taxpayers to reject the ex-ante bribe is more convenient, being the collection-based bonuses very high; a fraction $\theta$ of citizens will accept to pay the ex-ante bribe, while the remaining $1 - \theta$ individuals will prefer to stay clean. As a consequence, ex-ante corruption would be less widespread but at the same time the bribe level will be higher; rich types are going to be worse off because of a higher bribe, while the global level of corruption may increase or decrease depending on the parameters. These reforms may have perverse effects on the "level of corruption", defined here as the amount and frequency of bribes-exchange: on the one hand, shifting from a pooling equilibrium to a separating one implies a higher ex-ante bribe, on the other hand, it reduces the frequency of corruption.

**Transition from developing to developed societies**

1) Consider an ex-post corrupt regime. Moreover suppose penalties ($g$, $c$) and enforcement ability ($l$) are low enough such that $t \geq \frac{((g+c)^{14}}{(l-1)}$. This starting condition may refer to a developing country characterized by low income and inefficient deterrence of collusive activities.

---

\[^{14}\text{This condition guarantees that the likelihood of ex-ante corruption is increasing in taxpayer’s income.}\]
During the development process, increases in either $y_h$ or $y_l$ correspond to increases in per-capita income and lead to a richer society as a whole. At the beginning of the development stage, we would observe widespread ex-post corruption occurs, a type of corruption which is quite visible. All taxpayers accept to pay the ex-post bribe and the bribe level is relatively low: the typical speed money case, associated with high evasion. If $y_h$ ($y_l$) increases beyond a certain threshold the regime is going to switch to ex-ante corruption; this type of corruption is a less visible one but still fundamental\(^\text{15}\).

2) Consider an ex-ante corrupt regime where (1.1) doesn't hold. Suppose (1.25) to be positive, therefore society is located in a pooling equilibrium. During the development process and for a given level of inequality, increases in $\theta$ correspond to increases in per-capita income and lead to a richer society as a whole. At the beginning of the development stage, we would observe widespread corruption, all taxpayers accept to pay the ex-ante bribe and the bribe level is relatively low. As a consequence of the development process we would observe an increase in $\theta$. After a certain threshold, the regime will shift to a separating equilibrium consistent with a lower frequency of corruption and lower evasion. Corruption occurs only with probability $\theta$ and becomes more expensive. Moreover it is now exclusively related to rich individuals.

1.4 Conclusion

The contribution of this paper is to study the design of incentives mechanisms for corruptible inspectors, considering the effect of these incentive schemes on corruption, taxpayer compliance and tax revenue. Incorporating the possibility for the collusive contract to be established both before and after inspection, this paper offers new policy perspectives concerning the effect of privatization of law enforcement and economic development on corruption.

\(^{15}\)For example, Young et all (2001) find evidence that the fraction of individual income tax returns audited is significantly lower in districts that are important to the president electorally. These findings suggest the presence of ex-ante corruption in the contest of american IRS.
Bibliography


Chapter 2

Self Reporting Reduces Corruption in Law Enforcement

2.1 Introduction

The literature on law enforcement has long noted the widespread presence of self-reporting of criminal acts. Offenders choose to admit their misdeeds directly when they know the government will be lenient to them: they are better off paying a reduced fine for certain rather than facing a harsher sentence if detected. The government also prefers when law-breakers report their own crimes because it reduces enforcement costs. The theory behind self reporting in law enforcement has already been explored by Malik (1993) and Kaplow and Shavell (1994) (henceforth KS)\(^1\). Using the Becker model (1964), KS demonstrated that self reporting allows the government to save money by reducing the number of officers needed to monitor the population\(^2\). These cost savings hinge on two assumptions. First, that the enforcement agency can easily separate those who reported their crime from the rest of the population. This allows officers to concentrate their attention on the non-reporting population. Second, that since the monitored population has shrunk, fewer officers are employed for a given level of deterrence.

\(^1\)See also Polinsky and Shavell (2000) for a more general discussion on the law enforcement theory in economic literature.

\(^2\)A second advantage of self reporting is to reduce the risk borne by individuals. In the absence of self reporting, individuals would face uncertain sanctions. Innes (1999) shows that self-reporting enjoys an additional economic advantages if ex-post benefits of remediation are considered. In this case, the violator can undertake remediation in order to reduce the harm caused.
When both of these assumptions hold, KS finds that self reporting benefits law enforcement. Otherwise, self reporting may not be beneficial.

Despite these restrictions, self reporting is often found in areas of law enforcement that violate these assumptions. Hunting licenses in African national parks, nature reserves or game reserves provide one such compelling example. In many protected areas licensed hunting is introduced as a way for park authorities to cull wild animal populations in a controlled manner. However, in several African countries like Malawi, Tanzania and Zambia, hunting permits are used as a form of self reporting for a crime – poaching – that is difficult to control. For instance, in 1998 Tanzania introduced animal hunting licenses in game reserves and national parks during a period of falling wild animal populations, intense poaching and worries about enforcement efficacy. Since then, poaching has decreased dramatically and wild animal populations have recovered (Lamotte, 2008)\(^3\). In contrast, poaching in Kenya (where there are no licenses) has continued at sustained levels, and animal populations in the national parks continue to decline.

Despite the wide adoption of hunting permits in African parks, this form of self reporting cannot be justified by the standards set in KS. Hunters who purchase licenses can legally shoot wildlife within the park, but rangers cannot easily distinguish them from poachers. Because of this, the licenses do not affect how the park rangers patrol, and do not reduce the number of rangers needed. Why then do we observe such instances of self reporting in situations where enforcement cannot be reduced?

Our answer to that question is that self reporting has another important advantage than the ones highlighted by KS: self reporting can be used as a tool against corruption in law enforcement\(^4\).

In this paper, we show that when officers are corruptible, self reporting reduces revenue losses associated with corruption. The problem we consider is the following: enforcement of-

\(^3\)In Tanzania it is possible to legally hunt the vast majority of wild animals, if one is willing to pay the right price. A license to shoot a lion costs $40,000 and one for a leopard $12,000. The 1998 reform legalized hunting for local populations as well, with licenses selling at a much lower cost.

\(^4\)Whether the difference in the policy outcomes between Tanzania and Kenya can be prescribed to an increase in patrolling efficiency and corruption reduction is an issue that is worth exploring in further detail. That some legalization of wildlife hunting and trade can be helpful and beneficial for conservation is a concept that is not lost for several environmental groups. In their 2008 joint report, TRAFFIC and the WWF state that "Policies that criminalize the wild meat trade have not been effective to bringing it under control...greater consideration of alternative management scenarios, including legalizing hunting and trade of certain wild species for meat is therefore required" (Roe, 2008).
cers can accept bribes from apprehended wrong-doers in exchange for letting them go, thus challenging legal enforcement. To prevent this, the government may be compelled to offer incentives designed to keep enforcers honest. Thus, corruptible officers are able to extract payments from either miscreant or government. Self reporting provides an alternative to the offender, who can avoid paying either bribe or the full fine by admitting culpability and paying directly to the government. When the individual chooses this alternative, he avoids the law officer entirely, who –by being excluded– can claim neither bribe nor bonus.

From this perspective, the role of self reporting in reducing corruption is similar to the role that pre-paid or ‘top-up’ systems have in many utility companies in developing countries. Companies face the problem of dishonest bill collectors who collect bribes rather than payments. In Tanzania, customers could consume electricity for years without ever paying their bills. This was quite common until TANESCO introduced ‘smart’ electronic meters which require payment before supplying electricity, bypassing the corrupted workers. This system was borrowed from the cell phone industry and has been successfully adopted in the water, telecommunications and trash disposal sectors. In India, the *tatkal* system was introduced by the phone company to provide people with an alternative to paying bribes for quick phone line installations. The Bharat Sanchar Nigam company guaranteed speedy connections to those willing to sign up and pay into the program. The same *taktal* system was later adopted by the railways as a way for customers to avoid long waiting lists or paying bribes to intermediaries for a quick ticket.

A second issue we consider in this paper is that bribes do act as de-facto fines which discourage criminality. Because of that, certain governments may prefer an enforcement system based on corruption rather than one based on legal fines (Besley and McClaren 1993). Our paper shows that introducing self reporting benefits both clean and corrupt systems. When the system in place is clean, self reporting lowers the cost of incentives; when the system is

---

5 Self reporting has been found to have other important properties that are not discussed here. Motta and Polo (2003) show that leniency programs deter long term illegal relationships by creating a prisoner’s dilemma. In this regard, Buccirossi and Spagnolo (2005) show that leniency programs could help sustain occasional sequential illegal transactions.

6 Top-up systems or pre-paid services are different from self reporting in one important aspect: the client has no choice but to pre pay if he wants the service. In law enforcement, miscreants always have a choice not to self report.

7 While bribery discourages unlawful acts, it is true that it is not necessarily an effective deterrent when compared with a system based on fines (Becker and Stigler, 1974; Polinsky and Shavell 2001).
corrupt, it increases government revenues. Furthermore, we also show that once self reporting is introduced, it is always preferable to eliminate bribe exchanges and replace a corrupt regime with a clean one. Thus, governments can use self reporting as a way to clean up corrupt enforcement agencies. This result is true under specific assumptions which we relax in the later portion of the paper.

The rest of the paper is organized as follows: section 2.2 discusses the model and analyses the impact of self reporting on law enforcement and corruption; section 2.3 discusses the implications of some model assumptions; section 2.4 concludes.

2.2 The Model

2.2.1 Structure

There is a measure 1 of risk-neutral citizens who cause a harm to society of $h$ if they commit an unlawful act or crime. Each citizen derives a private gain $x$ from committing the act; the gain is distributed with a continuous density function $g(.)$ and cumulative distribution function $G(.)$. To minimize the number of crimes the government employs $p$ officers in the police force. They are responsible for monitoring the population and reporting violators to a court of law, which in turn imposes fines.

In the absence of any other consideration, citizens choose to commit the unlawful act only when their private benefit exceeds the expected sanction. Enforcement determines a threshold level of gain $\hat{x}$ such that only individuals whose private benefit exceeds $\hat{x}$ commit the act. The harm to society due to criminality is $[1 - G(\hat{x})] h$.

Enforcement and corruption

Monitoring by the police force is costly. The government pays a base wage $w \geq w^*$ to each officer, where $w^*$ is the reservation wage. Officers are risk neutral individuals who are potentially corruptible, meaning that they will consider accepting bribes. The pay cannot fall below their

---

8We refer to the police officer and the citizen respectively as "she" and "he".
reservation wage\(^9\).

Each officer is randomly matched with a citizen and learns whether he is an offender. When an offense is uncovered, the officer can report the violation to the judiciary system, which levies a fine \(f\) to the offender. Alternatively, the officer may offer a bribe \(b\) to the officer in exchange for her silence. We assume that if \(b \leq \sigma f\), where \(\sigma \in [0, 1]\), such bribe goes undetected or, if detected, unpunished. If a bribe is larger than \(\sigma f\), then the collusion is uncovered, the officer loses her wage, and the offender is made to pay the full fine \(f\). Because of this, officers are willing to consider bribes if \(b \leq \sigma f^{10}\).

The parameter \(\sigma\) plays an important role in the comparative statics of the model. We think of \(\sigma\) as a parsimonious description of how easy it is to detect collusion. The case where \(\sigma\) is small or zero corresponds to transparent societies where even small bribes are not tolerated; a high \(\sigma\) instead corresponds to potentially more corrupt societies, where even outrageous instances of bribery are not prosecuted\(^11\).

Aside from this exogenously given detection function, the government has no other ‘sticks’ to prevent corruption. The only anti-corruption measure would be a ‘carrot’: the payment of bonuses or incentives \(i\) that are paid when an officer reports an offender. Because acceptance of the bribe implies foregoing the incentive, the officer will consider only bribes that have the following characteristic: \(i \leq b \leq \sigma f\).

Finally, consider the choices available to an offender: if he is caught, he will prefer the payment of the bribe when such payment is less than the fine, i.e. \(b \leq f\).

The agreement on the bribe is reached through a bargaining process, which we will specify later. Regardless of how the bargain power is distributed between the two parties, to fully

---

\(^9\)This ‘limited liability’ assumption is reasonable when wages are understood in a more general sense to include ‘per officer costs’, which include expenditures on equipment (cars, computers, weapons) which cannot be eliminated without reducing the efficiency of the police force.

\(^{10}\)It is common in the literature on corruption to assume that the probability of detecting bribe exchanges is a choice variable of the government. We abstract from it in order to keep the model simple. For an analysis of endogenous bribe detection, see Mookherjee and Png (1995), Polinsky and Shavell (2001). In Adving and Moene (1990), the probability of detection depends on how corrupt the overall system is. Overlapping responsibilities among different bureaucrats may also increase the chances of detection and the size of bribes, since officials may need to solve a coordination game (Rose-Ackerman 1994). See Bhardan (1997) for an insightful review.

\(^{11}\)The parameter \(\sigma\) could also reflect the level of transactional imperfections between the colluding parties, i.e. the offender incurs a cost \((1 + \tau) b\) in transferring a bribe \(b\) to the inspector. In this case, once the offender is caught, he will consider offering a bribe only if \(b \leq \frac{1}{(1 + \tau)} f\), where \(\sigma = \frac{1}{(1 + \tau)}\). Notice that the transaction cost \(\tau b\) is equivalent to a deadweight loss in terms of social welfare, requiring a slight modification of the model presented here. See Tirole (1992) for a review of this transaction cost strand of the literature.
eliminate corruption the government must set incentives high enough according to the following no-collusion condition:

\[ i^{nc} \geq \sigma f. \] (2.1)

This result can be intuitively explained in the following fashion: to begin with, notice that the offender and the officer are willing to exchange a bribe only if \( i \leq b \leq \sigma f \). The offender must provide the officer with a bribe at least equal to value of the incentives offered by the government\(^{12}\): when the government sets \( i \) equal to \( \sigma f \), it is no longer convenient for both the officer and the offender to engage in a collusive agreement because it would certainly be uncovered.

In this simple model with homogeneous officers, there are only extreme outcomes. When \( i < i^{nc} \), all officers accept bribes, and all criminals offer them. Since delinquents are never reported, the government neither pays incentives, nor levies fines. When \( i \geq i^{nc} \), no bribes are ever exchanged, and therefore all violations are reported, all fines are levied, and all incentives are paid. We label these two states by the index \( j = c, nc \).

**Government expenditures**

Government expenditures \( B_j \) consist of wage payments net of revenues from fines collected in state \( j \). \( B_j \) is raised through distortionary taxation from the citizenship. Taxes cost taxpayers \( (1 + \lambda)B_j \), where \( \lambda \) is a parameter that measures the size of the dead weight loss. When \( \lambda = 0 \), resources are costlessly shifted from taxpayers to the government, resulting in no welfare losses; when \( \lambda > 0 \), there is a welfare loss equivalent to \( \lambda B_j \).

\(^{12}\)Notice that by offering higher incentives the government is effectively modifying the inspector’s outside option to colluding with the offender. The offender must forgo some of his rent to outbid the government offer.
Social welfare

The government maximizes a weighted average of the welfare of all members of society, which include both civilians and enforcement officers. The social welfare in state \( j \) is

\[
W_j = \int_{\hat{x}_j}^{\infty} (x - h)g(x)dx - \lambda B_j \quad \text{for } j = c, nc.
\]

(2.2)

The canonical approach in KS or Polinsky and Shavell (2001) doesn’t consider a government budget which include the revenues from fines collected and the effect of distortionary taxation. From this perspective, our framework is similar in nature to setups more commonly found in the literature of corruption, such as in Besley and McClaren (1993) and Laffont and Tirole (1993). This departure is justified by the nature of the problem we are analyzing: corruption affects efficiency and depends on incentives, and our paper analyzes the trade-offs between costly incentives and police force efficiency.

In the next section, we consider the welfare functions for \( j = c, nc \) under the assumption that self reporting is not possible. The government can affect the state \( j \) and the level of welfare \( W_j \) by changing the incentive paid to officers and their number \( \rho \).

2.2.2 No Self Reporting

Clean Regime

Consider the choice faced by a citizen when the regime is clean, i.e. officers are never corrupted (because they are paid high incentives, \( i \geq i^{nc} \)). Since his probability of being audited is \( p \), by committing the act he expects to pay a fine of \( pf \). This determines the threshold gain from the act, \( \hat{x}_{nc} \):

\[
\hat{x}_{nc} = pf.
\]

The threshold determines the number of unlawful acts, \( 1 - G(pf) \), and is also the probability

---

\(^{13}\)The fine \( f \) is also a policy instrument that is optimally set by the government. As in the rest of the enforcement literature, the fine is always maximal: \( f \) is set to coincide with the wealth of the citizen. For a proof, Becker (1968), Kaplow and Shavell (1994), Polinsky and Shavell (2001). There are some exceptions to this: Malik (1990) shows that when criminals engage in detection avoidance, fines are not maximal. Livernois and McKenna (1999) claim that under plausible condition higher compliance rates are achieved with lower fines. However, Ines (2001) shows that when self reporting is introduced, there is no need to engage in avoidance and therefore the Becker principle of maximal fines applies again.
that a given audit uncovers an offender. The threshold also helps to determine the expected wage \( w_{nc}^E \) for the officer, since she receives incentives only if an offender is caught:

\[
w_{nc}^E = w_{nc} + i \left[ 1 - G(pf) \right]. \tag{2.3}
\]

Next, consider the government budget. The government hires \( p \) officers, pays them \( w_{nc}^E \) on average, and collects fines from caught miscreants. We assume that every case brought to justice requires red tape and a lengthy bureaucratic process since guilt must be assessed in a court of law. Such process destroys a fraction \( (1 - \alpha) \) of the fine income. Hence, only \( \alpha f \) of the fine enters as revenue to the government\(^{14}\).

Accounting for wage expenses and fine income, the budget takes the form

\[
B_{nc}(p, w, i) = p \left\{ w_{nc}^E - \alpha f \left[ 1 - G(pf) \right] \right\}.
\]

Where \( w_{nc}^E \) is defined by (2.3), and subject to the no-collusion bonus pay (2.1). The budget is minimized when \( w_{nc}^E \) is minimized, that is, when officers’ base pay is the reservation wage \( w^* \) and the no-collusion constraint (2.1) binds. The objective function to be minimized is then reduced to a function of \( p \) only:

\[
W_{nc}(p) = \int_{\text{pf}}^{\infty} (x - h)g(x)dx - \lambda p \left\{ w^* + f(\sigma - \alpha)[1 - G(pf)] \right\}. \tag{2.4}
\]

The three terms in the welfare function are the following: the first term is the welfare loss due to illegal acts; the second term is the social cost of hiring \( p \) officers which are provided a base wage equal to \( w^* \); the last term is the net revenues to the government, that is, fine income minus incentive pay.

**Corrupt Regime**

Suppose that incentives are too low, i.e. \( i \leq \sigma f \): in this case there is scope for bribe exchange and the regime is corrupt. The agreement on the bribe is reached through Nash bargaining, with weights of \( \mu \). The equilibrium bribe is given by the following expression:

\[^{14}\text{A modified version of costly imposition of fines is found in Polinsky and Shavell (2000).}\]
\[ b = (1 - \mu)i + \mu f. \] (2.5)

The potential law breaker’s choice is either to do nothing, or commit the act and pay with probability \( p \) a bribe \( b \) as defined by (2.5). The threshold condition is

\[ \hat{x}_c = pb. \] (2.6)

Officers have two sources of income: their base wage and the bribes. Their expected earnings are then

\[ w^E_c = w_c + b[1 - G(pb)] \]

In this scenario, the government pays only base wages to its officers and receives no fine income, so the government budget is

\[ B_c(p, w, i) = pw_c. \]

Note that, in this case, incentives are offered but never paid out. However, they do play a role: they raise the equilibrium bribe and therefore increase the compliance threshold (2.6). In the optimum, the government increases \( i \) to the point where bribes are maximal (\( b = \sigma f \)) and \( \hat{x}_c = \sigma pf \). The corrupted loss function is then

\[ W_c(p) = \int_{\sigma pf}^{\infty} (x - h)g(x)dx - \lambda pw^*. \] (2.7)

**When is the corrupt regime optimal?**

The government can move from one regime to the other by modifying the incentive pay \( i \). The final outcome is determined by maximizing both \( W_{nc} \) and \( W_c \) in equations (2.4) and (2.7), and then choosing whichever is larger.

In this stylized model, the choice of regime is largely driven by the relative values of \( \alpha \) and \( \sigma \). When \( \alpha \geq \sigma \), officers collect in fines more than they cost in incentive pay, so it is never optimal to let these profitable workers collect bribes. When \( \alpha < \sigma \), fines collected are not sufficient to
cover the expense of keeping officers honest. It may be better to let them collect bribes\footnote{Governments often exert great effort to keep highly profitable enforcement agencies (such as tax administrations and customs) as honest as possible, even amid the widespread corruption of other forces (i.e., police forces and anti-crime units).}.

Aside from the profitability of officers, the government also finds that bribes are less desirable because they limit deterrence: for any given $p$, there are more criminals in a corrupt regime than in a clean one. Still, when society is highly tolerant of corruption ($\sigma$ is high), bribe amounts are high, and that provides a relatively large deterrence. Hence, a society characterized by high $\sigma$ could be better off adopting an enforcement system based on bribes.

Whichever regime is chosen, officers benefit from their corruptibility, since they are able to supplement their wage with rents in the shape of bonuses or bribes.

### 2.2.3 Self Reporting

We now introduce self reporting in both regimes $j = nc, c$. Following KS, we assume that an individual who committed an unlawful act may admit it before the inspection has taken place. As a reward for the admission, the judiciary imposes a reduced or discounted fine $r$, which is less than the fine $f$. By self reporting and paying $r$, the individual avoids paying either the fine $f$ or a bribe $b$. However, we diverge from the original Kaplow and Shavell model by assuming that those who self reported do not necessarily avoid enforcement audits, because officers cannot easily distinguish them from the general, non-self reporting population. By doing this, we abstract from the ‘enforcement reduction’ aspect of self reporting.

In general, the reduced fine follows a different judicial path than the full fine. A citizen who is reported to the judiciary by a law enforcer has the right to defend himself in a court of law, so demonstrating his culpability is time consuming and expensive. The same cannot be said when a person reports his unlawful act: by confessing, he gives up the right to proclaim his innocence. As a consequence, the bureaucratic procedure needed to process the reduced fine is more efficient and less expensive. This point, already highlighted by KS, adds to the appeal of self reporting. However, we prefer to abstract from it and show that the public finance benefits we highlight in this paper do not depend on the superior efficiency of self reporting. For simplicity, we assume that the government is able to appropriate only $\alpha r$ of the reduced fine,
the rest being lost to the bureaucracy. All results of the paper hold the same if the government is able to appropriate \( \alpha_r r \), where \( \alpha_r > \alpha \).

**Clean Regime**

Suppose that \( i \geq i^{nc} \), so that the regime is clean and there is no possibility of bribing officers. Then, an individual who commits an unlawful act can either accept the chance of being caught and pay a fine \( f \), or reports himself and pays the reduced fine \( r \) - whichever is more convenient (less expensive) in expectation to him. An individual with private gain \( x \) may commit the act if his private benefit from the act exceeds the cost:

\[
x \geq \min[r, pf] = \check{x}_{nc}^sr,
\]

where \( r \) is now part of the set of policy instruments available to the government. To get the optimal level of \( r \), consider first the case \( r > pf \). Because self reporting is more expensive than the expected full sanction, criminals do not report their act. In this case, self reporting is possible, but no one employs it.

Now suppose \( r \leq pf \). All unlawful acts are reported, and individuals pay only reduced fines to the government. Since \( \check{x}_{nc}^sr = r \), the total number of crimes committed is \( 1 - G(r) \). The welfare achieved is

\[
W_{nc}^{sr}(r, p, w, i) = \int_r^\infty (x - h)g(x)dx - \lambda \left\{ pw_{nc}^{sr,E} - \alpha r[1 - G(r)] \right\}, \tag{2.8}
\]

where \( w_{nc}^{sr,E} \) is again the expected wage paid per officer.

Proposition 1 explains the optimal self reporting policy.

**Proposition 2.1** When self reporting is adopted and corruption is not allowed:

(i) \( r = pf \) (the reduced fine is equal to the expected full fine when bribing is not possible);

(ii) \( i \geq \sigma f \) (the no-collusion condition (2.1) is implemented);

(iii) \( w_{nc}^{sr,E} = w^* \) (average wages, including incentive pay, are equal to the officers’ outside option).

**Proof.** see the Appendix.  ■
With this proposition, we can replace $r$ with $pf$, and $w_{nc}^{sr,E}$ with $w^*$. The welfare function to be maximized becomes

$$W_{nc}^{sr}(p) = \int_{pf}^{\infty} (x - h)g(x)dx - \lambda p \{ w^* - \alpha f[1 - G(pf)] \}. \quad (2.9)$$

The first term indicates the welfare loss due to crime; the second term represents the wage bill $w^*$ paid to officers; the third term is the revenues from self reporting.

Self reporting reduced the expected wage received by officers from $w_{nc}^E$ to $w^*$. This is because it separates innocent citizens from offenders without the direct intervention of the police force, and therefore deprives them of their bonuses\textsuperscript{16}. Still, the presence of officers does provide a credible threat against those criminals who fail to self report: such criminals could get caught and subjected to pay the full fine $f$. A simple example of this ‘threat’ can be found in the enforcement of parking regulations. Cities allow drivers to park in certain areas only if they pay a certain reduced fine at the curb by feeding a meter. People feed the meter in the off chance that a parking inspector passes by and fines those who did not.

**Corrupt Regime**

Suppose the regime is corrupt. The choice now facing a guilty individual is between self reporting $r_c$ and paying the bribe $b$ with probability $p$, where $b$ is still defined by (2.5). The threshold level of gain needed to choose between committing and not committing the act is

$$\tilde{x}_c^{sr} = \min[pb, r_c].$$

so that in order for self reporting to be used, $r_c \leq b$ and the welfare achieved is

$$W_c^{sr}(r_c, p, w, i) = \int_{r_c}^{\infty} (x - h)g(x)dx - \lambda \{ pw_{nc}^{sr,E} - \alpha r_c[1 - G(r_c)] \}, \quad (2.10)$$

where in this case, $\alpha r_c[1 - G(r_c)]$ is income earned from self reporting fines. We now establish the optimal corrupt policy under self reporting:

\textsuperscript{16}The same property of self reporting can be used in other settings: in a separate paper, we use a version of self reporting to create a collusion-proof mechanism that works in designing optimal contracts (Burlando and Motta 2008).
Proposition 2.2 When the regime is corrupted and self reporting is adopted:

(i) \( r_c = pb \)

(ii) \( b = \sigma f \)

(iii) \( w_{c, E}^{sr} = w_c = w^* \)

Proof. see the Appendix.

With this proposition, the objective function reduces to a function of \( p \) only:

\[
W_{c}^{sr}(p) = \int_{\alpha p f}^{\infty} (x - h)g(x)dx - \lambda p \{w^* - \alpha \sigma f [1 - G(\sigma pf)]\} \tag{2.11}
\]

Again, the first term is the social loss due to criminality, and the term in parenthesis is the wage paid net of revenues from self reporting. Note that, in essence, the revenues from self reporting are in fact the bribes that are now diverted from the hands of corrupted officials into the state’s coffers. This bribe diversion has no direct enforcement effect, but it has a significant revenue effect. It also has no direct effect on the corruptibility of officials, who would take bribes if in the position to do so.

2.2.4 Comparative Statics

We are now able to compare the different regimes with and without self reporting, and demonstrate the main result of the paper: that self reporting in a clean regime is the best enforcement policy. To reach that conclusion, we show in the next two lemmas that self reporting is more efficient in both clean and corrupt regimes. We start by evaluating self reporting in a clean regime.

Lemma 1 Self reporting improves welfare under a clean regime.

Proof. It is sufficient to show that for any \( \tilde{p} \) that is chosen under a policy without self reporting, a self reporting policy \( \tilde{r} = \tilde{pf} \) yields greater social welfare. Denoting the welfare without self reporting by \( W_{nc}(\tilde{p}) \) and the welfare with self reporting policy \( \tilde{r} \) by \( W_{nc}^{sr}(\tilde{p}) \), the increase in welfare under self reporting is

\[
W_{nc}^{sr}(\tilde{p}) - W_{nc}(\tilde{p}) = \lambda \tilde{p} \alpha \sigma f [1 - G(\tilde{pf})] > 0.
\]
The improvement is caused not by a reduction of crime rates (those remain constant), but by savings in enforcement expenditures. In fact, the difference in welfare is entirely made from the incentives that are no longer paid out to officers. Since trasgressors report their harmful acts directly, they do not need to be inspected, and so officers do no receive any bonuses.

Next, consider the corrupted regime with and without self reporting.

**Lemma 2** Self reporting improves welfare under a corrupt regime.

**Proof.** For any $\tilde{p}$ that is chosen under a policy without self reporting, the government can introduce self reporting by choosing $\tilde{r}_c = \sigma \tilde{p} f$. The gain in welfare is given by the difference between (2.7) and (2.10):

$$W_{sr}^c(\tilde{p}) - W_c(\tilde{p}) = \lambda \tilde{p} \sigma f [1 - G(\sigma \tilde{p} f)] > 0$$

Moving to a regime of self-reporting allows the government to add an income stream: bribes from officers that now find their way into the hands of the government. The principle established for the clean regime then translates also to the corrupt regime: officers cannot earn rents under self reporting.

The two lemmas show that self reporting improves welfare both under a clean and a corrupt regime. We now establish the third result: once self reporting is introduced, the clean regime always dominates.

**Proposition 2.3** When self reporting is introduced, for any policy that induces corruption, there exists another policy under a clean regime that is strictly preferable.

**Proof.** Consider a regime of corruption with self reporting, where the number of officers is $\tilde{p}$ and the reduced fine is $\tilde{r}_c = \sigma \tilde{p} f$. We now show that the government would be strictly better off if it eliminates corruption (by choosing $i \geq i^{nc}$), reduces the workforce from $\tilde{p}$ to $\bar{p} = \sigma \bar{p}$, and keeps the self reporting fine at $\tilde{r}_c$. The change in welfare is then

$$W_{sr}^{nc}(\bar{p}) - W_{sr}^c(\tilde{p}) = \lambda \tilde{p} (1 - \sigma) w^* > 0$$
The fact that corruption is never optimal should not come as a surprise: the main reason for allowing corruption when there is no self reporting is that the government would forgo the expense of paying bonuses to its officials; but under self-reporting, bonuses are never paid, and in either regime officers earn their outside wage $w^*$ only. With the main benefit of corruption gone, what is left is the negative aspect of corruption, namely, that it reduces deterrence. But the level of deterrence under a corrupt regime can be achieved with lesser expense (fewer officers) under a clean one.

2.3 Discussion and Limits to the Theory

2.3.1 Weakening of enforcement effort

The first limitation of self reporting in law enforcement is that it may create moral hazard problems of its own among the officers. In our model, the probability of detecting a criminal act does not depend on the officers’ effort. In reality, the intensity of effort exerted is likely to change the chance that an unlawful act is uncovered. If all criminal acts were self reported, law enforcers would see no benefit in working hard, and this would reduce the probability of detection for everyone. How this weakening in enforcement impacts the overall equilibrium and the implementation of self reporting depends on how effort is modeled.

While this limitation may be substantial in some settings, it may not be as important in instances where either officer exertion is unimportant or it is easily monitored by the enforcement agency. Effort may be unimportant when the officer must perform many tasks, and only one of them is to check whether an individual has committed a certain crime. For example, customs officials at a port of entry perform a series of tasks on a random selection of incoming containers, such as ensuring that contents match the documentation. In the process, they may determine whether other regulations have been violated without making significant extra effort: whether all import duties have been paid, whether illegal substances or restricted materials are found. In other instances, where effort matters, the government can monitor effort. For example, many tasks can be standardized and reduced to checklists or forms that must be completed by the officers. Many instances of tax evasion are captured in this way, since officers must first of all
check that forms sent from different sources match the income report.

2.3.2 Adverse selection of officers

A second aspect worth considering is adverse selection among officers. Officers may have different degrees of ability in performing their job: some may have a higher probability of uncovering offenders than others. Clearly, self reporting eliminates these differences, since the chance of encountering an unreported violation is zero for both ‘good’ and ‘bad’ officials. This may be a problem for the enforcement agency if selection is important in other aspects of its activities. For example, the agency may want to observe individual ability so that it can promote good workers to higher ranks. In that case self reporting is still worth it if the agency has other means to measure ability.

2.3.3 Failure to self report

In practice, offenders often fail to self report even when such an option is available. We can think of three reasons for this. First, it may be that individuals have heterogeneous probabilities of detection or levels of risk aversion (Innes 2000)\(^\text{17}\). Second, a person may have more to hide than the crime itself: self reporting on one crime may lead investigators to audit more thoroughly other aspects of a person’s life, the cost of which is not ‘priced in’ the self reporting fine. Thus, a driver may prefer to ‘hit and run’ a bystander rather than stop to help if he is carrying a stash of drugs with him, for which he is guaranteed a harsh punishment. Third, individuals may face some uncertainty when self reporting due to the complexity of the law. He cannot be sure that, after self reported for one crime, he won’t be held accountable for another act which he did not think was illegal.

Whenever these circumstances arise in a way that creates additional heterogeneity in citizens’ preferences, some individuals do not self report. The implication is that in a clean regime the government cannot avoid paying incentives to some of its officers. This reduces the effect of self reporting against corruption. While proposition 5 may therefore be violated, we can show that self reporting remains the optimal policy under either clean or corrupt regime: this

\(^{17}\text{It is also possible that individuals actively engage in avoidance (Malik 1990, Innes 2001); however, unless individuals have different cost functions of avoidance, there is no reason for agents to avoid self reporting.}\)
is because in either regime, the government can reduce the number of officers taking bribes or
bonuses. Moreover, the introduction of self reporting still allows some governments (but not all) to fully eliminate corruption\textsuperscript{18}.

2.3.4 Unconstrained officers

So far we have assumed that officers are constrained by some sort of ‘limited liability’ so that
their base pay cannot fall below $w^*$. A consequence of this is that their expected wages exceed
$w^*$.

Suppose now that the government can adjust the base pay such that, in expectation, officers
are always paid their outside option in either regime:

$$w_{nc}^E = w + E(i) = w^*$$
$$w_c^E = w_c + E(b) = w^*$$

In this case, officers are willing to accept a base pay $w$ and $w_c$ that is lower than their
outside option because they can make up the difference with either bonuses or bribes. Because
the wage can be adjusted downward, expected officers’ salaries equals their outside option even
in the absence of self reporting. Since the elimination of rents to officers was the key property
highlighted in this paper, self reporting ceased to be useful here. Nonetheless, suppose that
inspectors are risk averse. When there is no self reporting, they must be compensated for
bearing the risk of uncertain wages, so that the expected wage exceeds $w^*$ in both clean and
corrupt regime. By introducing self reporting, the government eliminates variations in wages,
and the risk premium with it. This restores the results of the paper\textsuperscript{19}.

2.3.5 More complex anti-corruption policies

Our theory does not allow the government to increase the probability of detecting bribe ex-
changes, nor do we fully allow for fines from bribery\textsuperscript{20}. Some other credible aspects of law

\textsuperscript{18}Notes available from the authors upon request.
\textsuperscript{19}We work out this model in a prior version of our paper, available upon request.
\textsuperscript{20}See Mookherjee and Png (1995) and Besley and McClaren (1993).
enforcement explored elsewhere, such as framing of innocent civilians by law enforcers, should also be included in the theory of self reporting.

2.4 Conclusions

We analyze the role of self reporting as an anti-corruption instrument in the practice of law enforcement when the enforcers are corruptible. Enforcement agencies which suffer from widespread corruption within the ranks are fairly common in many countries of the world, one reason being that cleansing can be painful and expensive. Our paper suggests that when reform aimed to eliminate corruption is implemented in conjunction with self reporting, some of the costs of reform can be eliminated. This is due to the self reporting propriety of cutting enforcement costs by reducing rents to officers.
Bibliography


[21] Rose-Ackerman, S., 1994, *Reducing Bribery in the Public Sector, in Corruption and democracy: Political institutions, processes and corruption in transition states in East-Central*

Appendix: Proof of Proposition 2.1

In what follows, we prove the three statements in Proposition 2.1, i.e. when the regime is clean and self reporting is allowed: (i) \( r = pf \), (ii) \( i \geq \sigma f \) and (iii) \( w_{sr,E}^{nc} = w^* \).

(i) Suppose \( r < pf \). Then, all criminals self report, and the government could slightly decrease \( p \) without changing the number of crimes. Neither the integral nor the last term in (2.8) would change. The second term would decrease if \( w_{sr,E}^{nc} \) is nonincreasing in \( p \). Since all offenders self report, there are no incentives paid and \( w_{sr,E}^{nc} = w \) which is independent of \( p \). Thus, welfare would increase. Since \( r < pf \) is not optimal, it must be that \( r \geq pf \). If \( r > pf \) self reporting is not binding, and therefore the proper welfare function that applies is still (2.4). When \( r = pf \) self reporting is used by criminals and the welfare function is given by (2.8). By comparing (2.4) and (2.8) it can be seen that society is better off when \( r = pf \); see lemma 3 for a detailed proof.

(ii) Suppose \( i < \sigma f \). An offender who does not pay the reduced fine \( r = pf \) expects to be apprehended with probability \( p \). In that case, the officer will accept the bribe \( b \) as defined by (2.5). It follows that the violator’s expected cost when not reporting himself equals \( p[(1-\mu)i + \mu \sigma f] \leq p(\sigma f) < pf = r \). In this case, the regime is corrupt. Therefore it must be that \( i \geq \sigma f \).

(iii) When offenders self report, officers only inspect innocent citizens, and therefore they earn no incentive pay: the only salary paid is \( w \geq w^* \). To maximize equation (2.8), the government sets base wage as low as possible, to \( w^* \).

Appendix: Proof of Proposition 2.2

In what follows, we show that when the regime is corrupted and self reporting is allowed: (i) \( r_c = pb \), (ii) \( b = \sigma f \) and (iii) \( w_{sr,E}^{c} = w^* \).

(i) Suppose \( r_c < pb \). Then, all criminals self report, and the government could slightly decrease \( p \) without changing the number of crimes. Neither the integral nor the last term in (2.10) would change. The second term would decrease if \( w_{sr,E}^{c} \) is non-increasing in \( p \). Since all offenders self report, there are no bribes and therefore \( w_{sr,E}^{c} = w_c \), which is independent of \( p \). Thus, reducing \( p \) increases welfare, which means that \( r < pf \) is not optimal. Hence, it must
be that \( r \geq pf \). When \( r > pb \), no offender self reports, so self reporting is not adopted and welfare function (2.7) applies. It follows that self report binds when \( r = pb \) and social welfare is given by (2.10). By comparing (2.7) and (2.10) it is easy to notice that for any given level of the control variables the social welfare is higher when \( r = pb \); see lemma 4 for the complete proof. We can then rewrite (2.10) as:

\[
W^{sr}_{c}(r, p, w, i) = \int_{pb}^{\infty} (x - h)g(x)dx - \lambda \{pw_{c} - pb[1 - G(pb)]\} \tag{2.12}
\]

(ii) To show that \( b = \sigma f \), we need to show that the government wants to set \( i = \sigma f \). Suppose that \( i < \sigma f \), so that \( b = (1 - \mu)i + \mu \sigma f < \sigma f \). Now consider a raise in \( i \) to \( \tilde{i} \), such that equilibrium bribe is \( \tilde{b} > b \), and a corresponding decrease of \( p \) to \( \tilde{p} \) such that \( \tilde{p}b = pb \). Then, the integral in equation (2.12) does not change, whereas the second term decreases. Thus, it must be that \( b \geq \sigma f \), which means that \( i \geq \sigma f \). If the inequality is strict, then there is no corruption, a situation that we exclude. Thus, \( b = \sigma f \) and \( i = \sigma f \).

(iii) So far, all criminal acts are reported to the government directly, so officers do not earn any bribes. Wages \( w_c \) enter the welfare function (2.12) negatively, so in order to maximize welfare, the base wage needs to be as low as the reservation wage \( w^* \).
Chapter 3

Conditional Delegation and Optimal Supervision

3.1 Introduction

In their seminal paper, Laffont and Tirole (1991) develop an agency-theoretic approach to interest-group politics. They consider a three-tier hierarchy where Congress (P) relies on information supplied by the agency (S) regarding the firm’s (F) technological type\(^1\); F can be either "efficient" or "inefficient". Laffont and Tirole (1991) conclude that the threat of producer protection, i.e. possibility of collusion between F and S, leads to low-powered incentive schemes with respect to the case in which producer protection is ignored. The intuitive reason for this result is the following: to prevent F from bribing S, P must provide S an incentive for reporting the true information. This incentive must be high enough such that the cost to F of compensating S for the income lost by not reporting exceed its stake. In other words, P must compete with F to have S reporting the true signal. To this purpose, P must reduce the efficient type rent under asymmetric information. Therefore, in Laffont and Tirole (1991) collusion reduces social welfare because the inefficient type is given an incentive scheme that is even less powerful than the corresponding scheme in the absence of collusion.

We take Laffont and Tirole (1991) setting seriously into consideration, by focusing on one

\(^1\)We refer to the S and F respectively as "she" and "it".

53
element that could have been naturally implemented in their approach: we allow F to choose between a regime free of supervision over a regime of supervision. Our mechanism works as follows: F can choose between two competing contracts. The first contract, which we call ‘fast contract’ for lack of a better word, directly specifies payouts to F, and involves no supervision at all. If this contract is rejected, then a grand contract involving P, S and F is implemented, in which S inspects F and reports to P.

Assuming that S and F cannot collude before the acceptance of the fast contract, this mechanism allows P to eliminate the cost related to collusion: intuitively, P can use the intervention of S in the grand contract stage as a credible threat for reducing F’s information rent in the fast contract stage, thereby maintaining P’s advantage of having a supervisory agency in the first place.

In the second part of the paper we show that this outcome is achieved conditional on P designing both contracts before F makes a selection. When P designs contracts sequentially or conditionally on the choices made by F, conditional supervision may still bring about better outcomes than the centralized supervision alone.

The key advantage of this model is its simplicity and its applicability. In practical terms, a fast contract is the ‘best case scenario’ contract that can be designed by P without knowing anything about F or S; it can easily be added into a more complex, contingency-based grand contract suggested by the literature.

Our objective here is to present a new mechanism under the very strict assumptions generally made in this strand of the literature, and leave the relaxation of those assumptions to later research. For instance, we retain frictions in the side-contracting stage of the model. While this assumption can be justified in practice (Laffont and Meleu 1997) and derived from models of repeated games (Martimort 1999), subsequent models of collusion in the presence of ‘soft’ information do not rely on them (Faure-Grimaud et al, 2003). Importantly, our model is measured against prior mechanisms under a fully centralized system, in which P contracts directly with both S and F. We then avoid, for now, the debate on whether centralized mechanisms performs better (Celik, 2008) or worse (Baliga and Sjorstrom 1998, Faure-Grimaud et al 2003) than full delegation to S. This is an important question that can be addressed in a more general setting.

\[^2\]In the Tirole setting, delegation and supervision are equivalent.
by the mechanism presented here\textsuperscript{3}.

Finally, it is important to emphasize that our model is restricted to unproductive supervision, since in many instances S is barred from entering into the contract. Our model thus cannot inform the debate on hierarchical models where multiple agents are involved in the production process (McAfee and McMillan 1995; Laffont and Martimort 1998; Mookherjee and Tsumagari 2004).

The rest of the paper is organized as follows. Section 3.2 defines the model and the benchmarks of collusion-free and collusion-proof supervision. Section 3.3 introduces conditional supervision, and measures it against the benchmarks. Section 3.4 considers sequential contracting, in which the fast contract and the grand contract are redacted sequentially instead of simultaneously. Section 3.5 concludes.

### 3.2 Benchmarks

We briefly present a simplified version of Laffont and Tirole (1993) setting - hereafter LT. More precisely, we adopt the framework proposed by Lambert-Mogiliansky (1998). In the exposition, we restrict our attention to those aspects that are relevant for the purposes of our analysis. We address the reader to the original articles for further details and proofs.

To begin with, we consider the case of full commitment, i.e. some technology is available to P so as to preclude renegotiations at any of the interim stages. Under this assumption, results are substantially identical to those in LT. We consider a three-tier hierarchy: F/S/P, all parties being risk neutral. In order to have F producing the good, P must pay a cost

\[
C = \beta - e, \tag{3.1}
\]

where \(\beta\) represents the technology parameter, which can take one of the two values: "efficient" \((\beta)\) with probability \(v\) and "inefficient" \((\overline{\beta})\) with probability \((1 - v)\). F knows the realization of \(\overline{\beta}\). By exerting effort \(e\), F reduces cost of production but it incurs an increasing and convex disutility \(\psi(e)\), where \(\psi' > 0, \psi'' > 0\) and \(\psi''' \geq 0\).

\textsuperscript{3}The conditional supervision presented in this model allows F to choose between a direct contract with P or a centralized contract; delegation can be obtained by having F contract directly either with P (as a direct contractor) or S (as a subcontractor).
S pays F’s costs and it also collects its revenue. \( t \) denotes the transfer from P to F. F’s utility \( U \) defines its participation constraint:

\[
PC : U = t - \psi(e) \geq 0, \tag{3.2}
\]

where we normalized F’s reservation utility to 0.

S receives a payout \( s \) from P. In order to accept the contract, its reservation utility must be met:

\[
PC_s : V = s \geq 0, \tag{3.3}
\]

where \( s \) is the income granted by P to S, while S’s reservation income equals zero. S receives a signal \( \sigma \) about F’s technology. With probability \( \xi \), S learns the true \( \beta \) (\( \sigma = \beta \)); with probability \( (1 - \xi) \), she learns nothing (\( \sigma = \emptyset \)). She reports \( r \in \{\sigma, \emptyset\} \). Finally, S may agree on a side contract with F, which involves receiving a transfer \( b \) by F. In doing so, the latter incurs a cost \( (1 + \lambda) b \), where \( \lambda \geq 0 \) denotes the shadow cost of transfers for F. Given LT setting, collusion can arise only if the retention of information benefits F, which happens in this model only for the efficient F. P observes neither \( \beta \) or \( \sigma \). It observes the cost \( C \) and receives S’s report \( r \). It designs incentive schemes \( s(C, r) \) and \( t(C, r) \) for S and F to maximize expected social welfare.

The timing of the game is as follows: At date 0, P learns that \( \beta \in \{\underline{\beta}, \overline{\beta}\} \) and F learns \( \beta \), where the probability parameters \( v \) and \( \xi \) are common knowledge. In the second stage of the game, date 1, P designs a contract for S and F. At date 2, S receives the signal and learns \( \sigma \). At date 3, S can then sign side contract with F. Next, at date 4, S makes a report to P, and F chooses its effort. Finally, transfers are operated as specified in the contract. Figure 3-1 summarizes.

The cost parameter \( \beta \) together with the signal received by the agency \( \sigma \in \{\underline{\beta}, \overline{\beta}, \emptyset\} \) define four states of the world:
Figure 3-1: Timing of the Game

\[ p_1 = \Pr(\beta = \beta, \sigma = \beta) = \nu \xi, \]
\[ p_2 = \Pr(\beta = \beta, \sigma = \beta) = (1 - \nu) \xi, \]
\[ p_3 = \Pr(\beta = \beta, \sigma = \emptyset) = \nu (1 - \xi), \]
\[ p_4 = \Pr(\beta = \beta, \sigma = \beta) = (1 - \nu)(1 - \xi), \]

where \( p_i \) is the probability of each correspondent state.

Let \( G \) denote the value of the good; we impose \( G \) to be sufficiently large to make production worthwhile in all the states of the world. Having this schedule in place, \( P \) expected net benefit of the project is

\[
\max_{\{e_i, t_i, s_i\}_{i=1,...,4}} W = G - \sum_{i=1}^{4} p_i (t_i + C_i + s_i), \tag{3.4}
\]

where \( C_i \) is defined by (3.1).

### 3.2.1 Collusion-free Supervision (CF)

We first consider the case in which \( S \) always reports truthfully. Henceforth, we denote this regime as the collusion-free regime \( CF \). It corresponds to the case in which \( P \) can directly supervise \( F \). The optimal contract will involve the production of the good in all four states. In
order for this production to occur, both F and S need to agree to the contract, and therefore their participation constraints (3.2) and (3.3) must be met.

Furthermore, the contract must meet the revelation principle: in each state of the world, F must reveal its technology parameter $\beta$ to P. Note that in states of the world 1 and 2, P knows from S the technology parameter of F. When the signal received is $\emptyset$, P must provide an incentive to the efficient F so that F does not mimic the inefficient (high cost) one.

The relevant incentive compatibility constraint (IC) involves the low type (efficient) F

$$IC_f : t_3 - \psi(e_3) \geq t_4 - \psi(e_4 - \Delta \beta).$$

(3.5)

Thus, the collusion-free contract CF is obtained by maximizing (3.4) with respect to constraints (3.2), (3.3) and (3.5). The standard solution (fully discussed in LM) has the following characteristics:

a) F earns zero rents when it is inefficient (state 2 and 4) and when it is efficient but the signal is informative (state 1): $t_i = \psi(e_i)$ for $i = 1, 2, 4$;

b) S never earns any rents: $s_i = 0$ for $i = 1, 2, 3, 4$;

c) when the signal is not informative and F is efficient (state 3), the information rent surrendered to F is $\Phi(e_4)$:

$$\Phi(e_4) = \psi(e_4) - \psi(e_4 - \Delta \beta)$$

d) finally, effort levels of F solve the following first order conditions:

$$\psi'(e_i^*) = 1 \quad \text{for } i = 1, 2, 3$$

$$\nu(1 - \xi)\Phi'(e_4^{CF}) + (1 - \nu)(1 - \xi)\psi'(e_4^{CF}) = (1 - \nu)(1 - \xi).$$

(3.6)

Where the superscript $CF$ denotes the collusion-free outcome. The equilibrium outcome is typical for this type of problems: the efficient F’s effort is always at the optimal level, whereas the inefficient F receives lower-powered incentive in one state of the world, since $e_4^{CF} < e_4^*$. 

To conclude, the equilibrium welfare level when S cannot be bribed is now

58
\[ W_{CF}^* = G - \sum_{i=1}^{3} p_i \{ \psi(e_i^s + \beta_i - e_i^s) - p_4 \{ \psi(e_4^{CF}) + \beta_4 - e_4^{CF} \} - \nu(1 - \xi)\Phi(e_4^{CF}) \}. \] (3.7)

### 3.2.2 Collusion-proof Supervision (CP)

We now proceed with the case in which S and F can collude. When this is the case, there is a possibility of bribe exchanges between S and F.

The problem of bribing arises when F is efficient. Under collusion-free supervision, F would earn a rent of \( \Phi(e_4) \) if S were to report \( r = \emptyset \) to P. In state 1, S knows that F is efficient, and may therefore want to share the information rent with F. It could do so by asking a bribe \( b \) to F in exchange for sending the message \( r = \emptyset \). Under the assumptions of the model so far, F is willing to pay the bribe as long as

\[ b \leq \frac{1}{1 + \lambda} \Phi(e_4). \]

When such an exchange occurs, S foregoes payment \( s_1 \) and receives instead payment \( s_3 \). Thus, to prevent F from bribing the agency, the agency’s income \( s_1 \) contingent on reporting an efficient F \((r = \beta)\) must exceed F’s stake in collusion,

\[ IC_s : s_1 \geq s_3 + \frac{1}{1 + \lambda} \Phi(e_4). \] (3.8)

LM show that with full commitment, there is no loss of generality in focusing on collusion-proof mechanism (henceforth, CP) where (3.8) is met. The mechanism CP is then characterized by a contract that maximizes (3.4), subject to constraints (3.2), (3.3) and (3.5), as in the CF problem. Moreover, collusion-proofness requires meeting (3.8). In that case, the solutions (a) and (d) in the CF problem remain the same. The additional constraint modifies part (b) as follows:

\[ b' \) for \( i = 2, 3, 4; \ s_1 = \frac{1}{1 + \lambda} \Phi(e_4) \]

That is, collusion proofness requires leaving some rents to S in state 1, in addition to the rents to F in state 3.

Part (d) is modified as follows:

\[ d' \]
\[ \psi'(e_i^*) = 1, \quad i = 1, 2, 3 \]

\[ \Phi'(e_{4CP}^C) \left[ \frac{\xi \nu}{1 + \lambda} + \nu(1 - \xi) \right] + (1 - \nu)(1 - \xi) \psi'(e_{4CP}^C) = (1 - \nu)(1 - \xi). \quad (3.9) \]

That is, there is an additional distortion away from efficiency that is due to the possibility of collusion between S and F: note that \( e_{4CP}^C < e_{4CF}^C \). This distortion is due to the trade off between allocation efficiency and F’s information rents in state 3, \( \Phi(e_4) \), as well as S’s transfer in state 1, \( s_1 = \frac{1}{1 + \lambda} \Phi(e_4) \).

Given the solutions indicated from (a) to (d’), the welfare level at equilibrium is

\[ W_{CP}^* = S - \sum_{i=1}^{3} p_i \left\{ \psi(e_i^*) + \beta_i - e_i^* \right\} - p_4 \left\{ \psi(e_{4CP}^C) + \beta_4 - e_{4CP}^C \right\} - \left\{ \frac{\xi \nu}{1 + \lambda} + \nu(1 - \xi) \right\} \Phi(e_{4CP}^C). \]

A comparison of \( CP \) and \( CF \) reveals that \( W_{CF}^* \geq W_{CP}^* \): indeed, for a given \( e_4 \) we have

\[ W_{CF}(e_4) = W_{CP}(e_4) + \frac{\xi \nu}{1 + \lambda} \Phi(e_4). \quad (3.10) \]

From (3.10) we get that the cost of collusive supervision is \( \frac{\xi \nu}{1 + \lambda} \Phi(e_4) \).

3.3 Conditional Supervision–CS

3.3.1 An intuitive Explanation

We now consider the conditional supervision mechanism (CS) as an alternative to the collusion proof mechanism. We will show that our mechanism improves on the collusion proof mechanism, and in fact is able to restore the collusion free outcome. We will present this in two ways. First, we will provide an intuition for our mechanism: it is useful to revisit the payoffs to S and F under collusion proofness (CP) and collusion free supervision (CF). When F is efficient, payoffs to F \( (U) \) and S \( (V) \), contingent on the probability of receiving an informative signal \( (\xi) \), are represented by a matrix:
A couple of aspects are worth noticing. First, note that in expectation the efficient F earns a rent of $(1 - \xi) \Phi(e_4)$ in both regimes, whereas S earns a rent of $\frac{\xi}{1 + \lambda} \Phi(e_4)$ in CP but of 0 in CF. Second, under CP, both F and P have reasons to dislike state 1. F dislikes it because it earns zero rents; P dislikes it because it must pay out S.

In conditional supervision, P can exploit this mutual dislike of state 1: when the efficient F accepts this fast contract, P guarantees a rent of $(1 - \xi) \Phi(e_4)$, and no supervision occurs. If F instead reveals to be inefficient, P then calls in S, a grand contract is implemented, and the game is played as in CP.

The outcome of the game is as follows: all efficient F choose to adopt the fast contract, under no supervision, and earn rents of $(1 - \xi) \Phi(e_4)$, whereas S is paid its outside option. All inefficient F choose instead to enter a CP contract with supervision. We have then a separating equilibrium.

To show that there are no profitable deviations, suppose that an efficient F chooses not to enter a fast contract. Then, it will receive supervision, and with a probability of $1 - \xi$ it will get caught and receive zero rents. Note that, since the lottery CP is collusion-proof, F cannot hope to collude with S.

Furthermore, it must be the case that inefficient F choose not to sign the fast contract. This is clearly the case in our model. A sufficient proof is that the inefficient F never chooses the efficient contract when this contract provides rents of $\Phi(e_4)$ to the efficient F. Therefore, this inefficient F will never choose a fast contract that provides even smaller rents of $(1 - \xi) \Phi(e_4)$.

A final aspect is worth noticing: if F happens to be risk averse, conditional supervision brings about a further gain with respect to both CP and CF mechanisms. Indeed, CS provides full insurance to the efficient F, which would otherwise face the supervision-lottery. As a
consequence, the efficient type is willing to surrender some of its rent in the fast contracting stage. P has to guarantee a reduced rent of

\[ (1 - \xi) \Phi(e_4) - j(\cdot), \]

where \( j(\cdot) \) denotes the risk premium. It is easy to notice that \( CS \) implements an outcome which is even better than the \( CF \) one. It follows that \( CS \) should be implemented even in the absence of collusion.

### 3.3.2 A Formal Model

**The contract**

P introduces two contracts: a fast contract between itself and F, and a grand contract which also involves S. In the fast contract, P makes a take-it-or-leave-it offer to F \((C_0, t_0)\) that does not depend on any input of either F or S.\(^4\)

If the contract is rejected by F, then the grand contract is offered. At this point, S receives the signal \( \sigma \in \{\varnothing, \beta\} \), which is also known by F, and makes a report \( r \in \{\varnothing, \beta\} \). The signal-contingent contract specifies a set of payments \( \{C(r), t(r), s(r)\} \). If either F or S rejects this grand contract, we have shutdown of production. The timing of the subsequent steps is as before. Figure 3-2 summarizes.

We stress the importance of timing here: the two contracts are drawn together, before the realization of the signal and before F agrees or rejects the fast contract. In the later section, we show what happens when this timing assumption is violated, and the two contracts are drawn sequentially.

With this setup, the state space has now expanded to eight possible states and is represented by a triplet \( \{\beta, \sigma, \kappa\} \), where \( \kappa = \{0, 1\} \) denotes whether F accepts (1) or rejects (0) the fast contract.

In practice, there are only five relevant states of the world; the other states are off of the

\(^4\)In practice, we do not think that the contract involves the supervisor at all. Technically, the supervisor can accept or reject the contract. If the supervisor rejects the contract, our model requires shutdown of production.
equilibrium paths. These five states are:

\[
\begin{align*}
0 & = \{\beta, \emptyset, 1\} \\
1 & = \{\beta, \beta, 0\} \\
2 & = \{\overline{\beta}, \overline{\beta}, 0\} \\
3 & = \{\beta, \emptyset, 0\} \\
4 & = \{\overline{\beta}, \emptyset, 0\}
\end{align*}
\]

The two states where \( \kappa = 1 \) and the signal is informative are excluded from our equilibrium because, when \( \kappa = 1 \), \( P \) proposes to not involve a \( S \); this leads to the impossibility by \( S \) to report a signal \( \sigma \neq \emptyset \) to \( P^5 \). The other restriction, in which inefficient \( F \) chooses self reporting, arises out of equilibrium: \( P \) formulates the problem in such a way that only low types want to report \( \kappa = 1 \). The incentive constraint that ensures this will be shown later in the section. Let \( \delta \in [0, 1] \) be the probability that such efficient \( F \) accepts the fast contract.

\(^{5}\)This assumption would also hold in equilibrium: the government would prefer to ignore the signal \( \sigma \) if \( \kappa = 1 \).
The constraints

**Grand contract stage.** Our model is a slight modification of LT, since all of the constraints in LT hold true in our model. The first set of constraints then apply in the grand contracting stage of the game, where F has chosen to reject the fast contract. As before, the participation constraints (3.2) and (3.3) are

\[
PC : t_i - \psi(e_i) \geq 0, \forall i \\
PC_s : s_i \geq 0, \forall i.
\]

The efficient F must still be discouraged from mimicking an inefficient F by meeting the incentive constraint (3.5),

\[
IC_f : t_3 - \psi(e_3) \geq t_4 - \psi(e_4 - \Delta \beta).
\]

And finally, S must be discouraged from entering into side agreements with F (constraint 3.8),

\[
s_1 \geq s_3 + \frac{1}{1+\lambda} b,
\]

where the bribe is at most the information rent of F in state 3.

**The fast contract.** The second set of constraints apply in the fast contract stage of the game. In order for the efficient F to choose the fast contract, the rents gained should at least equal those under the grand contract in expectation:

\[
IC_0 : t_0 - \psi(e_0) \geq \xi (t_1 - \psi(e_1)) + (1 - \xi)(t_3 - \psi(e_3)). \tag{3.11}
\]

Note that what we call \( IC_0 \) could be considered as a participation constraint. As long as participation constraints (3.2) and (3.3) bind, the outside option for F with respect to the fast contract is not to pull out of the market entirely but to pursue a strategy of hiding its own type and move into the grand-contract stage. Furthermore, in order for a fraction \( 1 - \delta \) of low type
F to choose not to self report, it must be the case that \( IC_0 \) binds with equality. The incentive constraint for the high type F as indicated here, ensures that high type F never report \( \kappa = 1 \), and is always met:

\[
IC_0 : t_0 - \psi(e_0) \leq \xi(t_2 - \psi(e_2)) + (1 - \xi)(t_4 - \psi(e_4)).
\]

### The maximization program

We are now ready to introduce the welfare function:

\[
\max_{\{e_i\}_{i=1, \ldots, 4}} W = G - \delta \nu \{ t_0 + s_0 + \beta - e_0 \} - \xi \nu (1 - \delta) \{ t_1 + s_1 + \beta - e_1 \} \\
- (1 - \nu) \xi \{ t_2 + s_2 + \beta - e_2 \} - \nu (1 - \xi) (1 - \delta) \{ t_3 + s_3 + \beta - e_3 \} \\
- (1 - \nu) (1 - \xi) \{ t_4 + s_4 + \beta - e_4 \},
\]

subject to constraints (3.2), (3.3), (3.5), (3.8) and (3.11).

The solution to the problem involves determining which participation constraints and incentive constraints bind. First off, participation constraints (3.2) for states \( i = 1, 2, 4 \) bind with equality: not only reducing \( t_i \) increases welfare, but it also makes it easier to facilitate \( IC_0 \). For similar reasons, the solution involves \( s_0 = s_2 = s_3 = s_4 = 0 \). Furthermore, \( IC_f \) binds with equality. This implies that rents to the efficient F in state 3 of the world are

\[
IC_f : t_3 - \psi(e_3) = \Phi(e_4),
\]

and the incentive pay paid to S in state 1 of the world is

\[
s_1 = \frac{1}{1 + \lambda} \Phi(e_4).
\]

Thus, the rents for F and S remain unchanged under the grand contract from the \( CP \) equilibrium. What is new now is that \( IC_0 \) also binds with equality:

\[
t_0 - \psi(e_0) = (1 - \xi) \Phi(e_4). \tag{3.13}
\]
Welfare is now reduced to a function of $e_i$ only:

\[
\max_{\{e_i\}, i=1,...,4} W = G - \delta \nu \left\{ \psi(e_0) + (1 - \xi)\Phi(e_4) + \beta - e_0 \right\} - \xi \nu (1 - \delta) \left\{ \psi(e_1) + \frac{1}{1+\lambda} \Phi(e_4) + \beta - e_1 \right\} - (1 - \nu) \xi \left\{ \psi(e_2) + \beta - e_2 \right\} - \nu (1 - \xi) (1 - \delta) \left\{ \psi(e_3) + \beta - e_3 + \Phi(e_4) \right\} - (1 - \nu)(1 - \xi) \left\{ \psi(e_4) + \beta - e_4 \right\}.
\]

We are now able to solve for the optimal contract. First order conditions are:

\[
\psi'(e_i) = 1, \ i = 0, 1, 2, 3
\]

\[
\nu (1 - \xi) \Phi'(e_4) + \nu \frac{(1 - \delta)\xi}{1+\lambda} \Phi'(e_4) + (1 - \nu)(1 - \xi) \psi'(e_4) = (1 - \nu)(1 - \xi). \quad (3.14)
\]

We are now able to derive the main point of the paper.

**Proposition 3.1** With Conditional Supervision the collusion-free supervision outcome is feasible.

**Proof.** When $\delta = 1$ the FOC (3.14) becomes

\[
\nu (1 - \xi) \Phi'(e_4) + (1 - \nu)(1 - \xi) \psi'(e_4) = (1 - \nu)(1 - \xi).
\]

Note that this FOC is perfectly identical to the FOC (3.6) in the $CF$ program: any additional cost associated with corruption is eliminated, and the optimal level of effort is $e_4^{CF}$. ■

All this implies that self-reporting makes it possible for $P$ to reach the second-best outcome.

**Proposition 3.2** With Conditional Supervision the collusion-free outcome is the best outcome achievable.

The proof is offered in Appendix 1.
With this proposition, we know that $\delta^* = 1$: it is best for $P$ to induce the efficient $F$ to sign the fast contract with probability 1. Note that this validates the example we saw in the previous section.

To conclude this section, we write again the Conditional Supervision outcome, which corresponds to the Collusion Free welfare:

$$W_{CS}^* = S - \sum_{i=1}^{3} p_i \{ \psi(e_i^*) + \beta_i - e_i^* \} - p_4 \{ \psi(e_{4F}^*) + \beta_4 - e_{4F}^* \} - \nu(1 - \xi)\Phi(e_{4F}^*).$$

We summarize the findings in the following proposition:

**Proposition 3.3** By adopting the conditional supervision mechanism with fast contract and grand contract:

1. All efficient $F$ choose the fast contract.
2. All inefficient $F$ decline the fast contract and accept the grand contract.
3. The agency receives zero rents in all realized states.
4. Principal fully eliminates the costs of collusion, and $W_{CS} = W_{CF} > W_{CP}$.
5. The inefficient $F$ (in state 4) is given a high-powered incentive scheme with respect to the case in which conditional supervision is not allowed ($e_{4F}^* > e_{4C}^*$).

### 3.4 Sequential Contracting—SC

A problem with the self reporting equilibrium shown here is that it is not interim-efficient. The introduction of a fast contract allows the government to naturally and costlessly separate efficient $F$ from inefficient $F$: once $F$ has chosen to reject the fast contract, it is obvious to everyone that this $F$ must be of the inefficient kind, and therefore the obvious contract that should be offered in the second stage is the first best contract. Therefore, the equilibrium grand contract is suboptimal, and $P$ will be tempted to renegotiate the contract. The ability of $P$ to renegotiate (or, conversely, the inability to commit to the original grand contract) causes the
outcome visited above to unravel: the efficient F suspects that renegotiation may happen, and thus require higher rents.

This problem arises if the contracting is sequential in nature. To see how sequential contracting may modify the model above, suppose that P offers two alternatives to F: a fast contract \((C_0, t_0, s_0)\), which is independent of the signal, or the possibility to negotiate a grand contract. Since the negotiation of a grand contract is conditional on a rejection by F to the fast contract, P will make use of the additional information to update its own beliefs regarding the type of F\(^6\).

To see how our result is modified, suppose that there is a proportion \(\delta\) of low type that choose to self report, while the remaining \(1 - \delta\) choose not to\(^7\). With this restriction in mind, note that the grand contract is changed because the probabilities of each state of the world are now changed. Denote by \(p'_i\) the ex-post distribution of types, after the fast contract was rejected. For a given \(\delta\), we have the following ex-post probabilities:

\[
\begin{align*}
    p'_1 &= \frac{\nu \xi (1 - \delta)}{(1 - \delta) \nu + (1 - \nu)} < p_1, \\
    p'_2 &= \frac{(1 - \nu) \xi}{(1 - \delta) \nu + (1 - \nu)} > p_2, \\
    p'_3 &= \frac{\nu (1 - \xi) (1 - \delta)}{(1 - \delta) \nu + (1 - \nu)} < p_3, \\
    p'_4 &= \frac{(1 - \nu) (1 - \xi)}{(1 - \delta) \nu + (1 - \nu)} > p_4.
\end{align*}
\]

These are the probabilities of each state of the world occurring conditional on F not self reporting its behavior.

The interim welfare function is the same as in LT, with the new set of probabilities \(p'_i\):

\[
\max_{\{e_i, t_i, s_i\}_{i=1,...,4}} W_{\text{interim}}(\delta) = G - \sum_{i=1}^{4} (t_i + C_i + s_i)p'_i.
\]

Subject to the same restrictions as in the LT model. Without going to all the steps, we

\(^6\)Throughout this paper, we maintain the assumption that P can commit to the grand contract once that has been drawn. Our results would be a little weaker if this assumption is removed, but the advantage of our model remains. See LM 1998 for an explanation of the renegotiation-proof mechanism in the standard model.

\(^7\)This can happen, for example, if efficient firms are heterogeneous in their risk aversion.
simply state the collusion-proof interim welfare function:

\[
\max_{\{e_i\}_{i=1,...,4}} W_{\text{interim}}(\delta) = G - p_1 \left\{ \psi(e_1) + \frac{1}{1 + \lambda} \Phi(e_4) + \beta - e_1 \right\} - p_2 \left\{ \psi(e_2) + \beta - e_2 \right\} \\
- p_3' \{ \psi(e_3) + \beta - e_3 + \Phi(e_4) \} - p_4' \{ \psi(e_4) + \beta - e_4 \}.
\]

The first order conditions for \(e_i\), \(i = 0, 1, 2, 3\) are, as usual, the optimal level \(e^*\). The level of effort in state 4, \(e_4(\delta)\), is now determined by the following:

\[
\nu \omega_\delta \Phi'(e_4(\delta)) + (1 - \nu)(1 - \xi) \psi'(e_4(\delta)) = (1 - \nu)(1 - \xi) + \nu \delta (1 - \xi) \Phi'(e_4(\delta)), \tag{3.15}
\]

where \(\omega_\delta = \left(1 - \xi + \frac{(1-\delta)\xi}{1+\lambda}\right)\).

Condition (3.15) defines the minimum effort \(e_4(\delta)\) to the inefficient F: any contract chosen by P in the first stage needs to involve \(e_4 \geq e_4(\delta)\) in order for this contract to be renegotiation proof. Denote by \(e^{SC}_4(\delta)\) the effort that meets this condition, where SC stands for "sequential contracting". It is easy to notice that \(e_4(\delta)\) is an increasing function of \(\delta\). Moreover if \(\delta = 0\) condition (3.15) turns out to be identical to (3.9): in this case \(e^{SC}_4(0) = e^{CP}_4\). This establishes that, at its worst, sequential contracting is no more distortionary than the collusion-proof contract. On the other hand, for any \(\delta > 0\) we have \(e^{SC}_4(\delta) > e^{CP}_4\). If \(\delta = 1\) condition (3.15) reduces to \(\psi'(e_4) = 1\), i.e. \(e^{SC}_4 = e^*_4\).

We now consider the global problem in which the contract is chosen ex-ante and is renegotiation proof. As before, the problem involves maximizing the welfare function (3.12) subject to all prior constraints (3.2), (3.3), (3.5), (3.8) and (3.11), plus the renegotiation constraint (3.15). Notice that (3.15) must also be binding in order to ensure the optimality of \(e_4(\delta)\). The resultant welfare equation is a function of \(\delta\).

\[
\max_{\delta \in [0,1]} W_{SC}(\delta) = G - \nu \left\{ \psi(e^*) + \beta - e^* \right\} - (1 - \nu)\xi \left\{ \psi(e^*) + \beta - e^* \right\} \\
-(1 - \nu)(1 - \xi) \{ \psi(e_4(\delta)) + \beta - e_4(\delta) \} - \nu \omega_\delta \Phi(e_4(\delta)),
\]
where \( e_4(\delta) \) solves (3.15). Given our assumptions on the disutility function \( \psi(\cdot) \), the welfare function is a concave function in \( \delta \). There is a level of \( \delta \in [0, 1] \) that would uniquely maximize this function. Next, we derive the condition for this \( \delta^* \).

By taking first order conditions with respect to \( \delta \), we get:

\[
\frac{\xi \nu}{1 + \lambda} \Phi(e_4(\delta)) + (1 - \nu)(1 - \xi) \frac{\partial e_4(\delta)}{\partial \delta} = \nu \omega \delta \Phi'(e_4(\delta)) \frac{\partial e_4(\delta)}{\partial \delta} + (1 - \nu)(1 - \xi) \psi(e_4(\delta)) \frac{\partial e_4(\delta)}{\partial \delta}.
\]

(3.16)

We can then substitute (3.15) into the RHS of this Euler equation, rearrange, and (3.16) becomes

\[
\frac{\xi \nu}{1 + \lambda} \Phi(e_4(\delta)) - \nu \delta \Phi'(e_4(\delta)) \frac{\partial e_4(\delta)}{\partial \delta} = 0.
\]

(3.17)

Condition (3.17) determines the optimal \( \delta^* \). Note that it is possible that the LHS is equal to 0, depending on the shape of the function \( \psi(e) \) and of \( \lambda \). When that is the case, we have an interior solution. It is also possible that for all values of \( \delta \), the LHS remains greater than zero. In that case, we have that \( \delta^* = 1 \), a condition that we explore below.

### 3.4.1 What happens when the efficient F chooses the fast contract with probability 1?

Suppose that \( \delta^* = 1 \). Then, clearly, (3.15) reduces to

\[
\psi'(e_4) = 1,
\]

and the optimal level of effort is always maximal: \( e_4 = e^* \). Social welfare then becomes:

\[
W_{SC}^*(\delta^* = 1) = G - \nu \left\{ \psi(e^*) + \beta - e^* \right\} - (1 - \nu) \left\{ \psi(e^*) + \bar{\beta} - e^* \right\} - \nu(1 - \xi) \Phi(e^*).
\]

This is the same expected utility for P as in \( CF \), provided that \( e_4^{CF} \) is replaced with the sub-optimal level \( e^* \). \( SC \) is, therefore, less advantageous than \( CF \) and, in some cases, less
advantageous than \( CP \).

### 3.4.2 On the attainability of welfare-improving Sequential Contracting

**The role of deadweight losses.** We have seen in the prior sub-section that, when sequential contracting leads to full separation, the fast contract leads to welfare inferior allocation. When the fast contract is never chosen, then sequential contracting reduces to \( CP \). Thus, in order for \( SC \) to be welfare-improving, it is necessary that \( \delta \in (0, 1) \). Even at its (interior) optimum, it is not necessary true that \( SC \) dominates \( CP \), since selecting the first over the second type of contract leads to a trade-off: a reduction of rents to \( S \) in exchange of an increase of rents to \( F \) (but also higher efficiency). \( SC \) may still be the better policy if deadweight losses from collusion are small (\( \lambda \) is small). When transfers from \( F \) to \( S \) are efficient, supervisory payouts are a large source of allocative inefficiency. Selecting a sequential contract eliminates this source of inefficiency. On the other hand, under a system where side contracting is cumbersome – say, due to high transparency requirements for \( F \) and the supervising agency, or because of aggressive auditing practices implemented by \( P \) – payouts to \( S \) are small enough to make sequential contracting a ‘fourth-best’ policy. This result enhances the applicability of Sequential Contracting in environments where the fight against corruption and collusion is difficult, such as in many developing countries.

### 3.5 Conclusions

This paper analyzes a simple modification of Laffont and Tirole’s (1993) standard mechanism in hierarchical structures, where an agent (a firm \( F \)) and his supervisor (\( S \)) can collude at the expense of the principal (\( P \)). By letting \( F \) choose between a regime free of supervision over a regime of supervision, our model yields results that are superior to the standard model. In fact, our mechanism allows \( P \) to eliminate all the costs associated with the threat of collusion.

These results must be mitigated in several ways. First, when \( P \) designs contracts sequentially or conditionally on the choices made by \( F \), our mechanism may still bring about better outcomes than centralized supervision alone, but that depends on the parameters of the model. Second, our mechanism devolves into the standard collusion proof mechanism if \( S \) and \( F \) can collude
before the fast contract is accepted or refused. This possibility arises if F can bribe S ex-ante, in exchange for an uninformative ex-post report to P. In order to do so S must be capable to commit to an outcome that is ex-post inferior: this may be reasonable if S and F interact repeatedly. In this case P may still be able to avoid the creation of the cartel by using additional strategies. For instance, job rotation can be used to insure that in the following period S will be moved to a different job with a different contractor. Alternatively, P could avoid to disclose F’s identity in the fast contracting stage: this precaution makes it difficult for S to collude in this stage of the game since she faces a potentially vast population of eligible F. Finally, P may decide to hire S only in the second stage of the game while in the first period no supervisor is in charge: for this solution to be effective S should not be able to anticipate that she will be hired for the job in the second stage of the game.
Bibliography


Appendix: Proof of proposition 3.2

We want to show that the self reporting solution necessarily involves \( \delta = 1 \). To show this, we will show that \( \delta = 1 \) maximizes the welfare function. First, we rearrange the FOC for \( e_4 \) (3.14):

\[
\nu \omega_\delta \Phi'(e_4(\delta)) + (1 - \nu)(1 - \xi) \psi(e_4(\delta)) = (1 - \nu)(1 - \xi) \\
where \quad \omega_\delta = \left( 1 - \xi + \frac{(1 - \delta)\xi}{1 + \lambda} \right).
\]

This first order condition defines a level of effort \( e_4(\delta) \) that is a function of the probability that a F self reports it of being a low type. It is straightforward to show that, as long as \( \psi''(.) > 0 \), the distortion of the inefficient type is reduced as \( \delta \) increases: \( \frac{\partial e_4(\delta)}{\partial \delta} > 0 \).

Second, we plug the optimized levels of \( e_i \) in the welfare function, to get a function that depends on \( \delta \) only:

\[
\max_{\delta \in [0,1]} W_{sr}(\delta) = G - \nu \{ \psi(e^*) + \beta - e^* \} - (1 - \nu)\xi \{ \psi(e^*) + \bar{\beta} - e^* \} \\
- (1 - \nu)(1 - \xi)\{ \psi(e_4(\delta)) + \beta - e_4(\delta) \} - \nu \omega_\delta \Phi(e_4(\delta)).
\]

The first order conditions are:

\[
\frac{\xi \nu}{1 + \lambda} \Phi(e_4(\delta)) + (1 - \nu)(1 - \xi) \left[ 1 - \psi'(e_4(\delta)) \right] \frac{\partial e_4(\delta)}{\partial \delta} = \nu \omega_\delta \Phi'(e_4(\delta)) \frac{\partial e_4(\delta)}{\partial \delta}.
\]

We can rewrite these as:

\[
\frac{\xi \nu}{1 + \lambda} \Phi(e_4(\delta)) + (1 - \nu)(1 - \xi) \frac{\partial e_4(\delta)}{\partial \delta} = \nu \omega_\delta \Phi'(e_4(\delta)) \frac{\partial e_4(\delta)}{\partial \delta} + (1 - \nu)(1 - \xi)\psi(e_4(\delta)) \frac{\partial e_4(\delta)}{\partial \delta}.
\]

Since these first order conditions hold at the optimal level of effort \( e_4 \), it must be the case that (3.14) binds at the optimum. We can then substitute the RHS and (3.18) becomes
\[
\frac{\xi \nu}{1 + \lambda} \Phi(e_4(\delta)) + (1 - \nu)(1 - \xi) \frac{\partial e_4(\delta)}{\partial \delta} = (1 - \nu)(1 - \xi) \frac{\partial e_4(\delta)}{\partial \delta}.
\]

or \[
\frac{\xi \nu}{1 + \lambda} \Phi(e_4(\delta)) = 0
\]

Which cannot be true, since \( \frac{\xi \nu}{1 + \lambda} \Phi(e_4(\delta)) > 0 \) for any level of \( \delta \in [0, 1] \). Hence, we have a corner solution and \( \delta \) is maximal.
Chapter 4

Collusion and Partial Delegation

4.1 Introduction

Within economic organizations, third party supervision is commonly observed: owners of a firm usually delegate to top managers the responsibility for supervising production, stockholders rely on auditors to acquire information about management conduct, managers ask employees to report on the performance of coworkers and governments make use of agencies to regulate firm with unknown cost.

In some instances, the need for such supervisory activities steams from the information asymmetry between the residual claimant of a productive activity (the principal) and who actually bears the cost of production (the agent). Under these circumstances, one alleged advantage of using an informed third party is to allow the principal to acquire information that would otherwise not be available, reducing his informational disadvantage with respect to the agent. Nonetheless, the introduction of this supervisor creates a potential for collusive behavior against the principal’s will. Whenever the coalition supervisor - agent is so efficient as to act like a single player, supervisory activity is hardly valuable for the principal. On the contrary, if the coalition performance suffers of some inefficiencies, the principal could potentially benefit from supervisor’s existence by adopting appropriate organizational responses. Contracting with both the supervisor and the agent through a grand contract constitutes the most general orga-

\(^1\) We refer to the supervisor and the agent respectively as "she" and "he".
nizational design for the principal, which we denote by centralized contracting. Special cases of this design would be respectively the principal contracting only with the supervisor, delegating authority to her over contracting with the agent, and the principal contracting with the agent directly, ignoring the presence of the supervisor. The former organizational response is denoted by decentralized contracting, while the latter by no-supervision contracting.

The contribution of this paper is to show that if collusion between the agent and the supervisor takes place after the acceptance of the grand-contract by both parties, the principal can eliminate all the costs associated with collusion by offering a centralized mechanisms where the agent can report his private information and the supervisor sends some "non-type" messages\(^2\); in the baseline model the "non-type" messages corresponds to supervisor’s acceptance or refusal of the grand-contract\(^3\). Indeed, if the principal holds non-passive beliefs following the supervisor’s refusal of accepting the grand-contract, the latter can be used to update the principal’s beliefs on the agent’s type. This centralized organization implies that rejection of the grand-mechanism is an equilibrium behavior for the supervisor in some states of the world\(^4\): the fact that the supervisor plays no role in the actual production process ensures that her presence is not indispensable for the principal’s purposes.

A couple of aspects regarding this "equilibrium rejection" grand-mechanism are worth noticing. First, the principal’s payoff is invariant in the accuracy of supervision information: soft supervisory information would be useful even when collusion takes place under symmetric information. Second, the agent’s bargaining power in side-contracting is not relevant either.

In order to test the robustness of the mechanism, its implementation is analyzed in two different frameworks. The first one considers an information structure for the supervisor that can be represented as a connected partition of the agent’s type space: Celik (2008) analyses

\(^2\)This is reminescent of Mookherjee and Reichelstein (1990) result, which states that if the mechanism designer seeks to ensure that all noncooperative equilibria of a mechanism achieve a given performance standard, then attention can be focused on a class of augmented revelation mechanisms. In these mechanisms agents can either report their private information, or send some auxiliary "non-type" message.

\(^3\)When a generic number of agent’s types are considered, the message space for \(S\) must contain a number of "non-type" messages that depends on her information set.

\(^4\)This aspect comports some difficulties in categorizing the principal’s contract: on the one hand, it clearly resembles a centralized contract (i.e. the principal contracts ex-ante with both the supervisor and the agent). On the other hand, it ends up realizing as a no-supervision contract in case the supervisor refuses the principal’s grand-mechanism.
centralized and decentralized contracting in this very setting. Whenever the optimal contract for the agent in the absence of the supervisor is strictly monotone, Celik (2008) concludes that the principal can increase his payoff with the introduction of the supervisor, as long as he offers the appropriate centralized contract.

The second information environment entails two possible agent types and the supervisor observing an informative signal which also takes two possible values. Faure-Grimaud, Laffont and Martimort (2003) (FLM, hereafter) analyses centralized and decentralized contracting in the same framework, concluding that decentralization is always equivalent to centralization.

Unlike the outcome implemented by the "equilibrium rejection" grand-mechanism presented here, the optimal outcome identified by both Celik (2008) and FLM (2003) fails to achieve the same expected payoff as the optimal collusion-free outcome. Moreover, in their organizational designs soft supervisory information is useless when collusion takes place under symmetric information.

These differences are driven by the restriction that both Celik (2008) and FLM (2003) impose on their grand-mechanisms under centralized contracting. That is, they assume that the principal offers a grand-contract to the supervisor and the agent; if any of them refuses, the game ends and no production takes place. Unlike them, the present work assumes that the only necessary condition for shutting down production is the agent’s refusal of the grand-contract: given that the supervisor is not essential for production, this is quite a natural assumption.

A similar point to the one presented here is made by Burlando and Motta (2008) that analyze a modification of the standard mechanism in hierarchical structures, where an agent and his supervisor can collude at the expense of the principal. They consider an environment characterized by supervision with hard information, imperfect inspection technology and limited liability requirement for the supervisor. By letting the agent choose between a regime free of supervision and a regime of supervision, the principal can eliminate the costs of collusion. This outcome is achieved on condition that the principal designs all the contracts before the agent makes a selection. When the principal designs contracts sequentially or conditionally on the
choices made by the agent, problem of commitment arises, reducing the principal’s expected payoff.

The way adopted for modelling collusion is due to the early literature’s standard set by Laffont and Martimort (1997, 2000) that consider collusion as a side-contract between the agent and the supervisor, which is unobserved by the principal and subject to asymmetric information within the coalition: furthermore, all enforceability constraints are ignored, by assuming self-enforcing relationships among the agent and the supervisor.

Another related article is the one by Mookherjee - Tsumagari (2004), which considers two productive agents and assumes that the supervisor is perfectly informed. They adopt this framework to explore the possibility that collusion may rationalize delegation to intermediaries uninvolved in production. They conclude that in case of supplier complementarity, the principal is strictly better off hiring the supervisor, and then can delegate to him at no cost. At the contrary, in the substitutes case, it is strictly better for the principal not to hire a supervisor and contract directly with both suppliers. Unlike the setting presented here, they allow agents to collude prior to making their decision to participate in the mechanism. Mookherjee (2006) provides an excellent survey of the active strand of the literature evaluating delegation when agents collude and authority delegates to expert intermediaries or managers who play no role in the actual production process.

Section 4.2 presents the model and discuss the organizational form consider in this paper. Section 4.3 and Section 4.4 describe the sets of collusion proof grand-mechanisms under the two different information environments. Section 4.5 concludes. All proof are given in Appendix.

4.2 The Model

The setting involves a productive agent (A) who bears the cost of production. A's utility function is given by

---

7 Other related articles are the following: Che and Kim (2006a) show that if collusion takes place between a subset of the agents, the second best payoff is implementable. Baliga and Sjostrom (1998), and Laffont and Martimort (1998) show that under certain conditions delegation is the optimal organizational response to collusion. Pavlov (2006) and Che and Kim (2006b) consider auctions where bidders collude prior to participating. Quesada (2004) studies collusion initiated by an informed party under asymmetric information.
\( t - \theta q, \)

where \( t \) denotes the transfer he receives from the principal \( (P) \), \( q \) is the output level and \( \theta \) represents the unitary cost of production, which takes three possible values from the set \( \Theta = \{\theta_1, \theta_2, \theta_3\} \), where \( 0 \leq \theta_1 \leq \theta_2 \leq \theta_3 \). The distribution of the cost, \( f(\theta) \), is common knowledge while \( A \) knows the realization of \( \theta \). The supervisor’s \((S)\) information structure is the partition \( \{\{\theta_1, \theta_2\}, \theta_3\} \) of \( A \)’s type space \( \theta \). It follows that \( S \) is able to tell whether \( A \) is the least efficient type \( \theta_3 \) or not, but she can’t distinguish types \( \theta_2 \) and \( \theta_1 \). \( S \)’s salary is \( s \), which represents her monetary transfer from \( P \), whose payoff for a given output \( q \), transfer level \( t \) and wage \( s \) is

\[
W(q) - t - s,
\]

where \( W'(q) > 0, W''(q) < 0 \), for all \( q \), and \( \lim_{q \to 0} W'(0) = \infty, \lim_{q \to \infty} W'(q) = 0 \). These conditions ensure positive production regardless of \( A \)’s cost type \( \theta \).

### 4.2.1 Organizational forms

\( P \) can choose different organizational forms in order to contract with the players involved in the game: below, the characteristics of the different organizational structures are briefly defined.

**Centralized contracting.** \( P \) directly contracts and communicates with both \( A \) and \( S \). A grand-mechanism consists in a triple \( GC = \{q(m_s, m_a), t(m_s, m_a), s(m_s, m_a)\} \), which defines the outcome and the monetary transfer respectively for \( A \) and \( S \) as a function of the \( S \)’s and \( A \)’s messages, which are respectively denoted as \( m_s \) and \( m_a \). In turn, \( m_s \) and \( m_a \) belong respectively to two message spaces \( M_s \) and \( M_a \). This organization structure may be subject to collusion between \( S \) and \( A \). More specifically \( S \) has all the bargain power and makes a take-it-or-leave-it offer to the agent. The side-contract is a pair \( SC_1 = \{\widehat{c}(.), b(.)\} \) where \( \widehat{c}(.) \) is a collective manipulation of the messages \( (m_s, m_a) \) sent to \( P \), while \( b(.) \) is the side-transfer from \( A \) to \( S \). As standard in this literature on collusion, this side-contract is assumed to be enforceable\(^8\).

\(^8\)Relaxation of the enforceability assumption is considered by Martimort (1999), Abdulkadiroglu and Chung (2003), and Khalil and Lawarree (2006).
Decentralized contracting. $P$ contracts only with $S$ who is free to subcontract with $A$. It follows that $A$ neither receives any transfer from $P$, i.e. $t(m_s,m_a) = 0$, nor communicates directly with $P$, i.e. $M_a = \emptyset$. The only player who communicates with $P$ is $S$ who does so through her message space $M_s$. A sub-contract links now $S$ and $A$: this sub-contract is denoted as $SC_2 = \{\tilde{c}(.),b(.)\}$ where $\tilde{c}(.)$ is the manipulation of $S$’s report $(m_s)$ to $P$, while $b(.)$ is the side-transfer from from $A$ to $S$.

No-supervision contracting. Under this organization form, $P$ contracts only with $A$. It follows that $S$ neither receives any transfer from $P$, i.e. $s(m_s,m_a) = 0$, nor communicates directly with $P$, i.e. $M_s = \emptyset$. $A$ can communicate directly with $P$ through his message space $M_a$. Per se, this organization structure is not subject to coalition formation between $S$ and $A$.

The organizational response considered in this paper could be broadly defined as centralized contracting: that is, $P$ contracts with both $S$ and $A$ by offering a centralized mechanism where $A$ can report a "type" message and $S$ sends some "non-type" messages. Anyway, this centralized organization implies that rejection of the grand-mechanism is an equilibrium behavior for the supervisor in some states of the world: when these states are realized, the grand-mechanism turns into a no-supervision contract (see Figure 4-1).

4.2.2 Timing of the game

The timing of the game under centralized contracting is as follows:

- $A$ learns $\theta$. $S$ learns the partition $\{\{\theta_1, \theta_2\}, \theta_3\}$.
- $P$ offers a grand-mechanism to $S$ and $A$.
- $S$ and $A$ accept or refuse the grand-mechanism. If $A$ refuses, the game ends. If $S$ refuses and $A$ accepts production and transfers take place as specified in the grand-contract. If both $S$ and $A$ accept the grand-mechanism $S$ may offer a take-it-or-leave-it side-contract to $A$ who accepts or refuses this contract. If $A$ refuses, the grand-mechanism is played non-cooperatively by $S$ and $A$. If $A$ accepts, the grand-mechanism is played according to the side-contract.
- Production and transfers take place.
4.2.3 Benchmark: Collusion free Supervision

Consider the benchmark case where collusion between $A$ and $S$ is not allowed. For a detailed analysis of this benchmark see Celik (2008). Here, a simple sketch is provided. Denote by $V_i$ the coalition information rent, which represents the sum of the utility levels for $A$ and $S$, whenever the agent’s type is $\theta_i$; $U_i$ represents the utility level for $A$. Under collusion free supervision $P$ can adopt centralized contracting and set $V_i = U_i$ for all $i \in \{1, 2, 3\}$; $S$ accepts the contract obtaining a surplus equal to zero regardless of her report to $P$. Moreover, $S$ reports the true partition cell she observes, allowing $P$ to extract her information for free. Accordingly $P$ can implement an output profile $\{q_i\}_{i \in \{1,2,3\}}$ with the utility levels

\[
V^cf_3 = U^cf_3 = 0, \\
V^cf_2 = U^cf_2 = 0, \\
V^cf_1 = U^cf_1 = (\theta_2 - \theta_1)q_2. \\
\tag{4.1}
\]

where the following condition must hold $q_1 \geq q_2$. 

---

**Figure 4-1: Organizational Responses**

- **Centralized Contracting (Standard)**
- **No Supervision Contracting**
- **Decentralized Contracting (with "double-tweet" message)**
As standard in this kind of design problem, the information rent to be paid to type $\theta_3$ is zero. Moreover, the presence of the supervisor provides $P$ with an additional advantage: that is, $A$ is no longer able to mimic a type belonging to a different partition cell. It follows that neither type $\theta_2$ nor type $\theta_1$ can misreport their types as $\theta_3$. Given that type $\theta_2$ is not willing to mimic type $\theta_1$, $P$ is able to leave zero rent to type $\theta_2$ as well as to type $\theta_3$. Yet, it is still profitable for type $\theta_1$ to imitate type $\theta_2$: in order to avoid this misrepresentation $P$ must forgo a positive rent to the most efficient type.

The standard treatment of this problem would suggest that an output profile $\{q_i\}_{i \in \{1,2,3\}}$ is implementable through a contact if and only if it is weakly decreasing, i.e. $q_1 \geq q_2 \geq q_3$. However, as long as $S$ can distinguish type $\theta_3$ from the other types, implementability requires the output profile to be monotonic only with respect to types $\theta_1$ and $\theta_2$ output levels, i.e. $q_1 \geq q_2$. Generally speaking, when supervisory activity is in place, there is no monotonicity requirement regarding two output levels in separate partition cells. Under collusion-free supervision, the optimal set of output levels is determined as

$$\left\{q_i^{\text{cf}}\right\}_{i \in \{1,2,3\}} \in \arg \max_{q_i} \sum_{i=1}^{3} f(\theta_i) \left[ W(q_i) - q_i - V_i^{\text{cf}} \right],$$

\[\text{s.t. } q_1 \geq q_2.\]

As soon as the possibility of collusion is introduced, the outcome presented above is no longer implementable. To see this point, consider the following profitable collective manipulation: type $\theta_2$ can increase her expected payoff by misreporting as type $\theta_3$. In order to do so he needs $S$’s cooperation. Notice that $S$ is indifferent between playing along with this manipulation or not: therefore the collective gain from this misreport is strictly positive. $S$ can simply offer a take-it-or-leave-it side-contract that asks for a bribe from $A$ for misreporting him as type $\theta_3$.

### 4.3 Grand-mechanism under collusion

This section proposes a grand-mechanism $GC^{\text{cf}}$, which implements the collusion free outcome also in the case where collusion is allowed. The design structure is as follows: after the grand-contract is proposed by $P$, $S$ and $A$ simultaneously make their acceptance decisions, which
are observed by all. The grand-contract itself is contingent on S’s acceptance decision. The smallest message spaces for S and A compatible with the implementation of this mechanism are respectively $M_s \in \{a_1, a_2\}$ and $M_a \in \{\theta_1, \theta_2, \theta_3\}$, where $a_1$ stands for "accept the grand contract" and $a_2$ indicates "refuse the grand contract". For the sake of exposition, denote by $s_{ij}$ (respectively $t_{ij}$, $q_{ij}$, $V_{ij}$ and $U_{ij}$) S’s wage (respectively the A’s transfer, the output target, the coalition information rent and A’s utility) when S reports $a_i$ and A reports that he has type $\theta_j$. Consider the following grand-contract:

$$
t_{13} = 0 \quad s_{13} = -(\theta_3 - \theta_2)q_{12}
$$
$$
t_{12} = \theta_2q_{12} \quad s_{12} = 0
$$
$$
t_{11} = (\theta_2 - \theta_1)q_{12} + \theta_1q_{11} \quad s_{11} = 0,
$$

and

$$
t_{23} = \theta_3q_{23} \quad q_{13} = 0 \quad q_{23} = q_{3f}^c
$$
$$
t_{22} = 0 \quad q_{12} = q_{2f}^c \quad q_{22} = 0
$$
$$
t_{21} = 0 \quad q_{11} = q_{1f}^c \quad q_{21} = 0.
$$

The following proposition encapsulates the first result of this paper.

**Proposition 4.1** $\text{GC}^{sb}$ allows $P$ to achieve the same expected payoff as the optimal collusion-free outcome: it follows that all the costs associated with collusion are fully eliminated.

The complete proof is relegated in the Appendix. Here, a simple sketch of the intuition is offered. To begin with, notice that the grand-contract is collusion-proof. The rationale of this result is straightforward: given that it is assumed that S offers the side-contract to A after the

---

9In principle, the message space for A should also contain the elements $\{a_1, a_2\}$; anyway, if A refuses the grand-contract the game ends with zero production and no monetary transfers.
acceptance of the grand contract by both parties, the collusion proofness of the mechanism must be tested only in case $S$ accepts the grand-contract, i.e. whenever she reports $a_1$ to $P$. Suppose that $S$ observes the partition $\{\theta_3\}$ and accepts the grand-contract. In this case $S$ would like to misreport type $\theta_3$ in the attempt to avoid the negative transfer $s_{13}$. A simple inspection reveals that $S$ is indifferent between paying $s_{13}$ to $P$ or offering a bribe $(\theta_3 - \theta_2)q_{12}$ to convince type $\theta_3$ to misreports his type as $\theta_2$. Also, $S$ strictly prefers paying $s_{13}$ than offering a bribe $(\theta_2 - \theta_1)q_{12} - (\theta_3 - \theta_1)q_{11}$ to induce type $\theta_3$ to mimic type $\theta_1$. On the other hand, suppose that $S$ observes the partition $\{\theta_2, \theta_1\}$ and decides to accept the grand-contract. In this case, there are no profitable manipulations available for the coalition because neither $S$ nor types $\theta_2$ and $\theta_1$ have a stake in misreporting. Accordingly, if $A$ and $S$ accept the grand-contract they shall respond to it in a non-cooperative fashion.$^{10}$

Moreover the grand-contract is also incentive compatible. Indeed, both the participation and the incentive compatibility constraints for all types hold when $S$ accepts the grand contract,

$$\begin{align*}
\text{IR } (\theta_j | a_1) \quad & U_{1j} \geq 0 \quad \text{for all } j = \{1, 2, 3\}, \quad (4.2) \\
\text{IC } (\theta_j | a_1) \quad & U_{1j} \geq U_{1j'} + (\theta_{j'} - \theta_j)q_{1j'} \quad \text{for all } j, j' = \{1, 2, 3\}. \quad (4.3)
\end{align*}$$

The collusion proofness and incentive compatibility of the mechanism ensures that $S$ accepts the grand-contract only when the partition $\{\theta_2, \theta_1\}$ is realized. Indeed, the participation constraints for $S$ holds when she observes the partition $\{\theta_2, \theta_1\}$

$$f(\theta_2)w_{12} + f(\theta_1)w_{11} \geq 0, \quad (4.4)$$

whereas it doesn’t hold when the partition $\{\theta_3\}$ is observed

$$w_{13} = -(\theta_3 - \theta_2)q_{12} < 0.$$

Given that $S$’s refusal of the grand contract indicates that the realized type is $\theta_3$, the only

---

$^{10}$In our framework this implies that $A$ sends his message non-cooperatively whenever the side-contract fails to be established.
relevant constraints when $S$ refuses the grand-contract are the participation and the incentive compatibility constraints for type $\theta_3$, which are given by

$$\text{IR} \left( \theta_3|a_2 \right) \quad U_{23} \geq 0, \quad (4.5)$$

$$\text{IC} \left( \theta_3|a_2 \right) \quad U_{23} \geq U_{2j'} + (\theta_{j'} - \theta_3)q_{2j'} \text{ for all } j' = \{1, 2\}. \quad (4.6)$$

It is easy to notice that both constraints hold.

Let assume that $P$ holds non-passive beliefs following $S$’s refusal of accepting the grand-contract. In this case, the acceptance or refusal of the grand-contract can be used to update $P$’s beliefs on $A$’s type\(^{11}\).

It follows that $P$’s expected utility is given by

$$f(\theta_1)[W(q_{11}) - \theta_1q_{11} - V_{11}] +$$

$$+ f(\theta_2)[W(q_{12}) - \theta_2q_{12} - V_{12}] +$$

$$+ f(\theta_3)[W(q_{23}) - \theta_3q_{23} - V_{23}]. \quad (4.7)$$

Having this schedule in place, since the output levels $q_{22} = q_{21} = q_{13} = 0$ are realized with probability zero because the correspondent strategies are off the equilibrium path, it is convenient to consider only the remaining output levels. The optimal set of relevant output levels is determined maximizing (4.8) with respect to $q_{11}, q_{12}, q_{23}$. A brief inspection reveals that

\(^{11}\)Accordingly, conditional probabilities become

$$p(\theta_j|a_2) = 0 \quad \text{for all } j = 1, 2$$

$$p(\theta_3|a_2) = 1$$

$$p(\theta_j|a_2) = \frac{f(\theta_j)}{f(\theta_1) + f(\theta_2)} \quad \text{for all } j = 1, 2$$

$$p(\theta_3|a_1) = 0$$
\[ V_{23} = V_{3}^{cf} = 0, \]
\[ V_{12} = V_{2}^{cf} = 0, \]
\[ V_{11} = V_{1}^{cf} = (\theta_2 - \theta_1)q_{22}, \]

which implies that (4.8) is the same objective function for \( P \) as in the collusion-free problem. The mechanism implements the same expected payoff for \( P \) as the optimal collusion-free one.

### 4.3.1 Remarks on the second-best grand-mechanism

**Risk Aversion**

As mentioned before, results wouldn’t change with risk averse \( A \) and \( S \) since their ex post participation and incentive constraints would be identical and only those constraints are relevant for the analysis.

**Generic number of agents’ types and partitions**

The key findings of this paper wouldn’t change if a generic number of \( A \)'s types were to be considered. Suppose the cost of production, \( \theta \), takes \( n \) possible values from the set \( \Theta = \{\theta_1, \theta_2, ..., \theta_n\} \), where \( \theta_n \geq \theta_{n-1} \geq ... \geq \theta_1 \). \( S \)'s information structure is the partition

\[ \{\{\theta_1, \theta_2, ..., \theta_m\}, \{\theta_{m+1}, \theta_{m+2}, ..., \theta_n\}\} \]

of \( A \)'s type space \( \Theta \), where \( n > m > 1 \). Note that \( S \) is able to tell whether \( A \) is less efficient than type \( \theta_m \) or not, but she can’t distinguish types inside each partition. All the other assumptions remain the same. Having this schedule in place, it is easy to notice that the design problem is not very different from the one analyzed before. \( P \) can still offer an incentive compatible and collusion-proof grand-contract that leaves \( S \) with a negative payoff when a certain partition is observed: this induces her to refuse the grand-contract in some states of the world and allows \( P \) to extract her information for free. The complete proof is offered in Appendix.
Bargaining power and side contracting

One aspect regarding bargaining power is worth noticing: the result do not depend on the distribution of the bargaining power allocation inside the coalition. Consider for example the polar case where A has all the bargaining power in designing the collusive agreement. As mentioned before, both the participation and the incentive compatibility constraints for all types hold when S accepts the grand contract. Moreover, A doesn’t require the participation of S to play along with a misreport. It follows that A is unable to find a profitable manipulation.

Collusion under symmetric information

The intensity of asymmetric information within the coalition does not affect the results of this paper. Soft supervisory information would be useful even when collusion takes place under symmetric information, i.e. S learns $\theta$. In this case, the fact that S is not required for the production of the good drives the result because $P$ uses $S$ merely to get information from her acceptance or refusal of the grand-contract.

Sequential implementation

The same second-best outcome could be obtained through a sequential mechanism which consists of decentralized and no-supervision contracting. See Motta (2008) for further details.

4.4 Alternative informational structure

In this section an alternative information structure is considered. Following FLM (2003), the unitary cost of production, $\theta$, takes two possible values from the set $\Theta = \{\theta_1, \theta_2\}$, where $\theta_2 - \theta_1 = \Delta \theta \geq 0$. The distribution of the cost, $f(\theta)$, is common knowledge while the realization of $\theta$ is A’s private information. S is uninformed about the agent’s type. Nonetheless, he receives a signal $\tau$ on the agent’s cost. $\tau$ is drawn from a discrete distribution on $T = \{\tau_1, \tau_2\}$. The joint probabilities on $(\theta_i, \tau_j)$ are defined as $p_{ij} = \text{Prob}(\theta = \theta_i, \tau = \tau_j)$ with $p_{ij} > 0$ for all $i, j$.

From the joint distribution above, one can derive the conditional probabilities $p(\theta_i|\tau_j)$. There is a positive correlation between signals and types when the monotone likelihood ratio property is satisfied

$$\frac{p(\theta_1|\tau_1)}{p(\theta_2|\tau_1)} = \frac{p_{11}}{p_{21}} \geq \frac{p(\theta_1|\tau_2)}{p(\theta_2|\tau_2)} = \frac{p_{12}}{p_{22}}.$$  

S is assumed to be risk averse and to have a CARA
utility function defined over his monetary payoff \( x \): \( U_s = \frac{1}{\rho} (1 - e^{-rx}) \). As before, \( A \)'s utility function is given by \( t - \theta q \) where \( t \) is the transfer he receives from \( P \), while \( S \)'s salary is denoted by \( s \). \( P \)'s utility for a given output \( q \), transfer level \( t \) and wage \( s \) is \( W(q) - t - s \), where \( W(.) \) is subject to same conditions as in the previous sections.

The timing of the game is as follows:

- \( A \) learns \( \theta \) and \( \tau \). \( S \) learns only \( \tau \).
- \( P \) offers a grand-mechanism to \( S \) and \( A \).
- \( S \) and \( A \) accept or refuse the grand-mechanism. If \( A \) refuses, the game ends. If \( S \) refuses and \( A \) accepts production and transfers take place as specified in the grand-contract. If both \( S \) and \( A \) accept the grand-mechanism \( S \) may offer a collusive side-contract to \( A \) who accepts or refuses this contract. If \( A \) refuses, the grand-mechanism is played non-cooperatively by \( S \) and \( A \).
- Production and transfers take place.

Given this new information structure, it exists a grand-mechanism \( GC_{sb}^1 \) that allows \( P \) to achieve the same expected payoff as the optimal collusion-free outcome: as before, the message spaces for \( S \) is \( M_s = \{a_1, a_2\} \) while \( A \)'s message space is given by \( M_a = \{\theta_1, \theta_2\} \). Consider a contract involving the following:

\[
\begin{align*}
t_{12} &= \theta_2 q_{12} & s_{12} &= \frac{p(\theta_1 | \tau_1)}{p(\theta_2 | \tau_1)} \xi \\
t_{11} &= \Delta \theta q_{12} + \theta_1 q_{11} & s_{11} &= \varepsilon,
\end{align*}
\]

and

\[
\begin{align*}
t_{22} &= \theta_2 q_{22} & s_{22} &= 0 \\
t_{21} &= \Delta \theta q_{22} + \theta_1 q_{21} & s_{21} &= 0,
\end{align*}
\]

where \( \varepsilon \in R^+ \) is small and the output profile is weakly decreasing, i.e. \( q_{i1} \geq q_{i2} \) for all \( i \in \{1, 2\} \).

The relevant constraints for \( A \) are the participation constraints (IR) and the incentive compatibility constraints (IC) for all types,
\[ U_{ij} \geq 0 \quad \text{for all } i, j = \{1, 2\}, \quad (4.8) \]

\[ U_{ij} \geq U_{ij'} + (\theta_{j'} - \theta_j)q_{ij'} \quad \text{for all } i, j, j' = \{1, 2\}. \quad (4.9) \]

It is easy to notice that the composition of the grand-mechanisms is incentive compatible. The efficient agent is indifferent between telling the truth or lying to \( P \) regardless of \( S \)'s acceptance or refusal. Indeed, the following incentive constraints hold

\[
\begin{align*}
t_{11} - \theta_1 q_{11} &= \Delta \theta q_{12} \geq t_{12} - \theta_1 q_{12} = \Delta \theta q_{12}, \\
 t_{21} - \theta_1 q_{21} &= \Delta \theta q_{22} \geq t_{22} - \theta_1 q_{22} = \Delta \theta q_{22}.
\end{align*}
\]

Similarly, the inefficient type prefers to tell the truth to \( P \) since

\[
\begin{align*}
t_{12} - \theta_2 q_{12} &= 0 \geq t_{11} - \theta_2 q_{11} = \Delta \theta (q_{12} - q_{11}), \\
 t_{22} - \theta_2 q_{22} &= 0 \geq t_{21} - \theta_2 q_{21} = \Delta \theta (q_{22} - q_{21}).
\end{align*}
\]

For \( \varepsilon \) small enough\(^{12}\), the participation constraints for \( S \) holds when the realized signal is \( \tau_1 \),

\[
p(\theta_1 | \tau_1) \frac{1}{r} \left( 1 - e^{-r\varepsilon} \right) + p(\theta_2 | \tau_1) \frac{1}{r} \left( 1 - e^{- \frac{r}{p(\theta_1 | \tau_1) \varepsilon}} \right) \geq 0, \]

whereas it doesn’t hold when the signal \( \tau_2 \) is observed,

\[
p(\theta_1 | \tau_2) \frac{1}{r} \left( 1 - e^{-r\varepsilon} \right) + p(\theta_2 | \tau_2) \frac{1}{r} \left( 1 - e^{- \frac{r}{p(\theta_1 | \tau_1) \varepsilon}} \right) < 0, \]

\(^{12}\)In this case \( S \)'s utility function is given by,

\[
\lim_{\varepsilon \to 0} U_i = p(\theta_1 | \tau_i) \varepsilon - p(\theta_2 | \tau_i) \left( \frac{p(\theta_1 | \tau_i)}{p(\theta_2 | \tau_i)} \varepsilon \right) \quad \text{for all } i = \{1, 2\}.
\]
where \( \frac{p(\theta_1 | \tau_1)}{p(\theta_2 | \tau_1)} > \frac{p(\theta_1 | \tau_2)}{p(\theta_2 | \tau_2)} \).

Notice that this centralized organization is not subject to coalition formation between \( S \) and \( A \): as long as the side-contract is offered after the acceptance of the grand contract by both parties and \( \varepsilon \) is small enough, \( S \) and \( A \) cannot find any profitable collective manipulation to play along with. Suppose that \( S \) accepts the grand-contract. In this case \( S \) would like to induce \( A \) to report he has type \( \theta_1 \) in the attempt to avoid the negative transfer \( s_{12} \). Anyway for small values of \( \varepsilon \), \( S \) strictly prefers paying \( s_{12} \) to \( P \) than offering a bribe \( \Delta \theta (q_{12} - q_{11}) \) to \( A \) for misreporting his type as \( \theta_1 \).

The collusion proofness and incentive compatibility of \( GC_{1b} \) implies that when \( \tau_2 \) (\( \tau_1 \)) is realized, \( S \) refuses (accepts) the grand-contract. Therefore when he has observed \( S \)'s decision, \( P \) updates his beliefs on the agent’s type.

It follows that \( P \) expected utility is given by,

\[ t_{12} = \theta_2 q_{12} \]
\[ t_{11} = \Delta \theta q_{12} + \theta_1 q_{11} + \Delta \]
\[ s_{12} = \varepsilon \]
\[ s_{11} = -\frac{p(\theta_2 | \tau_2)}{p(\theta_2 | \tau_1) + \varepsilon} \]

and

\[ t_{22} = \theta_2 q_{22} \]
\[ t_{21} = \Delta \theta q_{22} + \theta_1 q_{21} \]
\[ s_{22} = 0 \]
\[ s_{21} = 0 \]

where \( \varepsilon \in R^+ \) is small, the output profile is weakly decreasing, i.e. \( q_i \geq q_{i2} \) for all \( i \in \{1, 2\} \) and \( \Delta = \varepsilon + \frac{p(\theta_2 | \tau_2)}{p(\theta_2 | \tau_1) + \varepsilon} \).

This grand contract is collusion proof and incentive compatible: differently from the previous case, it induces \( S \) to accept (refuse) when \( \tau_2 \) (\( \tau_1 \)) is realized. When he has observed \( S \)'s decision, \( P \) updates his beliefs on the agent’s type: as in the previous case, for small values of \( \varepsilon \), this grand-contract allows \( P \) to implement the second best outcome.

\[ p(\theta_j | a_1) = p(\theta_j | \tau_1) \quad \text{for } j = \{1, 2\} \]
\[ p(\theta_j | a_2) = p(\theta_j | \tau_2) \quad \text{for } j = \{1, 2\} \]
The optimal contract solves

\[
p(\theta_1|a_1) [W(q_{11}) - \theta_1 q_{11} - U_{11}] + \quad (4.11) + \quad p(\theta_2|a_1) [W(q_{12}) - \theta_2 q_{12} - U_{12}] + \quad \]
\[
p(\theta_1|a_2) [W(q_{21}) - \theta_1 q_{21} - U_{21}] + \quad + \quad p(\theta_2|a_2) [W(q_{22}) - \theta_2 q_{22} - U_{22}].
\]

The optimal contract solves

\[
\max_{\{q_{1i}, q_{2i}, u_{1i}, u_{2i}\}_{i \in \{1,2\}}} \quad p(\theta_1|a_i) [W(q_{1i}) - \theta_1 q_{1i} - u_{1i}] + p(\theta_2|a_i) [W(q_{2i}) - \theta_2 q_{2i} - u_{2i}]
\]
\[
s.t. \quad q_{1i} \geq q_{2i} \quad \text{for all } i \in \{1, 2\}.
\]

The solution of this problem yields what FLM (2003) denote by conditionally-optimal second-best, which implements the first-best outputs \(q_{1i}^{fb} = q_{1i}^{fb}\) for an efficient agent and outputs \(q_{i2}^{sb}\) for an inefficient one, where

\[
W'(q_{1i}^{fb}) = \theta_1,
\]
\[
W'(q_{i2}^{sb}) = \theta_2 + \frac{p_{i1}}{p_{i2}} \Delta \theta.
\]

Given that the agent is more likely to be efficient when \(a_1\) is observed than when \(a_2\) is observed, the inefficient agent’s output is more distorted after the observation of \(\tau_1\) rather than after the observation of \(\tau_2\). Indeed, reducing the efficient agent’s information rent calls for a greater allocative inefficiency of the inefficient agent’s output:

\[
q_{21}^{sb} < q_{22}^{sb}.
\]

This result replicates the FLM (2003) optimal contracting outcome with direct supervision which is equivalent to the optimal centralized contracting outcome when \(S\) and \(A\) do not collude. This proves that \(GC_{1}^{sb}\) allows \(P\) to achieve the same expected payoff as the optimal
**collusion-free outcome:** all the costs associated with collusion are fully eliminated.

### 4.4.1 Collusion under symmetric information

There are a couple of aspects worth noticing. As before, the result doesn’t depend on the distribution of the bargain power allocation inside the coalition and it holds regardless of which party makes the take-it-or-leave-it side-offer. Moreover, the intensity of asymmetric information within the coalition is not relevant either. Soft supervisory information would be useful even when collusion takes place under symmetric information, i.e. $S$ learns $\theta$. In this case $P$ could implement the second-best outcome by offering the following grand contract $GC_{2}^{sb}$,

\[
\begin{align*}
t_{12} &= \theta_2 q_{12} \quad s_{12} = -\varepsilon \\
t_{11} &= \Delta \theta q_{12} + \theta_1 q_{11} \quad s_{11} = 0,
\end{align*}
\]

and

\[
\begin{align*}
t_{22} &= \theta_2 q_{22} \quad s_{22} = 0 \\
t_{21} &= \Delta \theta q_{22} + \theta_1 q_{21} \quad s_{21} = 0,
\end{align*}
\]

where $\varepsilon \in R^+$ is small and the output profile is weakly decreasing, i.e. $q_{i1} \geq q_{i2}$ for all $i \in \{1, 2\}$.

Notice that the relevant constraints for $A$ are unchanged with respect to previous case: it follows that this grand-contract is also incentive compatible. Moreover this centralized organization is not subject to collusion. That is, suppose that $S$ accepts the grand-contract. In this case $S$ would like to induce $A$ to report he has type $\theta_1$ in the attempt to avoid the negative transfer $s_{12}$: anyway for small values of $\varepsilon$, $S$ strictly prefers paying $s_{12}$ to $P$ than offering a bribe $\Delta \theta (q_{12} - q_{11})$ to $A$ for misreporting his type as $\theta_1$.

As before, the participation constraints for $S$ holds when the realized type is $\theta_1$,

\[
\frac{1}{r} (1 - e^{-r s_{11}}) \geq 0,
\]

whereas it doesn’t hold when the realized type is $\theta_2$. 

94
\[ \frac{1}{r} (1 - e^{\tau^*}) < 0. \]

It follows that when \( \tau_2 (\tau_1) \) is realized \( S \) refuses (accepts) the grand-contract and \( P \) updates his beliefs on the agent’s type based on \( S \)’s acceptance decision. \( P \)’s expected utility is given by,

\[
P(\theta_1|a_1) [W(q_{11}) - \theta_1 q_{11} - U_{11}] + \]
\[
+ p(\theta_2|a_1) [W(q_{12}) - \theta_1 q_{12} - U_{12}] + \]
\[
+ p(\theta_1|a_2) [W(q_{21}) - \theta_1 q_{21} - U_{21}] + \]
\[
+ p(\theta_2|a_2) [W(q_{22}) - \theta_1 q_{22} - U_{22}]. \tag{4.12}
\]

The optimal contract solves

\[
\max_{\{q_{i1}, q_{i2}, u_{i1}, u_{i2}\} \in \{1, 2\}} p(\theta_1|a_i) [W(q_{i1}) - \theta_1 q_{i1} - u_{i1}] + p(\theta_2|a_i) [W(q_{i2}) - \theta_2 q_{i2} - u_{i2}]
\]

s.t. \( q_{i1} \geq q_{i2} \) for all \( i \in \{1, 2\} \).

A simple inspection reveals that the solution of this problem is identical to the one presented in the previous section, i.e. the conditionally-optimal second-best outcome. All the costs associated with collusion are fully eliminated also in this case.

### 4.5 Conclusions

This paper analyzes the role of delegation when a principal delegates to expert intermediaries / managers, who play no role in actual production and when collusion between the intermediaries and the productive agents is allowed. If collusion is allowed only after the acceptance of the principal’s contract, it is not harmful for the principal. This result is robust to alternative information structures and it holds for a quite generic specification of agent’s type.
Bibliography


Appendix: Proof of Proposition 4.1

In this part of the Appendix the implementation of the grand-mechanism $GC^{sb}$ is analyzed. In the first part, it is proven that this grand-mechanism is collusion-proof; in the second part it is shown that $GC^{sb}$ allows $P$ to achieve the same expected payoff as the optimal collusion-free outcome.

Part 1

The grand-contract offered by $P$ entails,

\begin{align*}
    t_{13} &= 0, & s_{13} &= -(\theta_3 - \theta_2)q_{12}, \\
    t_{12} &= \theta_2 q_{12}, & s_{12} &= 0, \\
    t_{11} &= (\theta_2 - \theta_1)q_{12} + \theta_1 q_{11}, & s_{11} &= 0,
\end{align*}

and

\begin{align*}
    t_{13} &= \theta_3 q_{23}, \\
    t_{12} &= 0, \\
    t_{11} &= 0,
\end{align*}

where $q_{13} = q_{22} = q_{21} = 0$.

To start with, consider the collusion proofness of this grand contract. The process of collusion is formalized by assuming that $S$ offers a take-it-or-leave-it side-contract to $A$, after the acceptance of the grand contract by both parties. Therefore the collusion proofness of this mechanism must be tested only in case $S$ accepts the grand-contract, i.e. whenever she reports $a_1$ to $P$.

Under collusion, $\tilde{c}(\theta_i)$ denotes the misreport of type $\theta_i$ as type $\tilde{\theta}_i(\theta_i)$, where $\tilde{c} : \{1, 2, 3\} \rightarrow \{1, 2, 3\}$. Under this manipulation of $A$’s report, the coalition information rent is $V_{1i\tilde{\theta}_i} + (\tilde{\theta}_i(\theta_i) - \theta_i)q_{1i\tilde{\theta}_i}$. The share of the information rent for $A$ is given by $\tilde{u}(\theta_i)$, where $\tilde{u}(\theta_i) : i \rightarrow \mathbb{R}$ for $i = \{1, 2, 3\}$. Denote with $\{\tilde{c}(\theta_i), \tilde{\tau}(\theta_i)\}_{i \in \{1, 2, 3\}}$ the manipulation of the outcome $\{q_{1i}, V_{1i}, U_{1i}\}$. In order for such an outcome to be implementable, the expected value of $V_{1i} - U_{1i}$ must be weakly larger than the expected value of $V_{1i\tilde{\theta}_i} + (\tilde{\theta}_i(\theta_i) - \theta_i)q_{1i\tilde{\theta}_i}$ under any possible manipulation available to $S$. In order for the manipulation $\{\tilde{c}(\theta_i), \tilde{\tau}(\theta_i)\}_{i \in \{1, 2, 3\}}$ to be available to $S$, she must
leave the agents at least with the utility level correspondent to their outside option,

\[
\text{AIR} (\theta) : \; \tilde{u}(\theta_i) \geq U_{1i} \quad \text{for all } i = \{1, 2, 3\}. \tag{4.13}
\]

Since \( S \) cannot distinguish the different types within the same partition cell, she must provide \( A \) the incentive not to imitate the other type in the same partition cell. This issue is not relevant when \( A \) has type \( \theta_3 \). At the contrary, it is relevant when \( S \) observes the partition \( \{\theta_2, \theta_1\} \). In this case, for \( \{\tilde{c}(\theta_i), \tilde{r}(\theta_i)\}_{i \in \{1,2,3\}} \) to be an available misreport for \( S \) the following "agent’s collusion stage incentive compatibility constraint" must hold:

\[
\text{AIC} (\theta_{i'}|\theta_i) : \; \tilde{u}(\theta_i) \geq \tilde{u}(\theta_{i'}) + (\theta_{i'} - \theta_i) q_{1i} \tilde{c}(\theta_i) \quad \text{for all } i, i' = \{1, 2\}. \tag{4.14}
\]

What follows are the conditions for the feasibility of the outcome \( \{q_i, V_i, U_i\} \)

\[
\{\theta_3, U_{13}\} \in \arg \max_{\{\tilde{c}(\theta_3), \tilde{u}(\theta_3)\}} \left[ V_{13} \tilde{c}(\theta_3) + (\theta_{\tilde{c}(\theta_3)} - \theta_3) q_{13} \tilde{c}(\theta_3) - \tilde{u}(\theta_3) \right] \tag{4.15}
\]

\[
s.t. \; \text{AIR} (\theta_3),
\]

\[
\{\theta_i, U_{1i}\}_{i \in \{1, 2\}} \in \arg \max_{\{\tilde{c}(\theta_i), \tilde{u}(\theta_i)\}_{i \in \{1, 2\}}} \sum_{i \in \{1, 2\}} \left[ V_{1i} \tilde{c}(\theta_i) + (\theta_{\tilde{c}(\theta_i)} - \theta_i) q_{1i} \tilde{c}(\theta_i) - \tilde{u}(\theta_i) \right] \tag{4.16}
\]

\[
s.t. \; \text{AIC} (\theta_2|\theta_1), \; \text{AIC} (\theta_1|\theta_2), \; \text{AIR} (\theta_2) \; \text{and} \; \text{AIR} (\theta_1).
\]

Following Celik (2008), conditions (17) and (18) are defined as "collusion feasibility conditions" which guarantees that \( S \) is not able to find a profitable manipulation.

Now consider a manipulation where type \( \theta_3 \) is misreported as type \( \theta_1 \) such that \( \tilde{c}(\theta_1) = 1, \tilde{c}(\theta_2) = 2, \tilde{c}(\theta_3) = 1, \tilde{u}(\theta_1) = (\theta_2 - \theta_1) q_{12}, \tilde{u}(\theta_2) = 0 \) and \( \tilde{u}(\theta_3) = 0 \). This deviation satisfies \( \text{AIR} \) and \( \text{AIC} \) constraints. This deviation is not profitable because it does not improve the objective function in (17),

\[
V_{13} - U_{13} \geq V_{11} + (\theta_1 - \theta_3) q_{11}
\]
After substituting and rearranging, the former inequality is reduced to

\[-(\theta_3 - \theta_2)q_{12} \geq -(\theta_3 - \theta_1)q_{11}\]

This rules out type \(\theta_3\)'s imitation of type \(\theta_1\), because \(q_{11} \geq q_{12}\) and \((\theta_3 - \theta_1) \geq (\theta_3 - \theta_2)\).

Now consider a manipulation where type \(\theta_3\) is misreported as type \(\theta_2\) such that \(\tilde{c}(\theta_1) = 1\), \(\tilde{c}(\theta_2) = 2\), \(\tilde{c}(\theta_3) = 2\), \(\tilde{u}(\theta_1) = (\theta_2 - \theta_1)q_{12}\), \(\tilde{u}(\theta_2) = 0\) and \(\tilde{u}(\theta_3) = 0\). This deviation satisfies \textbf{AIR} and \textbf{AIC} constraints. This deviation is not profitable if the following inequality holds,

\[V_{13} - U_{13} \geq V_{12} + (\theta_2 - \theta_3)q_{12}.\]

After substituting and rearranging, the former inequality is reduced to

\[-(\theta_3 - \theta_2)q_{12} \geq -(\theta_3 - \theta_2)q_{12}.\]

This rules out type \(\theta_3\)'s imitation of type \(\theta_2\) because this deviation does not improve the objective function in (17).

Now consider a manipulation where type \(\theta_2\) is misreported as type \(\theta_1\) such that \(\tilde{c}(\theta_1) = 1\), \(\tilde{c}(\theta_2) = 1\), \(\tilde{c}(\theta_3) = 3\), \(\tilde{u}(\theta_1) = (\theta_2 - \theta_1)q_{12} + q_{11}\), \(\tilde{u}(\theta_2) = (\theta_2 - \theta_1)q_{11}\) and \(\tilde{u}(\theta_3) = 0\). This deviation satisfies \textbf{AIR} and \textbf{AIC} constraints. This deviation is not profitable if the following inequality holds:

\[f(\theta_1)(V_{11} - U_{11}) + f(\theta_2)(V_{12} - U_{12}) \geq [f(\theta_1) + f(\theta_2)] [V_{11} - (\theta_2 - \theta_1)(q_{12} + q_{11})].\]

After substituting and rearranging, the former inequality is reduced to

\[0 \geq -(\theta_2 - \theta_1)q_{11}.\]

This rules out type \(\theta_2\)'s imitation of type \(\theta_1\).

Now consider a manipulation where type \(\theta_2\) is misreported as type \(\theta_3\) such that \(\tilde{c}(\theta_1) = 1\), \(\tilde{c}(\theta_2) = 3\), \(\tilde{c}(\theta_3) = 3\), \(\tilde{u}(\theta_1) = (\theta_2 - \theta_1)q_{12}\), \(\tilde{u}(\theta_2) = 0\) and \(\tilde{u}(\theta_3) = 0\). This deviation satisfies \textbf{AIR} and \textbf{AIC} constraints. This deviation is not profitable if the following inequality holds:
\[ f(\theta_1) (V_{11} - U_{11}) + f(\theta_2) (V_{12} - U_{12}) \geq f(\theta_1) (V_{11} - U_{11}) + f(\theta_2)V_{13}. \]

After substituting and rearranging, the former inequality is reduced to

\[ 0 \geq - (\theta_3 - \theta_2) q_{12}. \]

This rules out type \( \theta_2 \)'s imitation of type \( \theta_3 \).

Now consider a manipulation where type \( \theta_1 \) is misreported as type \( \theta_2 \) such that \( \bar{c}(\theta_1) = 2 \), \( \bar{c}(\theta_2) = 2 \), \( \bar{c}(\theta_3) = 3 \), \( \bar{u}(\theta_1) = U_{12}, \bar{u}(\theta_2) = U_{12} - (\theta_2 - \theta_1) q_{12} \) and \( \bar{u}(\theta_3) = 0 \). This deviation satisfies \textbf{AIR} and \textbf{AIC} constraints. This deviation is not profitable if the following inequality holds:

\[ f(\theta_1) (V_{11} - U_{11}) + f(\theta_2) (V_{12} - U_{12}) \geq [f(\theta_1) + f(\theta_2)] [V_{11} + (\theta_2 - \theta_1) q_{12} - U_{12}]. \]

After substituting and rearranging, the former inequality is reduced to

\[ 0 \geq (\theta_2 - \theta_1) q_{12} - (\theta_2 - \theta_1) q_{12}. \]

This rules out type \( \theta_1 \)'s imitation of type \( \theta_2 \).

Now consider a manipulation where type \( \theta_1 \) is misreported as type \( \theta_3 \) such that \( \bar{c}(\theta_1) = 3 \), \( \bar{c}(\theta_2) = 2 \), \( \bar{c}(\theta_3) = 3 \), \( \bar{u}(\theta_1) = U_{12}, \bar{u}(\theta_2) = 0 \) and \( \bar{u}(\theta_3) = 0 \). This deviation satisfies \textbf{AIR} and \textbf{AIC} constraints. This deviation is not profitable if the following inequality holds:

\[ f(\theta_1) (V_{11} - U_{11}) + f(\theta_2) (V_{12} - U_{12}) \geq f(\theta_1)V_{13} + f(\theta_2) (V_{12} - U_{12}). \]

After substituting and rearranging, the former inequality is reduced to

\[ 0 \geq - (\theta_3 - \theta_2) q_{12}. \]

This rules out type \( \theta_1 \)'s imitation of type \( \theta_3 \).

Finally consider a manipulation where both types \( \theta_1 \) and \( \theta_2 \) are misreported as type \( \theta_3 \) such
that \( \tilde{c}(\theta_1) = 3, \tilde{c}(\theta_2) = 3, \tilde{c}(\theta_3) = 3, \tilde{u}(\theta_1) = U_{12}, \tilde{u}(\theta_2) = U_{12} \) and \( \tilde{u}(\theta_3) = 0 \). This deviation satisfies \textbf{AIR} and \textbf{AIC} constraints. This deviation is not profitable if the following inequality holds:

\[
f(\theta_1) (V_{11} - U_{11}) + f(\theta_2) (V_{12} - U_{12}) \geq [f(\theta_1) + f(\theta_2)] [V_{13} - U_{12}].
\]

After substituting and rearranging, the former inequality is reduced to

\[
0 \geq [f(\theta_1) + f(\theta_2)][-(\theta_3 - \theta_2)q_{12} - (\theta_2 - \theta_1)q_{12}].
\]

This rules out types \( \theta_1 \) and \( \theta_2 \)’s imitation of type \( \theta_3 \).

\textbf{QED}

**Part 2**

Given that the mechanism is collusion-proof, it follows that \( A \) and \( S \) respond to the grand-mechanism in a non-cooperative fashion. To begin with, notice that the participation and the incentive compatibility constraints for all types hold when \( S \) accepts the grand contract:

\[
\textbf{IR} (\theta_j|a_1) \quad U_{1j} \geq 0 \quad \text{for all } j = \{1,2,3\}, \tag{4.17}
\]

\[
\textbf{IC} (\theta_j|a_1) \quad U_{1j} \geq U_{1j'} + (\theta_{j'} - \theta_j)q_{1j'} \quad \text{for all } j,j' = \{1,2,3\}. \tag{4.18}
\]

The most efficient agent, \( \theta_1 \), prefers to tell the truth to \( P \): indeed, the following incentive constraints hold

\[
t_{11} - \theta_1q_{11} = (\theta_2 - \theta_1)q_{12} \geq t_{12} - \theta_1q_{12} = (\theta_2 - \theta_1)q_{12},
\]

\[
t_{11} - \theta_1q_{11} = (\theta_2 - \theta_1)q_{12} \geq t_{13} - \theta_1q_{13} = 0.
\]

Similarly, type \( \theta_2 \) prefers to tell the truth to \( P \) since

\[
t_{12} - \theta_2q_{12} = 0 \geq t_{13} - \theta_2q_{13} = 0,
\]
\[ t_{12} - \theta_2 q_{12} = 0 \geq t_{11} - \theta_2 q_1 = (\theta_2 - \theta_1) (q_{12} - q_{11}). \]

Finally, type \( \theta_3 \) prefers to tell the truth to \( P \) since

\[ t_{13} - \theta_3 q_{13} = 0 \geq t_{11} - \theta_3 q_{11} = (\theta_2 - \theta_1) q_{12} - (\theta_3 - \theta_1) q_{11}. \]

Therefore the participation constraints for \( S \) hold when she observes the partition \( \{ \theta_2, \theta_1 \} \)

\[ f(\theta_2)w_{12} + f(\theta_1)w_{11} \geq 0 \]  \hspace{1cm} (4.19)

whereas it doesn’t hold when the partition \( \{ \theta_3 \} \) is observed,

\[ w_{13} = -(\theta_3 - \theta_2) q_{12} < 0 \]

As long as the grand-contract is accepted by \( S \) only when the partition \( \{ \theta_2, \theta_1 \} \) is realized, it follows that when \( S \) refuses the grand contract the only relevant constraints are the participation and the incentive compatibility constraints for type \( \theta_3 \),

\[
\text{IR (} \theta_3 | a_2 \text{) } U_{23} \geq 0, \hspace{1cm} (4.20)
\]

\[
\text{IC (} \theta_3 | a_2 \text{) } U_{23} \geq U_{2j'} + (\theta_{j'} - \theta_3) q_{2j'} \text{ for all } j' = \{1, 2\}. \hspace{1cm} (4.21)
\]

Notice that also in this case type \( \theta_3 \) prefers to tell the truth to \( P \) since

\[ t_{23} - \theta_3 q_{23} = 0 \geq t_{22} - \theta_3 q_{22} = 0 \]

\[ t_{23} - \theta_3 q_{23} = 0 \geq t_{21} - \theta_3 q_{21} = 0 \]

Having this schedule in place, assume that \( P \) holds non-passive beliefs following \( S \)’s refusal
of accepting the grand-contract. In this case, the acceptance or refusal of the grand-contract can be interpreted as a cheap talk stage, which is used to update $P$’s beliefs on the $A$’s type\textsuperscript{15}.

It follows that $P$’s expected utility is given by:

$$
\begin{align*}
&f(\theta_1)[W(q_{11}) - \theta_1 q_{11} - V_{11}] + \\
&+ f(\theta_2)[W(q_{12}) - \theta_2 q_{12} - V_{12}] + \\
&+ f(\theta_3)[W(q_{23}) - \theta_3 q_{23} - V_{23}].
\end{align*}
$$

Given the output levels $q_{22} = q_{21} = q_{13} = 0$ are realized with probability zero, because the correspondent strategies are off the equilibrium path, it is possible to consider only the remaining output levels. The optimal set of relevant output levels is determined maximizing (24) with respect to $q_{11}, q_{12}, q_{23}$. A brief inspection reveals that:

$$
\begin{align*}
V_{23} &= V_{3}^{cf} = 0, \\
V_{12} &= V_{2}^{cf} = 0, \\
V_{11} &= V_{1}^{cf} = (\theta_2 - \theta_1)q_{22},
\end{align*}
$$

which implies that (24) is the same objective function for $P$ as in the collusion-free problem. The mechanism implements the same expected payoff for $P$ as the optimal collusion-free one.

QED.

\textsuperscript{15}Accordingly, conditional probabilities become

$$
\begin{align*}
p(\theta_j | a_2) &= 0 & \text{for all } j = 1, 2 \\
p(\theta_3 | a_2) &= 1 \\
p(\theta_j | a_2) &= \frac{f(\theta_j)}{f(\theta_1) + f(\theta_2)} & \text{for all } j = 1, 2 \\
p(\theta_3 | a_1) &= 0
\end{align*}
$$

105
Appendix: Generic number of types

This section of the Appendix studies the implementation of the grand-mechanism \(GC^{sbl}\) with finitely many cost levels. In the first part of the proof collusion-free supervision benchmark is studied, while in the second part it is shown that \(GC^{sbl}\) allows \(P\) to achieve the same expected payoff as the optimal collusion-free outcome.

Part 1

Suppose that \(S\) is not able to offer a side contract to \(A\). As in the standard case, \(P\) can set \(V_i = U_i\) for all \(i = \{1, 2, \ldots, n\}\) and obtain \(S\)'s participation in the mechanism. \(P\) can implement an output profile with the utility levels

\[
V_{n}^{cf} = U_{n}^{cf} = 0, \\
V_{n-1}^{cf} = U_{n-1}^{cf} = (\theta_{n} - \theta_{n-1})q_{n}, \\
\vdots \\
V_{m+1}^{cf} = U_{m+1}^{cf} = \sum_{j=m-2}^{n} U_{j}^{cf} + (\theta_{m-2} - \theta_{m-1})q_{m-2}, \\
V_{m}^{cf} = U_{m}^{cf} = 0, \\
V_{m-1}^{cf} = U_{m-1}^{cf} = (\theta_{m} - \theta_{m-1})q_{m}, \\
\vdots \\
V_{1}^{cf} = U_{1}^{cf} = \sum_{j=2}^{m} U_{j}^{cf} + (\theta_{2} - \theta_{1})q_{2}.
\]

As mentioned before, when \(S\) is introduced there is no monotonicity requirement regarding two output levels in separate partition cells. Under collusion free supervision, the optimal set of output levels are determined as

\[
\{q_{j}^{cf}\}_{j \in \{1, 2, \ldots, n\}} \in \arg\max \sum_{j=1}^{n} p(\theta_{j}) \left[ W(q_{j}) - \theta q_{j} - V_{j}^{cf} \right] \\
\text{s.t.} \\
q_{1} \geq q_{2} \geq \ldots \geq q_{m}, \\
q_{m+1} \geq q_{m+2} \geq \ldots \geq q_{n}.
\]

Part 2
The new message spaces are given by $M_s \in \{a_1, a_2\}$ and $M_a \in \{\theta_1, \theta_2, ..., \theta_n\}$, where again $a_1$ stands for "accept the grand contract" and $a_2$ indicates "refuse the grand contract". Denote by $s_{ij}$ (respectively $t_{ij}$, $q_{ij}$, $V_{ij}$ and $U_{ij}$) $S$’s wage (respectively the $A$’s transfer, the output target, the coalition information rent and $A$’s utility) when $A$ reports that he has type $\theta_j$ and $S$ reports $a_i$. The following grand-contract implements the second best outcome:

\[
\begin{align*}
    t_1n = 0 & \quad s_{1n} = -N \\
    t_{1n-1} = 0 & \quad s_{1n-1} = -N \\
    \quad \ldots & \quad \quad \ldots \\
    t_{1m+1} = 0 & \quad s_{1m-1} = -N \\
    t_{1m} = \theta_m q_{1m} & \quad s_{1m} = 0 \\
    t_{1m-1} = (\theta_m - \theta_{m-1}) q_{1m} + \theta_{m-1} q_{1m-1} & \quad s_{1m-1} = 0 \\
    \quad \ldots & \quad \quad \ldots \\
    t_{11} = \sum_{j=2}^{n} U_{1j} + (\theta_2 - \theta_1) q_{12} + \theta_1 q_{11} & \quad s_{11} = 0
\end{align*}
\]

and

\[
\begin{align*}
    t_{2n} = \theta_n q_{1n}, \\
    t_{2n-1} = (\theta_n - \theta_{n-1}) q_{1n} + \theta_{n-1} q_{1n-1}, \\
    \quad \ldots \\
    t_{2m+1} = \sum_{j=m-2}^{n} U_{1j} + (\theta_{m-2} - \theta_{m-1}) q_{1m-2} + \theta_{m-1} q_{1m-1}, \\
    t_{2m} = 0, \\
    t_{2m-1} = 0, \\
    \quad \ldots \\
    t_{21} = 0,
\end{align*}
\]

where $q_{1i} = 0$ for all $i = \{m + 1, m + 2, ..., n\}$, $q_{2i} = 0$ for all $i = \{1, 2, ..., m\}$, $q_{11} \geq q_{12} \geq \ldots \geq q_{1m}$, $q_{2m+1} \geq q_{2m+2} \geq \ldots \geq q_{2n}$ and $N = (\theta_m - \theta_{m-1}) q_{1m} - (\theta_n - \theta_{m-1}) q_{1m-1} \leq 0$.

Rewriting the grand-contract in terms of $A$’s utilities yields,
\[
U_{1n} = 0, \quad s_{1n} = -N, \\
U_{1n-1} = 0, \quad s_{1n-1} = -N, \\
\ldots \\
U_{1m+1} = 0, \quad s_{1m-1} = -N, \\
U_{1m} = 0, \quad s_{1m} = 0, \\
U_{1m-1} = (\theta_m - \theta_{m-1})q_{1m}, \quad s_{1m-1} = 0, \\
\ldots \\
\ldots \\
U_{11} = \sum_{j=2}^{n} U_{1j} + (\theta_2 - \theta_1)q_{12}, \quad s_{11} = 0,
\]

and

\[
U_{2n} = 0, \\
U_{2n-1} = (\theta_n - \theta_{n-1})q_{1n}, \\
\ldots \\
U_{2m+1} = \sum_{j=m-2}^{n} U_{1j} + (\theta_{m+2} - \theta_{m+1})q_{1m+2}, \\
U_{2m} = 0, \\
U_{2m-1} = 0, \\
\ldots \\
U_{21} = 0,
\]

To start with, consider the collusion proofness of this grand contract. The process of collusion is formalized by assuming that \( S \) offers a take-it-or-leave-it side-contract to \( A \), after the acceptance of the grand contract by both parties. Therefore the collusion proofness of the mechanism must be tested only in case \( S \) accepts the grand-contract, i.e. whenever she reports \( a_1 \) to \( P \).

Denote by \( \tilde{c}(\theta_i) \) the misreport of type \( \theta_i \) as type \( \theta_{\tilde{c}(\theta_i)} \), where \( \tilde{c} : \{1, 2, 3, \ldots, n\} \to \{1, 2, 3, \ldots, n\} \).

Under this manipulation of \( A \)'s report, the coalition information rent is \( V_{1\tilde{c}(\theta_i)} + (\theta_{\tilde{c}(\theta_i)} - \theta_i)q_{1\tilde{c}(\theta_i)} \).

The share of the information rent for \( A \) is given by \( \tilde{u}(\theta_i) \), where \( \tilde{u}(\theta_i) : i \to \mathbb{R} \) for \( i = \{1, 2, 3, \ldots, n\} \). Denote \( \{\tilde{c}(\theta_i), \tilde{r}(\theta_i)\}_{i \in \{1, 2, 3\}} \) as a manipulation of the outcome \( \{q_{1i}, V_{1i}, U_{1i}\} \).
In order for such an outcome to be implementable, the expected value of $V_{ii} - U_{ii}$ must be weakly larger than the expected value of $V_{i Geoff(\theta_{i})} + (\theta_{i Geoff(\theta_{i})} - \theta_{i})q_{1 Geoff(\theta_{i})}$ under any possible manipulation available to $S$. In order for the manipulation $\{\tilde{c}(\theta_{i}), \tilde{r}(\theta_{i})\}_{i \in \{1,2,3\}}$ to be available to $S$, she must leave the agents at least with the utility level correspondent to their outside option,

\[
\text{AIR} (\theta) : \quad \tilde{u}(\theta_{i}) \geq U_{ii} \quad \text{for all } i = \{1,2,3,...,n\}. \tag{4.23}
\]

Since $S$ cannot distinguish the different types within the same partition cell, she must provide $A$ the incentive not to imitate the other type in the same partition cell. In this case, for $\{\tilde{c}(\theta_{i}), \tilde{r}(\theta_{i})\}_{i \in \{1,2,3,...,n\}}$ to be an available misreport for $S$ the following "agent’s collusion stage incentive compatibility constraints" must hold:

\[
\text{AIC} (\theta'_{i} | \theta_{i}) : \quad \tilde{u}(\theta_{i}) \geq \tilde{u}(\theta'_{i}) + (\theta'_{i} - \theta_{i})q_{1 Geoff(\theta_{i})} \quad \text{for all } i, i' = \{1,2,...,m\}, \tag{4.24}
\]

\[
\text{AIC} (\theta'_{i} | \theta_{i}) : \quad \tilde{u}(\theta_{i}) \geq \tilde{u}(\theta'_{i}) + (\theta'_{i} - \theta_{i})q_{1 Geoff(\theta_{i})} \quad \text{for all } i, i' = \{m+1, m+2,...,n\}. \tag{4.25}
\]

What follows are the conditions for the feasibility of the outcome $\{q_{i}, \bar{V}_{i}, U_{i}\}$:

\[
\{\theta_{i}, U_{ii}\}_{i = \{1,2,...,m\}} \in \arg \max_{\{\tilde{c}(\theta_{i}), \tilde{r}(\theta_{i})\}_{i = \{1,2,...,m\}}} \sum_{i = \{1,2,...,m\}} \left[ V_{i Geoff(\theta_{i})} + (\theta_{i Geoff(\theta_{i})} - \theta_{i})q_{1 Geoff(\theta_{i})} - \tilde{u}(\theta_{i}) \right] \\
\text{s.t. } \text{AIC} (\theta_{i} | \theta'_{i}) \text{ and AIR} (\theta_{i}) \text{ for all } i = \{1,2,...,m\}, \tag{4.26}
\]

\[
\{\theta_{i}, U_{ii}\}_{i = \{m+1,...,n\}} \in \arg \max_{\{\tilde{c}(\theta_{i}), \tilde{r}(\theta_{i})\}_{i = \{m+1,...,n\}}} \sum_{i = \{m+1,...,n\}} \left[ V_{i Geoff(\theta_{i})} + (\theta_{i Geoff(\theta_{i})} - \theta_{i})q_{1 Geoff(\theta_{i})} - \tilde{u}(\theta_{i}) \right] \\
\text{s.t. } \text{AIC} (\theta_{i} | \theta'_{i}) \text{ and AIR} (\theta_{i}) \text{ for all } i = \{m+1,...,n\}. \tag{4.27}
\]

Following Celik (2008), conditions (28) and (29) are defined as "collusion feasibility
conditions" which guarantees that $S$ is not able to find a profitable manipulation.

First of all, when $S$ accepts the grand contract, the participation and incentive compatibility constraints hold for all types:

\[
\textbf{IR} \ (\theta_j|a_1) \quad U_{1j} \geq 0 \quad \text{for all } j = \{1, 2, 3, \ldots, n\}, \quad (4.28)
\]

\[
\textbf{IC} \ (\theta_j|a_1) \quad U_{1j} \geq U_{1j'} + (\theta_{j'} - \theta_j)q_{1j'} \quad \text{for all } j, j' = \{1, 2, 3, \ldots, n\}. \quad (4.29)
\]

To prove this, notice that types $\theta_i = \{m+1, \ldots, n\}$ have no stake in mimicking each other; indeed, the payoff would be zero regardless of the misreport. Moreover, types $\theta_i = \{1, 2, \ldots, m\}$ have no stake in misreporting as types $\theta_i = \{1, 2, \ldots, m\}$; indeed, the following incentive compatibility constraints for type $\theta_n$ hold,

\[
U_{1n} = 0 \geq U_{1m-1} + (\theta_{m-1} - \theta_n)q_{1m-1} \\
= (\theta_m - \theta_{m-1})q_{1m} - (\theta_n - \theta_{m-1})q_{1m-1},
\]

\[
U_{1n} = 0 \geq U_{1m-2} + (\theta_{m-2} - \theta_n)q_{1m-2} \\
= (\theta_m - \theta_{m-1})q_{1m} + (\theta_{m-1} - \theta_{m-2})q_{1m-1} - (\theta_n - \theta_{m-2})q_{1m-2},
\]

\[
U_{1n} = 0 \geq U_{11} + (\theta_{1} - \theta_n)q_{11} \\
= (\theta_m - \theta_{m-1})q_{1m} + (\theta_{m-1} - \theta_{m-2})q_{1m-1} + \ldots \\
+ (\theta_2 - \theta_1)q_{12} - (\theta_n - \theta_1)q_{11},
\]

Incentive compatibility constraints also hold for type $\theta_{n-1}$,
\[ U_{1n-1} = 0 \geq U_{1m-1} + (\theta_{m-1} - \theta_{n-1})q_{1m-1} \]
\[ = (\theta_m - \theta_{m-1})q_1m - (\theta_{n-1} - \theta_{m-1})q_{1m-1}, \]
\[ U_{1n-1} = 0 \geq U_{1m-2} + (\theta_{m-2} - \theta_{n-1})q_{1m-2} \]
\[ = (\theta_m - \theta_{m-1})q_1m + (\theta_{m-1} - \theta_{m-2})q_{1m-1} - (\theta_{n-1} - \theta_{m-2})q_{1m-2}; \]
\[
\vdots
\]
\[ U_{1n-1} = 0 \geq U_{11} + (\theta_1 - \theta_{n-1})q_11 \]
\[ = (\theta_m - \theta_{m-1})q_1m + (\theta_{m-1} - \theta_{m-2})q_{1m-1} + \]
\[ + (\theta_2 - \theta_1)q_{12} - (\theta_{n-1} - \theta_1)q_{11}. \]

The same applies for the all types \( \{\theta_i\}_{i=m+1,...,n}. \)

On the other hand, types \( \{\theta_i\}_{i=1,2,...,m} \) have no stake in misreporting as types \( \{\theta_i\}_{i=m+1,...,n} \) because they would obtain a payoff equal to zero, which corresponds to their outside option.

Types \( \{\theta_i\}_{i=1,2,...,m} \) have also no stake in mimicking each other, since the following incentive compatibility constraints hold for type \( \theta_m: \)

\[ U_{1m} = 0 \geq U_{1m-1} + (\theta_{m-1} - \theta_m)q_{1m-1} \]
\[ = (\theta_m - \theta_{m-1})q_1m - (\theta_m - \theta_{m-1})q_{1m-1}, \]
\[ U_{1m} = 0 \geq U_{1m-2} + (\theta_{m-2} - \theta_m)q_{1m-2} \]
\[ = (\theta_m - \theta_{m-1})q_1m + (\theta_{m-1} - \theta_{m-2})q_{1m-1} - (\theta_m - \theta_{m-2})q_{1m-2}; \]
\[
\vdots
\]
\[ U_{1m} = 0 \geq U_{11} + (\theta_1 - \theta_m)q_{11} \]
\[ = (\theta_m - \theta_{m-1})q_1m + (\theta_{m-1} - \theta_{m-2})q_{1m-1} + \]
\[ + (\theta_2 - \theta_1)q_{12} - (\theta_m - \theta_1)q_{11}. \]

Incentive compatibility constraints also hold for type \( \theta_{m-1} \)
$U_{1m-1} = (\theta_m - \theta_{m-1})q_{1m} \geq U_{1m} + (\theta_m - \theta_{m-1})q_{1m}$

$= (\theta_m - \theta_{m-1})q_{1m} - (\theta_m - \theta_{m-1})q_{1m} = 0$

$U_{1m-1} = (\theta_m - \theta_{m-1})q_{1m} \geq U_{1m-2} + (\theta_{m-2} - \theta_{m-1})q_{1m-2}$

$= (\theta_m - \theta_{m-1})q_{1m} + (\theta_{m-1} - \theta_{m-2})q_{1m-1} - (\theta_{m-1} - \theta_{m-2})q_{1m-2} = 0$

$...$

$U_{1m-1} = (\theta_m - \theta_{m-1})q_{1m} \geq U_{11} + (\theta_1 - \theta_{m-1})q_{11}$

$= (\theta_m - \theta_{m-1})q_{1m} + (\theta_{m-1} - \theta_{m-2})q_{1m-1} + ...$

$+ (\theta_2 - \theta_1)q_{12} - (\theta_{m-1} - \theta_1)q_{11} = 0$

The same applies for the all types $\{\theta_i\}_{i=1,2,...,m}$.

From $S$’s perspective, the misreporting of types $\{\theta_i\}_{i=1,2,...,m}$ as types $\{\theta_i\}_{i=1,2,...,m}$ is not profitable because $S$ is going to receive a payoff equal to zero anyway. Moreover $S$ strictly prefers not to misreport types $\{\theta_i\}_{i=1,2,...,m}$ as types $\{\theta_i\}_{i=m+1,...,n}$ because this implies a negative payoff. The only profitable manipulation available to $S$ is to misreport types $\{\theta_i\}_{i=m+1,...,n}$ as $\{\theta_i\}_{i=1,2,...,m}$ but this involves paying a positive bribe to $A$ when the partition $\{\theta_m, \theta_{m+1}, ..., \theta_n\}$ is realized. In order to be certain of convincing all types $\{\theta_i\}_{i=m+1,...,n}$ to misreport as types $\{\theta_i\}_{i=1,2,...,m}$ $S$ has to pay a bribe\(^{16}\) at least equal to $(\theta_m - \theta_{m-1})q_{1m} - (\theta_n - \theta_{m-1})q_{1m-1}$, which leaves $S$ indifferent between proposing this bribe to $A$ or paying the negative transfer to $P$. Therefore when $S$ observes the partition $\{\theta_{m-1}, ..., \theta_{n-1}, \theta_n\}$, $S$ refuses the grand-contract, i.e. the participation constraint for $S$ doesn’t hold when the partition $\{\theta_{m-1}, ..., \theta_{n-1}, \theta_n\}$ is observed. On the other hand if $S$ observes $\{\theta_1, \theta_2, ..., \theta_{m-1}\}$ she accepts the grand-contract. Accordingly, $P$ updates his beliefs on the

\(^{16}\)Notice that this bribe is sufficient to convince type $\theta_n$ to misreport as type $\theta_m$ which implies that all the other types $\{\theta_i\}_{i=(m+1),...,(n-1)}$ are also willing to accept this manipulation.
agent’s type based on $S$’s acceptance or refusal. Conditional probabilities become

\[
p(\theta_j | a_1) = \begin{cases} 
0 & \text{for all } j \in \{m-1, \ldots, n-1, n\} \\
f(\theta_j) & \text{for all } j \in \{1, 2, \ldots, m\} \\
\frac{f(\theta_j)}{\sum_{i=m-1}^{n} f(\theta_i)} & \text{for all } j \in \{m-1, \ldots, n-1, n\} \\
0 & \text{for all } j \in \{1, 2, \ldots, m\}
\end{cases}
\]

It follows that $P$ maximizes the following expected utility function

\[
p(\theta_1) [W(q_{21}) - \theta_1 q_{21} - V_{11}] + \\
+ p(\theta_2) [W(q_{22}) - \theta_2 q_{22} - V_{12}] + \\
+ \ldots \\
+ p(\theta_{m-1}) [W(q_{1m-1}) - \theta_3 q_{1m-1} - V_{2m+1}] + \\
+ \ldots \\
+ p(\theta_n) [W(q_{1n}) - \theta_3 q_{1n} - V_{2n}].
\]

Having this schedule in place, the output levels $q_{1m-1}, q_{1m-2}, \ldots, q_{1n}$ and $q_{21}, q_{22}, \ldots, q_{2m}$ are realized with probability zero because the correspondent strategies are off the equilibrium path. Therefore, it is possible to consider only the remaining output levels. The optimal set of relevant output levels is determined maximizing (32) with respect to $q_{2m-1}, q_{2m-2}, \ldots, q_{2n}$ and $q_{11}, q_{12}, \ldots, q_{1m}$ subject to

\[q_{11} \geq q_{12} \geq \ldots \geq q_{1m}\]

\[q_{2m+1} \geq q_{2m+2} \geq \ldots \geq q_{2n}\]

A fast inspection reveals that:
\(V_{2n} = U_{n}^{cf} = 0,\)

\(V_{2n-1} = U_{n-1}^{cf} = (\theta_n - \theta_{n-1})q_n,\)

\(\ldots\)

\(V_{2m+1} = U_{m+1}^{cf} = \sum_{j=m-2}^{n} U_{j}^{cf} + (\theta_{m-2} - \theta_{m-1})q_{m-2}\)

\(V_{1m} = U_{m}^{cf} = 0,\)

\(V_{1m-1} = U_{m-1}^{cf} = (\theta_m - \theta_{m-1})q_m,\)

\(\ldots\)

\(V_{11} = U_{1}^{cf} = \sum_{j=2}^{m} U_{j}^{cf} + (\theta_2 - \theta_1)q_2.\)

It follows that the mechanism implements the same expected payoff for \(P\) as the optimal collusion-free one, provided that the set of optimal output levels under collusion-free supervision is weakly decreasing.