Collusion and Selective Supervision*

Alberto Motta†

November 2008

Abstract

This paper studies the role of a policy of inducing *selective supervision* in combating collusion within organizations, or in regulatory setups. In a mechanism-design problem involving a principal-supervisor-agent we show the role of endogenous selection of supervisory activity by the principal. One simple example is a mechanism in which the agent bypasses the supervisor and contracts directly with the principal in some states of the world. If collusion between supervisor and agent can occur only after they have decided to participate in the mechanism, this can costlessly eliminate collusion. This result is robust to alternative information structures, collusive behaviors and specification of agent's types. Applications include self-reporting of crimes, tax amnesties, immigration amnesties, work contracts specifying different degrees of discretion, mechanisms based on recommendation letters, embassies issuing immigration permits, and hiring committees.

Key Words: Collusion, supervision, delegation, mechanism design.

JEL Classification: D82, C72, L51.

*This paper was written during my visit at Boston University. I wish to express my thanks to the Department of Economics for its hospitality and support. I am especially grateful to my advisors Antonio Nicolò and Dilip Mookherjee for their support, comments and encouragement. I also thank David Martimort, Gorkem Celik, Vianney Dequiedt, Gregory Pavlov, Alfredo Burlando, Patrick Rey, Barton Lipman, Michael Manove and Stefan Napel for helpful discussions. This paper has benefited with comments from seminar participants at University Toulouse 1, Boston University and University of Padua. This contribution is related to other two papers belonging to my Ph.D. thesis, see Burlando and Motta (2008a, 2008b).

†University of Padua, Department of Economic Sciences, 33 via del Santo Padua, Italy, E-mail: alberto.motta@unipd.it .

1 Introduction

Within economic organizations, third-party supervision is commonly observed: owners of a firm usually delegate the responsibility for supervising production to top managers, stockholders rely on auditors to acquire information about management conduct, managers ask employees to report on the performance of coworkers, Governments make use of agencies to regulate firms, auditors to examine tax returns, and inspectors to detect illegal immigration.

The need for supervisory activity originates in an information asymmetry between the residual claimant of a productive activity (the *principal*) and who actually carries out the productive activity (the *agent*). The role of the *supervisor*[1] is to provide the principal with information concerning actions or characteristics of the agent. This creates a potential for collusion between supervisor and agent, wherein the agent bribes the supervisor to conceal information from the principal.[2] Most authors conclude that collusion is a problem, and eliminating it is costly to the principal. In this context, the role of collusion in limiting the scope for incentives, value of hiring supervisors, and delegation to supervisors have been examined by many authors in a standard framework that includes asymmetric information between the colluding parties, and the inability to collude prior to making a decision to participate in the mechanism.

This paper focuses on a tool for combating collusion that has been previously overlooked. This tool is based on the idea of *selective supervision*, where the supervisor may not be engaged by the principal in certain states of the world. Take, for example, a simple mechanism where the agent selects between a regime with supervision and a regime without it. In some states of the world, the agent can opt out of being supervised and contract directly with the principal. This reveals useful information to the principal and reduces the scope of collusion. In the standard framework we show that it costlessly eliminates collusion.

There are many real-world examples of such *selective supervision* mechanism. One of them is self-reporting of illegal acts, wherein offenders can choose to report their illegal acts directly to principal by choosing a mechanism that bypasses the supervisor. The literature on law enforcement has long highlighted that self-reporting allows the government to save money by reducing enforcement costs.[3] This paper tackles the issue from a different angle, suggesting a new and different advantage to the use of self-reporting: namely, the reduction of the costs associated with the threat of collusion. Another example is that of tax amnesties where the agent is induced to report his type directly to the principal, bypass-

[1] We refer to the supervisor and the agent respectively as *she* and *he*.

[2] See for example Tirole (1986), Laffont and Tirole (1991,1993), Lambert-Mogilianksy, (1998), Faure-Grimaud and Martimort (2001), Faure-Grimaud, Laffont and Martimort (2003) and Celik (2008).

[3] See Kaplow and Shavell (1994), Innes (1999) and Innes (2001).

ing the supervisor's inspection. The same applies to immigration amnesties.[4] Finally the strategic tool presented in this paper may shed light on the formation of hierarchy structures within the firms, where the scope and intensity of supervision varies depending on the agent's characteristics.

Generally speaking, the precise implementation of *selective supervision* depends on both the nature and the timing of the supervisor's information. First, consider the case where the information is binary. If the supervisor receives this information after the acceptance of the principal's offer, the example of the agent choosing between a regime with or without supervision applies.[5] On the other hand, if the supervisor receives her information before the acceptance decision[6] and the agent discovers his type only later on, then a different mechanism is required. In this case, the principal offers a mechanism in which the supervisor can opt out in some states of the world. An example fitting this case is that of an advisor who is asked to write a recommendation letter for a student who is not fully aware of his skills related to the job he intends to apply for. The advisor may refuse to write the letter, revealing some information about the agent's type. By the same token, foreigner embassies have the discretion to refuse immigration permits to applicants that they do not consider suitable for admission. Similarly, hiring committees may refuse to offer interviews to certain candidates. Failure to receive interviews signal a portion of the private information available to the committees. In all these cases, the supervisor's decision to opt out conveys information about the applicants' characteristics. This happens in the standard framework as well - but there the Revelation Principle applies, and there exists an equivalent mechanism where the supervisor always participates. In our setting, the Revelation Principle does not apply, owing to the presence of collusion.

When the nature of the supervisor's information is not binary, the implementation of *selective supervision* becomes more nuanced. In this case, the principal proposes a menu

[4] Some of these issues are explored in a separate paper by Burlando and Motta (2008a). They analyze the impact of self reporting on law enforcement when officers are corruptible. They show that a budget-constrained government may prefer an enforcement system based on corruption rather than one based on legal fines. They conclude that the government can use self reporting as a way to clean up corrupt enforcement agencies. Unlike our paper, they do not adopt a mechanism design approach. Moreover our contribution considers a larger class of mechanisms, wherein self-reporting with binary information structure for the supervisor is only one simple application.

[5] Burlando and Motta (2008b) consider a framework with two types of agents and limited liability for the supervisor. Unlike this paper, they consider hard-information supervision where with some probability the supervisor learns the true agent's type, otherwise she learns nothing. They show that there exists a mechanism that eliminates agency costs by providing the productive agent with the possibility of avoiding inspection. When the productive agent is risk averse, the mechanism also provides him with an insurance coverage: as a consequence, this mechanism would be worthwhile even abstracting from collusion.

[6] In this case, the supervisor is an "informed third party" or a "witness" who happened to learn some information about the agent even before the principal had a chance to offer her a grand-contract.

of grand-contracts. Each grand-contract specifies a different scope for supervisory activity, where the scope of supervision refers to the dimension of the message space available to the supervisor. Applications include self-reporting schemes limited to certain crimes (i.e., the supervisor can still report on some issues but not all of them), letters of recommendation with different degrees of approbation, restricted visa permits, tax amnesties for specific types of evasions, or work contracts subject to different degrees of discretion.

The results presented in this paper are robust to many aspects of collusive behavior. First, collusion-proof implementation does not rest on special assumptions about the accuracy of the supervisor's information: the principal can attain the second-best payoff even when there is no residual asymmetric information between the supervisor and the agent. Second, collusion-proof implementation does not require any restrictions on the allocation of bargaining power inside the coalition. Third, the result does not depend on the identity of the coalition member who offers and initiates the collusive agreement. The mechanism proposed in this paper holds for a quite generic specification of agent type and does not rely on special assumptions about players' utility functions. In order to further test the robustness of the mechanism, its implementation is analyzed in both the information frameworks proposed by Celik (2008) and Faure-Grimaud, Laffont and Martimort (2003) (FLM, hereafter.) The first one considers an information structure for the supervisor that can be represented as a connected partition of the agent's type space. The second information environment entails the supervisor observing an informative signal. Unlike the outcome implemented here, the optimal outcomes identified by both Celik (2008) and FLM (2003) fail to achieve the same expected payoff as the optimal collusion-free outcome.[7] This is due to the fact that they restrict attention to direct revelation mechanisms with full participation. The Revelation Principle is usually invoked to support the idea that full participation is without loss of generality. However, the assumption of *no collusion in participation decisions* implies a solution concept that is no longer compatible with the very foundation of the Revelation Principle. These issues are explained in more detail in Section 3.

The concept of *selective supervision* shares some similarities with the mechanism proposed by Dequiedt (2006) and Celik and Peters (2008b). The latter study an example of a mechanism-design problem where the players can coordinate their actions in a default game. They show that some allocation rules are implementable only with mechanisms which will be rejected on the equilibrium path. It may be useful to re-label aspects of their framework to highlight the similarities with the environment considered here. The two players in their model can be thought of as the supervisor and the agent in our framework. The default game corresponds to the principal's mechanism in this paper, whereas the coordination-

[7]FLM's (2003) mechanism implements the second-best outcome when the supervisor is risk neutral.

mechanism corresponds to the collusive side-contract between the agent and the supervisor. Consequently, the scope of the present contribution goes beyond the one proposed by Celik and Peters (2008b) in that it considers endogenously determined "default" games. Even though their setting is different from the one considered here, there is one aspect which is common to both contributions: participation decisions convey information about the types of the players. Dequiedt (2006) considers a similar point in the mechanism design literature that assumes that each agent has a veto power.

In a related paper, Che and Kim (2006) study a general collusion setup where agents cannot collude prior to making their decision to participate in the mechanism. They conclude that the second-best payoff is implementable when players are risk neutral. Given the restrictions they impose on the correlation of information of the colluding parties, their result does not apply to the setup considered in this paper. Their implementation strategy would not work in our setting because the information structure for the supervisor in our model is far more complex than the one they adopt. For example, when the supervisor's information structure is represented by a connected partition of the agent's type space, Che's and Kim's strategy of "selling the firm to the coalition" does not work because the supervisor would definitely refuse to become the residual claimant of production in some states of the world. On the contrary, it remains an open and intriguing question whether or not the strategy proposed in our framework applies to no-supervision setups such as the one they proposed.

The assumption of no collusion in participation decisions is plausible in many realistic situations: oftentimes, the agent and the supervisor are matched together after they have decided to participate. Under these circumstances, their failure to coordinate participation decisions is due to the impossibility of signing a preemptive side-contract with all eligible supervisors. The principal could also decide to hire the supervisor after the agent has made his participation decision. In some cases, the use of job rotation for supervisors achieves the same result. In some other cases, the principal can avoid the disclosure of the agent's identity at the participation stage: this precaution makes it difficult for the supervisor to collude since she faces a potentially vast population of eligible agents.

Nonetheless, the fact that in our setting collusion can be costlessly eliminated, is in contrast to the persistence of real-world collusion. This seems to suggest that allowing for some degree of coordination in participation decisions may be realistic. We suspect that *selective supervision* is useful in a setting where collusion occurs prior to participation, though in that context it is unlikely to costlessly eliminate collusion. This issue remains to be explored in future research. A few interesting papers have already studied the implementation of collusion-proof mechanisms when agents can collude on their participation decisions,[8] but

[8]Pavlov (2008), Che and Kim (2008) and Dequiedt (2007) consider auctions where bidders collude prior to

none of them have addressed this question yet. Among them, Mookherjee and Tsumagari (2004) analyze this problem in a supervision setup.[9] However they focus on a different question with respect to one analyzed in our contribution. Namely, they consider two productive agents and explore the possibility that collusion may rationalize delegation to intermediaries uninvolved in production. In particular, they do not focus on the identification of the optimal mechanism in the presence of collusion.

The remainder of the paper is organized as follows. Section 2 proposes the baseline model. Section 3 presents two illustrative examples. Section 4 introduces the general model and the notion of collusion-proof implementation. Section 5 provides some extensions and additional comments. Section 6 concludes. All proofs are given in the Appendix.

2 The Baseline Model

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

$$t - \theta q,$$

where t denotes the transfer he receives from the principal (P), q is the output level and θ 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 θ. The supervisor's (S) information structure is the partition $\{\{\theta_1, \theta_2\}, \theta_3\}$ of A's type space θ. It follows that S is able to tell whether A is the least efficient type θ_3 or not, but she can't distinguish types θ_2 and θ_1. In the baseline model S is assumed to be risk neutral. 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 θ. P can commit to a

participating. Che and Kim (2006) study an optimal collusion-proof auction in an environment where subsets of bidders may collude not just on their bids but also on their participation. They find that informational asymmetry facing the potential colluders can be significantly exploited to reduce their possibility to collude. Dequiedt (2007) considers two bidders with binary types. He finds that the seller can, at most, collect her reserve price when a bidder's valuation exceeds that price, if and only if a cartel can commit to certain punishment. Pavlov (2008) independently studies a problem similar to Che and Kim (2008) and reaches similar conclusions. Quesada (2004) studies collusion initiated by an informed party under asymmetric information.

[9] Mookherjee (2006) provides an excellent survey of this strand of the literature.

grand-contract, consisting in a triple

$$\Gamma = \{q(m_s, m_a), t(m_s, m_a), s(m_s, m_a)\}.$$

This grand-contract defines the outcome and the monetary transfer respectively for A and S as a function of S's and A's messages, which are denoted as m_s and m_a and belong respectively to the message spaces M_s and M_a. If the grand-contract is rejected, the game ends with zero production and no monetary transfer to the players. In other words, the outside option is normalized to zero for both S and A.

2.1 Collusion-free Supervision

Consider the benchmark case where collusion between A and S is not allowed. The timing of the game is as follows:

- A learns θ. S learns the partition $\{\{\theta_1, \theta_2\}, \theta_3\}$.
- P offers a grand-contract to S and A.
- S and A accept or refuse the grand-contract. If A refuses, the game ends. Otherwise production and transfers take place as specified in the grand-contract.

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 θ_i; U_i represents the utility level for A. In what follows, $\{q_i, V_i, U_i\}_{i\in\{1,2,3\}}$ is referred to as an **outcome**. Under collusion-free supervision P can 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

$$\begin{aligned} V_3^{cf} &= U_3^{cf} = 0, \\ V_2^{cf} &= U_2^{cf} = 0, \\ V_1^{cf} &= U_1^{cf} = (\theta_2 - \theta_1)q_2. \end{aligned} \tag{1}$$

where $q_1 \geq q_2$.

As standard in this kind of design problem, the information rent to be paid to type θ_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 θ_2 nor type θ_1 can misreport their types as θ_3. Given that type θ_2 is not willing to mimic type θ_1, P is able to leave zero rent to type θ_2 as well as to type θ_3. Yet, it is still profitable for type θ_1 to imitate type θ_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, if S can distinguish type θ_3 from the other types, implementability requires the output profile to be monotonic only with respect to types θ_1 and θ_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

$$\left\{q_i^{cf}\right\}_{i\in\{1,2,3\}} \in \arg\max \sum_{i=1}^{3} f(\theta_i)\left[W(q_i) - \theta_i q_i - V_i^{cf}\right],$$
$$s.t. \qquad q_1 \geq q_2.$$

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 θ_2 can increase his expected payoff by misreporting as type θ_3. In order to do so, A may need 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 side-contract that asks for a bribe from A for misreporting him as type θ_3.

3 Two Illustrative Examples

In this section we propose some simple mechanisms based on *selective supervision.* These mechanisms implement the collusion-free outcome even when the supervisor and the agent are allowed to collude. In order to further test the robustness of these mechanisms, their implementation is analyzed in two different information frameworks. The first one considers the information structure for the supervisor presented in our baseline model. Celik (2008) analyzes different organizational responses in this setting.[10]

Contracting with both the supervisor and the agent through a grand-contract constitutes the most general organizational design for the principal, which is usually denoted as *centralized contracting* in the literature. Special cases of this design would be respectively, the principal contracting only with the supervisor, delegating authority to her over contracting

[10]Faure-Grimaud and Martimort (2001) also analyse delegation within an identical information structure. Differently from this paper, they assume that the highest cost agent is so inefficient that it is never optimal for him to produce.

with the agent, and the principal contracting with the agent directly, ignoring the presence of the supervisor. The former organizational response is denoted as *decentralized contracting,* while the latter is denoted by *no-supervision contracting.* 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 we consider entails two possible agent types and the supervisor observing an informative signal which also takes two possible values. FLM (2003) analyze *centralized* and *decentralized contracting* in this framework, concluding that decentralization is always equivalent to centralization.

Unlike the outcome implemented by the mechanisms presented here, the optimal outcomes identified by both Celik (2008) and FLM (2003) fail to achieve the same expected payoff as the optimal collusion-free outcome. This is due to the fact that Celik (2008) and FLM (2003) restrict their attention to direct revelation mechanisms with full participation. Focusing on this class of mechanisms is not without loss of generality under the assumption of *no collusion in participation decisions.* In fact, this assumption imports some kind of sequential rationality in the mechanism design; the agent and the supervisor may collude, but only for a given participation decision. Therefore, the relevant solution concept is no longer compatible with the very foundation of the Revelation Principle,[11] which is usually invoked to support the idea that full participation is without loss of generality.

3.1 Celik's (2008) Model

This section proposes a couple of simple mechanisms Γ and Γ_0, which implement the collusion-free outcome. The process of collusion is formalized by assuming that S and A can stipulate a side-contract, after the acceptance of the grand-contract by both parties. The side-contract is a pair $SC = \{\widehat{c}(.), b(.)\}$ 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.[12]

Γ illustrates an example of *selective supervision* where the principal offers a mechanism in which the supervisor can opt out in some states of the world. The design structure is as follows: after the grand-contract is proposed by P, S and A simultaneously make their acceptance decisions. 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

[11] The Revelation Principle applies to solution concepts, such as Bayesian equilibria, with no *sequential rationality* restrictions on the *out-of-equilibrium* path events.

[12] Relaxation of the enforceability assumption is considered by Martimort (1999), Abdulkadiroglu and Chung (2003), and Khalil and Lawarree (2006).

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."*[13] For the sake of exposition, denote by s_{ij} (respectively t_{ij}, q_{ij}, V_{ij} and U_{ij}) S's wage (respectively 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 θ_j. Consider the following grand-contract:

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

and

$$\begin{aligned} t_{23} &= \theta_3 q_{23}, \\ t_{22} &= 0, \\ t_{21} &= 0, \end{aligned}$$

where

$$\begin{aligned} q_{13} &= 0 & q_{23} &= q_3^{cf}, \\ q_{12} &= q_2^{cf} & q_{22} &= 0, \\ q_{11} &= q_1^{cf} & q_{21} &= 0. \end{aligned}$$

The following proposition encapsulates the first result of this paper.

Proposition 1 Γ *allows P to achieve the same expected payoff as the optimal collusion-free outcome.*

Proof. See Appendix. ■

The complete proof is relegated in the Appendix. Here, a simple sketch of the intuition is offered. To begin with, recall that collusion takes place after the acceptance of the grand-contract by both parties. Therefore the threat of coalition formation arises only in the case where S accepts the grand-contract, i.e., whenever she reports a_1 to P. Under these conditions, A and S cannot find any profitable collective manipulation. Suppose that S accepts the grand-contract; the collusive coalition clearly has no stake in misreporting types θ_2 and θ_1. On the other hand, S would like to misreport type θ_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 θ_3 to misreport his type as θ_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

[13] In principle, the message space for A should also contain the elements $\{a_1, a_2\}$. If A refuses the grand-contract the game ends with zero production and no monetary transfers.

type θ_3 to mimic type θ_1. Accordingly, if A and S accept the grand-contract they shall respond to it in a non-cooperative fashion.[14] A crucial aspect of this mechanism is related to S's participation decision. Notice that the participation constraint for S holds when she observes the partition $\{\theta_2, \theta_1\}$, whereas it doesn't hold when the partition $\{\theta_3\}$ is observed. Given that S's refusal of the grand contract indicates that the realized type is θ_3, the only relevant constraints when S refuses the grand-contract are the participation and the incentive compatibility constraints for type θ_3. This extra information, conveyed by S's participation decision, is acquired by P without forgoing any rents to S. Similarly, only the participation and the incentive compatibility constraints for types θ_2 and θ_1 must hold when S accepts the grand contract. One final remark. Recall that under collusion-free supervision there is no monotonicity requirement regarding two output levels in separate partition cells. This propriety is present in Γ. Indeed, the only monotonicity requirement regards output levels q_{11} and q_{12}.

Γ is not the only mechanism that allows the principal to implement the collusion-free outcome. For example, Γ_0 illustrates a case of *selective supervision* second-best-mechanism, wherein the agent has the possibility to choose between a regime with or without supervision. P offers a menu of two grand-contracts and A selects one of them. S chooses whether to participate in the mechanism or not. S and A simultaneously make their acceptance decisions, which are not subject to collusion. Following A's selection of the grand-contract, A and S must respond to the selected grand-contract. The message spaces for S and A are respectively $M_s \in \{\tau_1, \tau_3\}$ and $M_a \in \{\theta_1, \theta_2, \theta_3\}$, where τ_1 and τ_3 correspond respectively to the partitions $\{\theta_1, \theta_2\}$ and $\{\theta_3\}$ observed by the supervisor.

In the first stage of the game, A selects the grand-contract he wants to take part in. In order to do so, he sends a message from the message space $M_{\overline{a}} \in \{a_1, a_2\}$, where a_1 stands for *"supervision regime"* and a_2 indicates *"no-supervision regime."* Denote by s_{kij} (respectively t_{kij}, q_{kij}, V_{kij} and U_{kij}) S's wage (respectively A's transfer, the output target, the coalition information rent and A's utility) when A selects a_k and reports that he has type θ_j and S reports τ_i. Consider the following grand-contract:

$$
\begin{aligned}
t_{133} &= \theta_3 q_{133}, & s_{133} &= 0, \\
t_{112} &= \theta_2 q_{112}, & s_{112} &= (\theta_3 - \theta_2)\, q_{133}, \\
t_{111} &= (\theta_2 - \theta_1) q_{112} + \theta_1 q_{111}, & s_{111} &= (\theta_3 - \theta_2)\, q_{133},
\end{aligned}
$$

[14] In other words, A sends his message non-cooperatively whenever the side-contract fails to be established.

and

$$t_{23} = 0,$$
$$t_{22} = \theta_2 q_{22},$$
$$t_{21} = (\theta_2 - \theta_1) q_{22} + \theta_1 q_{21},$$

where

$$q_{133} = q_3^{cf}, \qquad q_{23} = 0,$$
$$q_{112} = q_2^{cf}, \qquad q_{22} = q_2^{cf},$$
$$q_{111} = q_1^{cf}, \qquad q_{21} = q_1^{cf}.$$
$$q_{113} = 0,$$
$$q_{132} = 0,$$
$$q_{131} = 0,$$

Notice that Γ_0 allows P to achieve the same expected payoff as the optimal collusion-free outcome. The formal demonstration is analogous to the proof of Proposition 1 and is omitted. Intuitively, when A selects a regime subject to supervision he keeps in mind that the information rent arising from the fact that P cannot observe S's signal is entirely captured by the latter. Moreover, S is always induced to report truthfully in a way that is robust to any possible side-contract available to the coalition. In other words, A obtains no extra-profit from the asymmetric information between P and S. Consequently, A is willing to self-report S information in the first stage of the game because self-reporting this information has no effect whatsoever on his payoff. In equilibrium, the most inefficient type opts for a regime subject to supervision, while the medium and the most efficient types select the no-supervision regime.

3.2 FLM's (2003) Model

In this section an alternative information structure is considered. Following FLM (2003), the unitary cost of production, θ, 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 θ is A's private information. S is uninformed about the agent's type. Nonetheless, he receives a signal τ on the agent's cost. τ is drawn from a discrete distribution on $T = \{\tau_1, \tau_2\}$. The joint probabilities on (θ_j, τ_i) are defined as $p_{ij} = Prob(\tau = \tau_i, \theta = \theta_j)$ with $p_{ij} > 0$ for all i, j. From the joint distribution above, one can derive the conditional probabilities $p(\theta_j|\tau_i)$. 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 her monetary payoff x: $U_s = \frac{1}{r}(1 - e^{-rx})$. All other assumptions are unchanged. The timing of the game is as follows:

- A learns θ and τ, while S learns only τ.
- P offers a grand-contract to S and A.
- S and A accept or refuse the grand-contract. 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-contract, S may offer a collusive side-contract to A who accepts or refuses this side-contract. If A refuses, the mechanism is played non-cooperatively by S and A.
- Production and transfers take place.

As before, the message spaces for S is $M_s = \{a_1, a_2\}$, where a_1 stands for *"accept the grand contract"* and a_2 indicates *"refuse the grand contract."* A's message space is given by $M_a = \{\theta_1, \theta_2\}$. Denote by s_{ij} (respectively t_{ij}, q_{ij}, V_{ij} and U_{ij}) S's wage (respectively 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 θ_j. Consider a contract Γ_1 involving the following:

$$\begin{array}{ll} t_{12} = \theta_2 q_{12} & s_{12} = -\frac{p(\theta_1|\tau_1)}{p(\theta_2|\tau_1)}\varepsilon, \\ t_{11} = \Delta\theta q_{12} + \theta_1 q_{11} & s_{11} = \varepsilon, \end{array}$$

and

$$\begin{array}{l} t_{22} = \theta_2 q_{22}, \\ t_{21} = \Delta\theta q_{22} + \theta_1 q_{21}, \end{array}$$

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

Proposition 2 Γ_1 *allows P to achieve the same expected payoff as the optimal collusion-free outcome.*

Proof. See Appendix. ■

The intuition behind Γ_1 implementation is very simple. As in the previous example, no collusion in participation decisions implies the possibility of coalition formation only if S reports a_1 to P. A fast inspection reveals that GC_1^{sb} is not subject to collusive coalition: if ε is small enough, S and A cannot find any profitable collective manipulation to play along with. Suppose S accepts the grand-contract. In this case S would like to induce type θ_2 to report he has type θ_1 in the attempt to avoid the negative transfer s_{12}. For small values of ε, S strictly prefers paying s_{12} to P rather than offering a bribe $\Delta\theta\,(q_{12} - q_{11})$ to type θ_2

for misreporting his type as θ_1. It is easy to notice that when τ_2 (τ_1) is realized, S refuses (accepts) the grand-contract. Therefore, following S's participation decision, P extracts S's information with no extra costs. This allows P to achieve the same expected payoff as the *optimal collusion-free outcome*: all the costs associated with collusion are fully eliminated.

3.3 An Extension of Celik (2008)

The illustrative examples presented in this section rely on two special assumptions. First, S is not indispensable for production. Second, S's information must be binary. Before analysing the general model, it may be useful to sketch the intuition by considering an extension of Celik (2008). Let Θ be the set of possibile agent's types with finite cardinality $n \geq 2$, i.e. $\Theta = \{\theta_1, ..., \theta_n\}$. S's information structure is the connected partition Ω of A's type space Θ. Denote with m the number of possible partitions with $1 \leq m \leq n$. Note that S is able to tell whether A's type belongs to a certain partition or not, but she can't distinguish types inside each partition. All the other assumptions remain the same. Having this schedule in place, the design problem is similar to the one analyzed before.

Proposition 3 *Consider any pair* (Θ, Ω). *There exists a mechanism that implements the optimal collusion-free outcome.*

Proof. See Appendix. ■

This result will be clearer after reconsidering the first illustrative example. To this purpose, a different interpretation must be attributed to S's message space. Consider a mechanism where P offers a menu of grand-contracts and S can accept only one of them: in the baseline example this implies that a_1 would stand for *"accept the grand-contract 1"* and a_2 would indicate *"accept the grand-contract 2,"* where accepting the *grand-contract 1* implicitly comports refusing the *grand-contract 2* and vice versa. The resulting mechanism is similar to Γ, provided that the *grand-contract 2* must be incentive-compatible for all A's types[15] and specify the following payoffs for S, $s_{23} = 0$ and $s_{22} = s_{21} = -\varepsilon$. If S and A simultaneously make their acceptance decisions (i.e., S chooses which grand-contract to accept and A chooses whether to participate or not in the mechanism) and collusion is not allowed at this stage, the outcome is identical to the one analysed in Proposition 1.

The intuition behind this mechanism also applies to the case with a generic number of A's types and a generic number of connected partitions; this requires a menu that contains a number of grand-contracts equal to the number of partitions observable by S. In this case

[15] Accordingly the transfer's levels are given by $t_{23} = \theta_3 q_{23}$, $t_{22} = (\theta_3 - \theta_2) q_{23} + \theta_2 q_{22} + \varepsilon$ and $t_{21} = (\theta_3 - \theta_2) q_{23} + (\theta_2 - \theta_1) q_{22} + \theta_1 q_{21} + \varepsilon$.

S's participation decision becomes a proxy for sending a report. P can always induce S to choose the grand-contract which corresponds to the partition he truly observed, by offering grand-contracts that leave S with a negative payoff in all other cases.

4 The General Model

In order to generalize the result, the current section advances on this front by proposing a setting with a generic number of A's types and S's information sets. This setting accommodates as special cases both Celik's (2008) and FLM's (2003) frameworks. The general model has the following features: The unitary cost of production, θ, takes n possible values from the set $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$, where $\theta_n \geq \theta_{n-1} \geq .... \geq \theta_1 \geq 0$. S receives a signal τ on the agent's cost. τ is drawn from a discrete distribution on $T = \{\tau_1, \tau_2, ..., \tau_n\}$. The joint probabilities on (θ_j, τ_i) are defined as $p_{ij} = Prob(\tau = \tau_i, \theta = \theta_j)$ with $p_{ij} \geq 0$ for all i, j and $\sum_{j=1}^{n} p_{ij} = 1$ for all i. From the joint distribution above, one can derive the conditional probabilities $p(\theta_j|\tau_i)$. There is a positive correlation between signals and types when the monotone likelihood ratio property is satisfied, i.e., $p(\theta_i'|\tau_i')p(\theta_i|\tau_i) - p(\theta_i'|\tau_i)p(\theta_i|\tau_i') \geq 0$ for all $\tau_i' \geq \tau_i$ and $\theta_i' \geq \theta_i$.[16] Finally, A and S can be either risk neutral or risk averse; results are not affected by this characteristic of the players.[17] All the other assumptions remain the same. The precise implementation of *selective supervision* depends crucially on the timing of the supervisor's information. Let us denote *Timing 1* the framework in which the supervisor receives her information before the acceptance decision and the agent discovers his type only later on. In this case, the supervisor is an *informed third party* or a *witness* who happened to learn some information about the agent even before the principal had a chance to offer her a grand-contract. On the other hand, let us denote *Timing 2* the setting in which the supervisor receives this information after the acceptance of the principal's offer.

4.1 Timing 1

In rest of this section, the following implementation strategy is considered. P offers a menu of mechanisms to S who has to select one of them.[18] S's choice unilaterally determines which mechanism is played. A is not allowed to select a specific mechanism; he can simply choose whether to participate in the whole menu of mechanisms or not. S and A simultaneously

[16] It is straightforward to see that this framework reduces to FLM (2003) when $p_{ij} > 0$ and $n = 2$. Moreover, it also includes Celik (2008) as a special case. If $n = 3$, $p_{31} = p_{32} = 0$, $p_{33} = 1$, $p_{i3} = 0$ for all $i \in \{1, 2\}$, $p_{11} = p_{21}$, $p_{12} = p_{22}$, the information structure for S reduces to the standard connected partition case, $\{\{\theta_1, \theta_2\}, \theta_3\}$. One aspect is worth noticing at this point. The general model presented here not only allows for a generic number of types and partitions; it further extends Celik (2008) by allowing S's information to be represented by both connected and non-connected partitions.

[17] For simplicity, A is assumed to be risk neutral and S risk averse.

[18] S could also choose not to accept any mechanism.

make their acceptance decisions, which are not subject to collusion. Following S and A participation decisions, A reports his type. The timing of the game is as follows,

- At date -1, S learns τ.
- At date 0, P offers a menu of mechanisms to S and A.
- At date 1, S has to accept one of these mechanisms. A accepts or refuses the whole menu of mechanisms.
- At date 2, A learns θ and τ.
- At date 3, S and A can stipulate a side-contract.
- At date 4, A reports his type.
- At date 5, production and transfers take place.

Each mechanism offered by P in the participation stage corresponds to a grand-contract. Throughout the whole game, the message spaces for S and A are respectively $M_s \in \{a_1, a_2, ..., a_n\}$ and $M_a \in \{\theta_1, \theta_2, ..., \theta_n\}$, where a_i stands for *"accept the grand-contract i."* For the sake of notation simplicity, denote by s_{ij} (respectively t_{ij}, q_{ij}, V_{ij} and U_{ij}) S's wage (respectively 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 θ_j. Having this schedule in place, the next section illustrates the concept of collusion-proofness.

4.1.1 Collusion-proof Implementation

At this stage, it is useful to formalize the concept of collusion-proofness. Recall that S and A can stipulate a side-contract, after the acceptance of the grand-contract by both parties. The side-contract is a pair $SC = \{\widehat{c}(.), b(.)\}$ 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. This side-contract is assumed to be enforceable. If A or S refuse the side-contract, the game is played non-cooperatively. Consider a generic i-grand-contract, which corresponds to the message a_i. Let $\widehat{c}(\theta_j)$ denotes the misreport of type θ_j as type $\theta_{\widehat{c}(\theta_j)}$, where $\widehat{c} : \{1, 2, ..., n\} \to \{1, 2, ..., n\}$. Under this manipulation of players' reports, the coalition information rent is $V_{i\widehat{c}(\theta_j)} + (\theta_{i\widehat{c}(\theta_j)} - \theta_j)q_{i\widehat{c}(\theta_j)}$. Consider the outcome $\{q_{ij}, V_{ij}, U_{ij}\}_{j\in\{1,2,...,n\}}$. In order for such an outcome to be implementable, the value of V_{ij} must be weakly larger than the value of $V_{i\widehat{c}(\theta_j)} + (\theta_{i\widehat{c}(\theta_j)} - \theta_j)q_{i\widehat{c}(\theta_j)}$ under any possible manipulation available to A and S. What follows is the condition for the feasibility of the outcome $\{q_{ij}, V_{ij}, U_{ij}\}$:

$$\{\theta_j, V_{ij}\}_{j\in\{1,2,...,n\}} \in \arg \max_{\{\widehat{c}(\theta_j)\}_{j\in\{1,2,...,n\}}} V_{i\widehat{c}(\theta_j)} + (\theta_{\widehat{c}(\theta_j)} - \theta_j)q_{i\widehat{c}(\theta_i)}. \tag{2}$$

Condition (2) is defined as **"strong collusion feasibility condition"** which guarantees that for a given i-grand-contract the players are unable to find a profitable manipulation. This choice of words follows the terminology of Celik (2008) who define the property of "collusion feasibility condition" to indicate the set of outcomes such that S is unwilling to offer a side contract that misreports A's type. The notion of collusion-proofness considered in this paper is stronger in that it requires no collective gain from any possible type of misreport. In other words, collusion-proofness must hold true regardless of (i) the allocation of bargaining power inside the coalition, (ii) the residual asymmetric information between S and A, and (iii) the identity of the coalition member who offers and initiates the collusive agreement.

Up to this point, we considered collusion feasibility for a given i-grand-contract. S's participation decision must also be taken into account in the notion of collusion-proofness. The following condition guarantees that S accepts the i-grand-contract only if she has observed the signal τ_i,

$$\begin{cases} U_s\left(\sum_{j=1}^{n} p(\theta_j|\tau_z)V_{ij}\right) \geq 0 & \text{for } z = i, \\ U_s\left(\sum_{j=1}^{n} p(\theta_j|\tau_z)V_{ij}\right) < 0 & \text{for all } z \neq i. \end{cases} \tag{3}$$

The other relevant constraints for A are the participation constraints and the incentive compatibility constraints for all types j that have a positive probability to be realized following a signal τ_i. To this purpose, define the set

$$Y = \{j \mid 1 \leq j \leq n \text{ and } p(\theta_j|\tau_i) \neq 0\}.$$

Participation constraints (**IR**) and the incentive compatibility constraints (**IC**) must be satisfied for all i-grand-contracts,

$$\textbf{IR} \quad U_{ij} \geq 0 \qquad \text{for all } i \in [1,n] \text{ and all } j \in Y, \tag{4}$$

$$\textbf{IC} \quad U_{ij} \geq U_{ij'} + (\theta_{j'} - \theta_j)q_{ij'} \qquad \text{for all } i \in [1,n] \text{ and all } j, j' \in Y. \tag{5}$$

The strong collusion feasibility condition, the participation and incentive compatibility constraints define the outcome implementability.

Definition 1 $\{q_{ij}, V_{ij}, U_{ij}\}_{i,j\in\{1,2,\ldots,n\}}$ *is a collusion-proof outcome if it satisfies the participation constraints (4), (3), the incentive compatibility constraint (5) and the strong collusion feasibility condition (2).*

This notion of collusion-proofness intentionally overstresses one aspect of participation decisions. That is, it requires that S must *strictly* prefer not to accept the generic i'-grand-contract when she observes the signal τ_i. This approach is adopted to highlight and strengthen the applicability of the results presented in this paper. Formally, the same conclusion could be obtained with a leaner notation requiring S to weakly prefer not to accept the generic i'-grand-contract, following the realization of τ_i. Nonetheless this approach would be unsatisfactory: the resulting mechanism would not be robust to epsilon-implementation. Consequently, any small perturbations could destroy it. Moreover, the feasibility of epsilon-implementation is not at all trivial under certain conditions. Recognizing these conditions helps to shed light on those aspects of collusion which are more difficult than others to deal with. In the next section some of these conditions are analyzed.

The central result of the paper is presented in the following Proposition:

Proposition 4 *There exists a mechanism that implements a collusion-proof outcome, which allows P to achieve the same expected payoff as the optimal collusion-free outcome.*

Proof. See Appendix. ■

The proof is presented in the Appendix. The implementability of the collusion-free outcome is driven by two essential elements: First, the centralized organization. Second, the assumption of no collusion in participation decisions. This assumption has an implication which has been ignored elsewhere in the literature. That is, if the optimal collusive agreement is contingent on a given participation choice, an element of sequential rationality must be introduced in the solution concept. Therefore restricting attention to direct-revelation mechanisms with full participation is no longer without loss of generality. P can achieve the optimal collusion-free outcome by offering a menu of grand-contracts such that some of these grand-contracts are going to be rejected in equilibrium by the supervisor.

The tension inside the coalition not only derives from the impossibility of coordinating participation decisions: it is accentuated by the centralized organizational response adopted by P. Indeed, under centralized contracting, A has the opportunity of rejecting the collusive side-contract offered by S and playing P's game non-cooperatively. Consequently, A gains bargaining power, improving his outside option. The extent to which he does so depends endogenously on P's grand-contract. On the contrary, if the principal delegates subcontracting to S, then A do not have the option of rejecting the subcontract and dealing with P directly. Therefore, he effectively loses bargaining power.

The implementability of the collusion-free outcome requires that all the bargaining power is attributed to A. This creates the maximal tension inside the coalition: S makes her acceptance decisions keeping in mind that in the second stage of the game A shall have all the bargaining power and all the incentives to truthfully reports his type. Consequently S reports truthfully as well.

4.2 Timing 2

The general model presented above rests on the assumption that S receives her informative signal in the first stage of the game. This assumption is unsatisfactory because it outlines a type of supervision that fits a limited range of cases. Under this assumption, S can be thought as an "informed third party" or a "witness" who happened to learn some information about A even before P had shown any interest in contracting with him. In reality, S's information is often acquired after an inspection or lenghty investigation, which takes place following the acceptance of P's mechanism. In this case our analysis would change. To see how, suppose that S learns her signal after accepting P's mechanism. Clearly, the mechanism proposed in Proposition 4 is no longer collusion-proof because S has not enough information to make the informed participation decision that is crucial for the implementation of the mechanism.

In order to circumvent this problem, P can adopt a different strategy. The mechanism-design works as follows: P offers a menu of mechanisms to A who has to select one of them.[19] A's choice unilaterally determines which mechanism is played. S has not the possibility to select a specific mechanism; she can simply accept or refuse the whole menu of mechanisms. S and A simultaneously make their acceptance decisions, which are not subject to collusion. Following participation decisions, A reports his type and S reports her signal. The timing of the game is as follows,

- At date -1, A learns θ.
- At date 0, P offers a menu of mechanisms to S and A.
- At date 1, A has to accept one of these mechanisms. S accepts or refuses the whole menu of mechanisms.
- At date 2, S learns τ.
- At date 3, S and A can stipulate a side-contract.
- At date 4, A reports his type and S reports her signal.
- At date 5, production and transfers take place.

[19] A could also choose not to accept any mechanism.

Notice that S is assumed to learn her information at date 2, after the acceptance of P's mechanism.[20] This timing-assumption constitutes the substantial difference with respect to the model presented in the previous section. All other assumptions stay the same. We have that:

Proposition 5 *Under the alternative timing where S learns her signal after accepting P's mechanism, there still exists a mechanism that implements the optimal collusion-free outcome.*

Proof. See Appendix. ■

P can implement the collusion-free outcome by inducing A to self report part of his private information in the participation stage. Selecting the appropriate mechanism, A conveys extra information about S's signal. A is willing to do so because he keeps in mind two things. First, under any mechanism that he can select in the participation stage, S is always induced to report truthfully in a way that is robust to any possible side-contract available to the coalition. Second, under any mechanism that he can select in the participation stage, S obtains all the extra rent related to the fact that P cannot directly observe S's signal. Consequently, A obtains no extra-profit from the asymmetric information between P and S. These considerations are sufficient to induce A to truthfully self report part of his private information in the participation stage, because self-reporting this information has no effect whatsoever on his payoff.

5 Remarks on Collusion-Proof Implementation

5.1 Bargaining Power and Collusive Behaviors

A couple of issues concerning bargaining power and collusive behaviors are worth noticing. As we mentioned before, the results do not depend on the distribution of the bargaining power allocation inside the coalition nor do they rest on the identity of the coalition member who offers and initiate the collusive agreement. This generality is guaranteed by the strong notion of collusion-proofness adopted in the paper. Moreover collusion-proof implementation does not depend on special assumptions about the accuracy of S's information. For example, suppose S learns A's cost, i.e., there is no residual asymmetric information between A and S. The mechanism simply reduces to a special case where $p(\theta_i|\tau_i) = 1$ for all i.

[20] The mechanism proposed in this section would work also in the standard case where S receives her information at date -1. In this sense the mechanism presents in this section has a wider applicability with respect to the one presented in Proposition 4.

5.2 Feasibility of Epsilon-Implementation

This paper considers a class of mechanisms which are robust to epsilon-implementation. This choice of solution concept is neither redundant nor irrelevant for the analysis of collusion. As a matter of fact, recognizing the conditions under which epsilon-implementation is not feasible, can help furthering our understanding of those aspects of collusion that are harder to fight. Heterogenous supervision is one of these conditions. Consider this issue in the context of the general model presented under *Timing 1*. Suppose there are many S's types, each one receiving a specific signal τ on the A's cost. This problem is relatively easy to solve if epsilon-implementation is not required. P can simply offer as many menus of grand-contracts as the number of possible S's type, each one tailored on one specific S's types. Similarly to the case of collusion-free supervision, P can design the mechanism in such a way that S obtains a surplus equal to zero regardless of the grand-contract she selects. Being indifferent between telling the truth or lying, S reports the true partition cell she observes, allowing P to extract her information for free. Accordingly P can implement an outcome that yields the second-best payoff. If epsilon-implementation is introduced, the former mechanism would not be collusion-proof. The technical reason is the following. P is not able to offer a salary schedule for S such that she strictly prefers to accept both the appropriate menu *and* the appropriate i-grand-contract within that menu. The formal analysis of this case is beyond the scope of this paper and is left to future research. The point here was only to stress that requiring epsilon-implementation can provide some extra insights on the applicability and robustness of collusion-proof mechanisms.

5.3 Implication for Decentralization

The recent literature evaluating delegation when agents collude offers an intriguing puzzle. FLM (2003) and Celik (2008) represents two influential papers in this literature. Despite the very similar setting they consider, the results of these papers are strikingly different: FLM (2003) find that delegation is always equivalent to centralization, whereas Celik (2008) finds that centralization is superior in general. The results of this paper confirms that centralization performs better than delegation.[21] The crucial assumption of no collusion in participation decisions drives this result. P can improve his payoff by contracting directly with S and A. This is due to the fact that participation decisions can be exploited to extract supplementary information. This is not possible under decentralized contracting: under *Timing 1* S would have a stake in misreporting his signal and then offering a subcontract

[21] On the contrary, Baliga and Sjostrom (1998), and Laffont and Martimort (1998) consider a setup that does not involve supervision, showing that under certain conditions delegation is the optimal organizational response to collusion.

to A which instructs him to misreport his type. Given that S has all the bargaining power, A has no other choice than to accept S's subcontract. Otherwise, he has to renounce the production altogether. On the other hand, centralized contracting allows P to actively influence A's outside option, eliminating the scope for collusion. Therefore, the standard argument in favor of centralized contracting applies also to the present contribution.

6 Conclusions

This paper analyzes the role of supervision in organizations involving both supervisory and productive tasks when these two tasks are performed by different collusive agents. The main contribution of the paper is to show the role of endogenous selection of supervisory activity by the principal. If collusion between supervisor and agent can occur only after they have decided to participate in the mechanism, endogenous selection of supervisory activity can costlessly eliminate collusion. This conclusion is robust to alternative information structures, collusive behaviors and specification of agent's types. Surprisingly, the cost related to collusion can be fully eliminated even when there is no residual asymmetric information between the agent and the supervisor. This paper, therefore, does not build on the important insight gained from Laffont and Martimort (1997, 2000) that agents' asymmetric information constitutes an obstacle to collusive arrangements. Quite the opposite, this paper highlights a different "Achilles' heel" of collusive coalition: the inability to collude prior to making a decision to participate in the mechanism. The full implications of this assumption have been overlooked by the literature so far. A valuable lesson from the current paper lies in the following observation: under the assumption of *no collusion in participation decisions*, restricting attention to direct revelation mechanisms with full participation is no longer without loss of generality. As a matter of fact, assuming *no collusion in participation decisions* imports some kind of sequential rationality in the design problem; the agent and the supervisor may collude, but only after the participation decisions have already been made. This is the reason why the partial participation mechanism proposed in this paper improves over the optimal outcomes identified by both Celik (2008) and Faure-Grimaud, Laffont and Martimort (2003).

The present contribution also develops a unified and general model, which includes as special cases the frameworks considered in Celik (2008) and Faure-Grimaud, Laffont and Martimort (2003). The results of this paper seem to confirm Celik's (2008) intuition that centralized contracting performs better than decentralized contracting.

Admittedly, the collusion-proof implementation presented in this paper heavily relies on the assumption of *no collusion in participation decision.* Far from strenuously trying to make a case in favor of this assumption, which is nevertheless plausible in many realistic situations, this paper intends to shed light on those factors that make collusion truly problematic by identifying the ones which are less so. From this perspective, the present contribution seems to suggest that allowing for collusion on participation decisions may be a more interesting way of thinking about collusion. We suspect that *selective supervision* is useful in a setting where collusion occurs prior to participation, though in that context it is unlikely to costlessly eliminate collusion. This issue remains to be explored in future research.

7 Appendix

7.1 Proof of Proposition 1

In this part of the Appendix the implementation of the grand-mechanim Γ is analyzed. The proof is divided into two parts. The first part proves that this grand-mechanism is not subject to coalition formation between S and A. The second part shows that Γ allows P to achieve the same expected payoff as the optimal collusion-free outcome.

7.1.1 Part 1

Assuming *no collusion in participation decisions* implies that collusion is a problem only in case S accepts the grand-contract, i.e. whenever she reports a_1 to P. Now consider a manipulation where type θ_3 is misreported as type θ_1. This deviation is not profitable because it does not improve the coalition utility,

$$V_{13} \geq V_{11} + (\theta_1 - \theta_3)q_{11}.$$

After substituting and rearranging, the former inequality is reduced to

$$(\theta_3 - \theta_1)(q_{11} - q_{12}) \geq 0.$$

This rules out type θ_3's imitation of type θ_1, because $q_{11} \geq q_{12}$ and $\theta_3 \geq \theta_1$. Now consider a manipulation where type θ_3 is misreported as type θ_2. This deviation is not profitable if the following inequality holds,

$$V_{13} \geq V_{12} + (\theta_2 - \theta_3)q_{12}.$$

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 θ_3's imitation of type θ_2. Now consider a manipulation where type θ_2 is misreported as type θ_1. This deviation is not profitable if the following inequality holds:

$$V_{12} \geq V_{11} + (\theta_1 - \theta_2)q_{11}.$$

After substituting and rearranging, the former inequality is reduced to

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

This rules out type θ_2's imitation of type θ_1. Now consider a manipulation where type θ_2 is misreported as type θ_3. This deviation is not profitable if the following inequality holds:

$$V_{12} \geq V_{13} + (\theta_3 - \theta_2)q_{13}.$$

After substituting and rearranging, the former inequality is reduced to

$$0 \geq (\theta_3 - \theta_2)\left(q_{13} - q_{12}\right).$$

This rules out type θ_2's imitation of type θ_3. Now consider a manipulation where type θ_1 is misreported as type θ_2. This deviation is not profitable if the following inequality holds:

$$V_{11} \geq V_{12} + (\theta_2 - \theta_1)q_{12}.$$

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 θ_1's imitation of type θ_2. Now consider a manipulation where type θ_1 is misreported as type θ_3. This deviation is not profitable if the following inequality holds:

$$V_{11} \geq V_{13} + (\theta_3 - \theta_1)q_{13},$$

After substituting and rearranging, the former inequality is reduced to

$$0 \geq (\theta_3 - \theta_2)\left(q_{13} - q_{12}\right).$$

This rules out type θ_1's imitation of type θ_3.

7.1.2 Part 2

Given that the mechanism is not subject to coalition formation between S and A, it follows that A and S respond to the grand-mechanism in a non-cooperative fashion. Consequently, the participation constraints for S hold when she observes the partition $\{\theta_2, \theta_1\}$

$$\frac{f(\theta_2)}{f(\theta_2) + f(\theta_1)} w_{12} + \frac{f(\theta_1)}{f(\theta_2) + f(\theta_1)} w_{11} \geq 0, \tag{6}$$

whereas it does not 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 θ_3,

$$\textbf{IR } (\theta_3|a_2) \quad U_{23} \geq 0, \tag{7}$$

$$\textbf{IC } (\theta_3|a_2) \quad U_{23} \geq U_{2j'} + (\theta_{j'} - \theta_3) q_{2j'} \text{ for all } j' = \{1, 2\}\,. \tag{8}$$

Notice that type θ_3 prefers to tell the truth to P since

$$t_{23} - \theta_3 q_{23} = 0 \geq t_{22} - \theta_3 q_{22} = 0,$$

$$t_{23} - \theta_3 q_{23} = 0 \geq t_{21} - \theta_3 q_{21} = 0,$$

Consider now the participation and the incentive compatibility constraints for all types when S accepts the grand contract

$$\textbf{IR } (\theta_j|a_1) \quad U_{1j} \geq 0 \quad \text{ for all } j = \{1, 2, 3\}\,, \tag{9}$$

$$\textbf{IC } (\theta_j|a_1) \quad U_{1j} \geq U_{1j'} + (\theta_{j'} - \theta_j) q_{1j'} \text{ for all } j, j' = \{1, 2, 3\}\,. \tag{10}$$

The most efficient agent, θ_1, prefers to tell the truth to P: indeed, the following incentive constraints hold

$$t_{11} - \theta_1 q_{11} = (\theta_2 - \theta_1) q_{12} \geq t_{12} - \theta_1 q_{12} = (\theta_2 - \theta_1) q_{12},$$

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

Similarly, type θ_2 prefers to tell the truth to P since

$$t_{12} - \theta_2 q_{12} = 0 \geq t_{13} - \theta_2 q_{13} = 0,$$

$$t_{12} - \theta_2 q_{12} = 0 \geq t_{11} - \theta_2 q_1 = (\theta_2 - \theta_1)\left(q_{12} - q_{11}\right).$$

Finally, type θ_3 prefers to tell the truth to P since

$$t_{13} - \theta_3 q_{13} = 0 \geq t_{12} - \theta_3 q_{12} = -(\theta_3 - \theta_2) q_{12},$$

$$t_{13} - \theta_3 q_{13} = 0 \geq t_{11} - \theta_3 q_{11} = (\theta_2 - \theta_1) q_{12} - \left(\theta_3 - \theta_1\right) q_{11}.$$

Following S's participation decision, P updates his beliefs on A's type. Accordingly, conditional probabilities become

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

It follows that P's expected utility is given by

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

Given that 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 (11) with respect to q_{11}, q_{12}, q_{23}. A brief inspection reveals that:

$$
\begin{aligned}
V_{23} &= V_3^{cf} = 0, \\
V_{12} &= V_2^{cf} = 0, \\
V_{11} &= V_1^{cf} = (\theta_2 - \theta_1) q_{22},
\end{aligned}
$$

which implies that (11) 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.

7.2 Proof of Proposition 2

The relevant constraints for A are the participation constraints (**IR**) and the incentive compatibility constraints (**IC**) for all types,

$$
\textbf{IR} \quad U_{ij} \geq 0 \qquad \text{for all } i, j = \{1, 2\}\,, \tag{12}
$$

$$
\textbf{IC} \quad U_{ij} \geq U_{ij'} + (\theta_{j'} - \theta_j) q_{ij'} \qquad \text{for all } i, j, j' = \{1, 2\}\,. \tag{13}
$$

It is easy to notice that the composition of the mechanism 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

$$t_{11} - \theta_1 q_{11} = \Delta\theta q_{12} \geq t_{12} - \theta_1 q_{12} = \Delta\theta q_{12},$$

$$t_{21} - \theta_1 q_{21} = \Delta\theta q_{22} \geq t_{22} - \theta_1 q_{22} = \Delta\theta q_{22}.$$

Similarly, the inefficient type prefers to tell the truth to P since

$$t_{12} - \theta_2 q_{12} = 0 \geq t_{11} - \theta_2 q_{11} = \Delta\theta\left(q_{12} - q_{11}\right),$$

$$t_{22} - \theta_2 q_{22} = 0 \geq t_{21} - \theta_2 q_{21} = \Delta\theta\left(q_{22} - q_{21}\right).$$

For ε small enough, S's utility function is given by,

$$\lim_{\varepsilon\to 0} U_s = p(\theta_1|\tau_i)\varepsilon - p(\theta_2|\tau_i)\left(\frac{p(\theta_1|\tau_1)}{p(\theta_2|\tau_1)}\varepsilon\right) \qquad \text{for all } i = \{1,2\}.$$

The participation constraints for S holds when the realized signal is τ_1,

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

whereas it doesn't hold when the signal τ_2 is observed,

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

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 ε 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 type θ_2 to report that he has type θ_1 in an attempt to avoid the negative transfer s_{12}. This deviation is not profitable because it does not improve the coalition utility,

$$V_{12} \geq V_{11} + (\theta_1 - \theta_2)q_{11},$$

After substituting and rearranging, the former inequality is reduced to

$$-\varepsilon\left(\frac{p(\theta_1|\tau_1)}{p(\theta_2|\tau_1)} + 1\right) \geq \left(\theta_2 - \theta_1\right)\left(q_{12} - q_{11}\right).$$

For small values of ε, this rules out type θ_2's imitation of type θ_1. Now consider a manipulation where type θ_1 is misreported as type θ_2. This deviation is not profitable if the following inequality holds:

$$V_{11} \geq V_{12} + (\theta_2 - \theta_1) q_{12}.$$

After substituting and rearranging, the former inequality is reduced to

$$\varepsilon \geq -\frac{p(\theta_1|\tau_1)}{p(\theta_2|\tau_1)}\varepsilon.$$

Type θ_1's imitation of type θ_2 is ruled out. This is sufficient to prove that the grand-contract is not subject to coalition formation. Consequently, when τ_2 (τ_1) is realized, S refuses (accepts) the grand-contract. Following S's decision, P updates his beliefs on the agent's type. Conditional probabilities become,

$$\begin{aligned} p(\theta_j|a_1) &= p(\theta_j|\tau_1) \quad \text{for } j = \{1, 2\}, \\ p(\theta_j|a_2) &= p(\theta_j|\tau_2) \quad \text{for } j = \{1, 2\}. \end{aligned}$$

It follows that P expected utility is given by,

$$\begin{aligned} & p(\theta_1|a_1)\left[W(q_{11}) - \theta_1 q_{11} - U_{11}\right] + \\ & + p(\theta_2|a_1)\left[W(q_{12}) - \theta_2 q_{12} - U_{12}\right] + \\ & + p(\theta_1|a_2)\left[W(q_{21}) - \theta_1 q_{21} - U_{21}\right] + \\ & + p(\theta_2|a_2)\left[W(q_{22}) - \theta_2 q_{22} - U_{22}\right]. \end{aligned} \tag{14}$$

The optimal contract solves

$$\begin{aligned} & \max_{\{q_{i1}, q_{i2}, u_{i1}, u_{i2}\}_{i \in \{1,2\}}} \quad p(\theta_1|a_i)\left[W(q_{i1}) - \theta_1 q_{i1} - u_{i1}\right] + p(\theta_2|a_i)\left[W(q_{i2}) - \theta_2 q_{i2} - u_{i2}\right], \\ & s.t. \quad q_{i1} \geq q_{i2} \quad \text{for all } i \in \{1, 2\}. \end{aligned}$$

The solution of this problem yields what FLM (2003) denote by conditionally-optimal second-best, which implements the first-best outputs $q_{i1}^{sb} = q_1^{fb}$ for an efficient agent and outputs q_{i2}^{sb} for an inefficient one, where

$$\begin{aligned} W'(q_1^{fb}) &= \theta_1, \\ W'(q_{i2}^{sb}) &= \theta_2 + \frac{p_{i1}}{p_{i2}}\Delta\theta. \end{aligned}$$

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 τ_1 rather than after the observation of τ_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 Γ_1 allows P to achieve the same expected payoff as the optimal collusion-free outcome: all the costs associated with collusion are fully eliminated.

7.3 Proof of Proposition 3

Consider the following grand-mechanism, Γ_2. P offers a menu of grand-contracts and S can select one of them: accepting one of these grand-contracts implicitly comports refusing all the other ones.[22] A chooses whether to participate in the grand-mechanism or not. S and A simultaneously make their acceptance decisions, which are not subject to collusion. The smallest message spaces for S and A compatible with the implementation of this grand-mechanism, are, respectively, $M_s \in \{a_1, a_2, ..., a_m\}$ and $M_a \in \{\theta_1, \theta_2, ..., \theta_n\}$, where a_i stands for *"accept the grand contract i."* Each message a_i corresponds to a generic i-partion whose first element is θ_{h_i}. Let $k_i \in N^+$ denotes the number of A's types included in the i-partition, which can be represented as $\{\theta_{h_i}, \theta_{h_i+1}, ..., \theta_{h_i+k_i}\}$. For the sake of simplicity, denote by s_{ij} (respectively t_{ij}, q_{ij}, V_{ij} and U_{ij}) S's wage (respectively 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 θ_j. First, consider the generic i-grand-contract, for $1 < i < m$:

$$\begin{aligned}
&t_{ij} = 0, && s_{ij} = -\varepsilon, && \text{for } h_i + k_i < j \leq n,\\
&t_{ij} = \sum_{z=j+1}^{h_i+k_i+1} U_{iz} + (\theta_{j+1} - \theta_j) q_{ij+1} + \theta_j q_{ij}, && s_{ij} = 0, && \text{for } h_i \leq j \leq h_i + k_i,\\
&t_{ij} = \sum_{z=j+1}^{h_i+k_i+1} U_{iz} + (\theta_{j+1} - \theta_j) q_{ij+1} + \theta_j q_{ij} + \varepsilon, && s_{ih_i-1} = -\varepsilon, && \text{for } 1 \leq j < h_i,
\end{aligned}$$

and

$$\begin{aligned}
&q_{ij} = 0, && \text{for } j > h_i + k_i,\\
&q_{ij} = q_{ij}^{cf}, && \text{for } 1 \leq j \leq h_i + k_i,
\end{aligned}$$

[22] S could also choose not to accept any grand-contracts, by rejecting the grand-mechanism proposed by P.

where $\varepsilon \in R^+$ is small and the output profile q_{ij}^{cf} is weakly decreasing, i.e., $q_{i1}^{cf} \geq q_{i2}^{cf} \geq ... \geq q_{ih_i+k_i}^{cf}$. The relevant constraints for A are the participation constraints (**IR**) and the incentive compatibility constraints (**IC**) for all types,

$$\textbf{IR} \quad U_{ij} \geq 0 \qquad \text{for all } j, \tag{15}$$

$$\textbf{IC} \quad U_{ij} \geq U_{ij'} + (\theta_{j'} - \theta_j) q_{ij'} \quad \text{for all } j, j'. \tag{16}$$

Notice that the composition of the i-grand-contract is incentive compatible and meets the participation constraints for A. Moreover it induces S to participate only if she has observed the i-partition. In other words, the i-grand-contract induces thruthful revelation of both A's type and S's partition. In order to prove this point, suppose that S accepts the i'-grand-contract when she observed the i-partition. If A truthfully reveals his type, S receives a negative payoff. To avoid this negative payoff, S could offer a bribe to convince A to misreport his type as one belonging to the i-partition, i.e., θ_j where $h_i + k_i \geq j \geq h_i$. Nontheless, there is no collective gain from this misreport. In order to formally demonstrate the collusion proofness of the i-grand-contract, it is sufficient to prove that S has no stake in bribing the i-partition's upper and lower adjacent type, i.e., respectively types $\theta_{h_i+k_i+1}$ and θ_{h_i-1}. Indeed, bribing any other type would be even more costly: if S is not willing to bribe types $\theta_{h_i+k_i+1}$ and θ_{h_i-1}, she would definitely not consider to bribe anyone else either. To start with, consider type $\theta_{h_i+k_i+1}$; by mimicking type $\theta_{h_i+k_i}$ he would obtain a negative payoff equal to $(\theta_{h_i+k_i} - \theta_{h_i+k_i+1}) q_{ih_i+k_i}^{cf}$. Mimicking any other type inside the i-partition would be even more costly. S must compensate A's loss from misreporting. Clearly, S prefers paying $-\varepsilon$ to P than offering $(\theta_{h_i+k_i} - \theta_{h_i+k_i+1}) q_{ih_i+k_i}^{cf}$ to A in order to misrepresent his type. If we were to consider a less efficient A's type, the collective loss from this misreport would be even bigger: for example, if type $\theta_{h_i+k_i+2}$ mimics type $\theta_{h_i+k_i}$ he obtains a negative payoff equal to $(\theta_{h_i+k_i} - \theta_{h_i+k_i+2}) q_{ih_i+k_i}^{cf} \leq (\theta_{h_i+k_i} - \theta_{h_i+k_i+1}) q_{ih_i+k_i}^{cf}$. The same applies to types that are more efficient than the ones in the i-partition. Consider type θ_{h_i-1}; by mimicking type θ_{h_i} he would forgo a payoff equal to ε. Mimicking any other type inside the i-partition would be even more costly because the output profile q_{ij}^{cf} is weakly decreasing. Clearly, S is indifferent between paying $-\varepsilon$ to P or offering ε to A to misrepresent his type. If we were to consider a more efficient A's type, the collective loss from this misreport would be even bigger: for example, if type θ_{h_i-2} mimics type θ_{h_i} he forgoes a payoff equal to $(\theta_{h_i-1} - \theta_{h_i-2}) \left(q_{ih_i-2}^{cf} - q_{ih_i-1}^{cf} \right) \geq \varepsilon$. This proves that the i-grand-contract is collusion-proof and induces truthful revelation of both A's type and S's partition.

Two i-grand-contracts are left to be analyzed, i.e., $i = 1$ and $i = m$. The former correponds to the first partition $\{\theta_1, \theta_2, ..., \theta_{k_1}\}$, while the latter corresponds to the last partition $\{\theta_{n-k_m}, ..., \theta_n\}$. Consider the 1-grand-contract,

$$\begin{array}{lll} t_{ij} = 0 & s_{ij} = -\varepsilon & \text{for } k_1 \leq j \leq n, \\ t_{ij} = \sum_{z=j+1}^{k_i+1} U_{iz} + (\theta_{j+1} - \theta_j) q_{ij+1} + \theta_j q_{ij} & s_{ih_i - 1} = 0 & \text{for } 1 \leq j \leq k_1, \end{array}$$

and

$$\begin{array}{ll} q_{ij} = 0 & \text{for } k_1 \leq j \leq n, \\ q_{ij} = q_{ij}^{cf} & \text{for } 1 \leq j \leq k. \end{array}$$

The generic m-grand-contract,

$$\begin{array}{lll} t_{ij} = \sum_{z=j+1}^{n} U_{iz} + (\theta_{j+1} - \theta_j) q_{ij+1} + \theta_j q_{ij} & s_{ij} = 0 & \text{for } n - k_m \leq j \leq n, \\ t_{ij} = \sum_{z=j+1}^{n} U_{iz} + (\theta_{j+1} - \theta_j) q_{ij+1} + \theta_j q_{ij} + \varepsilon & s_{ij} = -\varepsilon & \text{for } 1 \leq j < n - k_m, \end{array}$$

and

$$q_{ij} = q_{ij}^{cf} \qquad \text{for all } j,$$

where $\varepsilon \in R^+$ is small and the output profile q_{ij}^{cf} is weakly decreasing for both contracts. It is straightforward to notice that the proof presented above applies also to these grand-contracts.

Up to this point, it has been proved that the i-grand-contract induces truthful revelation of both A's type and S's partition, meaning that it induces S to participate only if she has observed the i-partition. More than that, P is able to obtain S's information without forgoing any rents to her. In other words, the grand-mechanism Γ_2 works as if P would directly learns S's signal. Finally, the grand-mechanism implements the optimal collusion-free outcome. To prove this point, it is sufficient to notice that P leaves A with the same information rent he would receive in the optimal collusion-free outcome. P can achieve this outcome because each i-grand-contract offers zero production and zero transfers to all types that are less efficient than the i-partition's ones. This reduces the information rents to the i-partition's types accordingly. One final remark. Recall that the information environment considered in Celik (2008) entails that when supervisory activity is in place (and in the absence of collusion), there is no monotonicity requirement regarding two output levels in separate partition cells. This proprierty is still present in Γ_2. Indeed there is no monotonicity requirement regarding two output levels in different i-grand-contracts. QED.

7.4 Proof of Proposition 4

Consider the following mechanism, Γ_3. P offers a menu of grand-contracts and S selects one of them: accepting one of these grand-contracts implicitly comports refusing all the other ones.[23] A chooses whether to participate in the grand-mechanism or not. S and A simultaneously make their acceptance decisions, which are not subject to collusion. The smallest message spaces for S and A compatible with the implementation of this grand-mechanism is respectively $M_s \in \{a_1, a_2, ..., a_n\}$ and $M_a \in \{\theta_1, \theta_2, ..., \theta_n\}$, where a_i stands for *"accept the grand contract i."* The message a_i corresponds to the signal τ_i. Denote by s_{ij} (respectively t_{ij}, q_{ij}, V_{ij} and U_{ij}) S's wage (respectively 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 θ_j. Moreover, define $Y = \{j \mid 1 \leq j \leq n \text{ and } p(\theta_j|\tau_i) \neq 0\}$ as the collection of all types j that have zero probability to be realized for a given signal τ_i. Consider a salary w_{ij} for S that solves the following system,

$$\begin{cases} \sum_{j=1}^{n} p(\theta_j|\tau_z) w_{ij} = 0, & \text{for } z = i, \\ \sum_{j=1}^{n} p(\theta_j|\tau_z) w_{ij} = -\varepsilon, & \text{for all } z \neq i, \end{cases}$$

where $\varepsilon \in R^+$ is small. Two aspects are worth noticing. First, this is a linear system of n equations in n variables: consequently, there are n well defined solutions. Second, this system guarantees that S receives a monetary expected payoff equal to zero when she accepts the i-grand-contract after observing the signal τ_i (see first equation). On the other hand, this i-grand-contract leaves S with a negative expected monetary payoff when she observes any signal different from τ_i (see second equation). Moreover, w_{ij} can be rescaled without affecting any relevant implications. For example, divide both expressions for a generic positive number N. It follows that

$$\begin{cases} \sum_{j=1}^{n} p(\theta_j|\tau_z) \frac{w_{ij}}{N} = 0, & \text{for } z = i, \\ \sum_{j=1}^{n} p(\theta_j|\tau_z) \frac{w_{ij}}{N} = -\frac{\varepsilon}{N}, & \text{for all } z \neq i. \end{cases}$$

For example, suppose S's utility function is given by $U_s = \frac{1}{r}\left(1 - e^{-rx}\right)$, where x represents S's monetary payoff. It is easy to notice that

$$\lim_{N \to \infty} U_s = \sum_{j=1}^{n} p(\theta_j|\tau_i) \frac{w_{ij}}{N}.$$

[23] S could also choose not to accept any grand-contracts, by rejecting the grand-mechanism proposed by P.

By offering a salary equal to $s_{ij} = \frac{w_{ij}}{N}$ for all $j \in [1, n]$, P can induce S to accept the i-grand-contract only if she has observed the signal τ_i. This clearly holds for a generic "concave" utility function as well. Consider now the i-grand-contract; the production and transfer levels are given by,

$$
\begin{aligned}
t_{ij} &= \sum_{z=j+1}^{n} U_{iz} + (\theta_{j+1} - \theta_j) q_{ij+1} + \theta_j q_{ij} + |s_{ij}|, &\qquad \text{for } j \in Y, \\
t_{ij} &= 0, &\qquad \text{for } j \notin Y,
\end{aligned}
$$

and

$$
\begin{aligned}
q_{ij} &= q_{ij}^{sb}, &\qquad \text{for } j \in Y, \\
q_{ij} &= 0, &\qquad \text{for } j \notin Y,
\end{aligned}
$$

where q_{ij}^{sb} is weakly decreasing, i.e., $q_{i1}^{sb} \geq q_{i2}^{sb} \geq ... \geq q_{in}^{sb}$. Recall that the relevant constraints for A are the participation constraints and the incentive compatibility constraints for all types $j \in Y$. Participation constraints (**IR**) and the incentive compatibility constraints (**IC**) must be satisfied for all i-grand-contracts,

$$
\textbf{IR} \quad U_{ij} \geq 0 \qquad \text{for all } i \in [1, n] \text{ and all } j \in Y,
$$

$$
\textbf{IC} \quad U_{ij} \geq U_{ij'} + (\theta_{j'} - \theta_j) q_{ij'} \qquad \text{for all } i \in [1, n] \text{ and all } j, j' \in Y.
$$

Notice that the composition of the i-grand-contract is incentive-compatible and meets the participation constraints for A. What is left to prove is that the outcome meets the “strong collusion feasibility condition.” Suppose S accepts the i-grand-contract. Notice that there is no collective gain from any possible misreports if s_{ij} is small enough. The coalition utility is given by,

$$
\begin{aligned}
V_{ij} &= \sum_{z=j+1}^{n} U_{iz} + (\theta_{j+1} - \theta_j) q_{ij+1} + \theta_j q_{ij} + 2s_{ij}, &\qquad \text{for } j \in Y \text{ and } s_{ij} > 0, \\
V_{ij} &= \sum_{z=j+1}^{n} U_{iz} + (\theta_{j+1} - \theta_j) q_{ij+1} + \theta_j q_{ij}, &\qquad \text{for } j \in Y \text{ and } s_{ij} \leq 0, \\
V_{ij} &= 0, &\qquad \text{for } j \notin Y,
\end{aligned}
$$

A fast inspection reveals that this outcome satisfies the “strong collusion feasibility condition.” The coalition has no incentive to deviate from reporting truthfully. Up to this point, it has been proved that the outcome $\{q_{ij}, V_{ij}, U_{ij}\}_{i,j \in \{1,2,\ldots,n\}}$ is collusion-proof. More importantly, P is able to obtain S's information without forgoing any rents to her. It is as if P would directly learns S's signal. Based on this signal P offers a standard incentive-compatible

i-grand-contract. It follows that Γ_3 implements the optimal collusion-free outcome. QED.

7.5 Proof of Proposition 5

Consider the following mechanism, Γ_4. P offers a menu of grand-contracts and A selects one of them: accepting one of these grand-contracts implicitly comports refusing all the other ones.[24] S chooses whether to participate in the mechanism or not. S and A simultaneously make their acceptance decisions, which are not subject to collusion. After the acceptance of the mechanism, S learns her signal. Following A's selection of the grand-contract, A and S must respond to the selected grand-contract. The message spaces for S and A are respectively $M_s \in \{\tau_1, \tau_2, ..., \tau_n\}$ and $M_a \in \{\theta_1, \theta_2, ..., \theta_n\}$.

In the first stage of the game, A selects the grand-contract he wants to take part in. In order to do so, he sends a message from the message space $M_{\overline{a}} \in \{a_1, a_2, ..., a_n\}$, where a_k stands for *"accept the grand contract k."* Denote by *k-grand-contract* the grand-contract selected by A when he reports a_k. The k-grand-contract specifies zero transfers and zero production for any message τ_b sent by S where $b \in [k+1, ..., n]$. Denote by s_{kij} (respectively t_{kij}, q_{kij}, V_{kij} and U_{kij}) S's wage (respectively A's transfer, the output target, the coalition information rent and A's utility) when A select the *k-grand-contract* and reports that he has type θ_j and S reports τ_i. Moreover, define $Y = \{j \mid 1 \leq j \leq n \text{ and } p(\theta_j|\tau_i) \neq 0\}$ as the collection of all types j that have zero probability to be realized for a given signal τ_i.

Consider now the k-grand-contract. The production and transfer levels are given by,

$$\begin{aligned}
t_{kij} &= \sum_{z=j+1}^{n} U_{kiz} + (\theta_{j+1} - \theta_j) q_{kij+1} + \theta_j q_{kij}, && \text{for } j \in Y \text{ and } i = \{1, ..., k\} \\
t_{kij} &= 0, && \text{for } j \notin Y \text{ and } i = \{k+1, ..., n\},
\end{aligned}$$

and

$$\begin{aligned}
q_{kij} &= q^{sb}_{kij}, && \text{for } j \in Y \text{ and } i = \{1, ..., k\}, \\
q_{kij} &= 0, && \text{for } j \notin Y \text{ and } i = \{k+1, ..., n\},
\end{aligned}$$

where q^{sb}_{kij} is weakly decreasing, i.e., $q^{sb}_{ki1} \geq q^{sb}_{ki2} \geq ... \geq q^{sb}_{kin}$. Recall that the relevant constraints for A are the participation constraints and the incentive compatibility constraints for all types $j \in Y$. Participation constraints (**IR**) and the incentive compatibility constraints (**IC**) must be satisfied for all k-grand-contracts,

$$\textbf{IR} \quad U_{kij} \geq 0 \qquad \text{for all } i \in [1, n] \text{ and all } j \in Y,$$

[24] A could also choose not to accept any grand-contracts, by rejecting the grand-mechanism proposed by P.

$$\textbf{IC} \quad U_{kij} \geq U_{kij'} + (\theta_{j'} - \theta_j)q_{kij'} \quad \text{for all } i \in [1,n] \text{ and all } j,j' \in Y.$$

Notice that the composition of the k-grand-contract is incentive-compatible and meets the participation constraints for A. Γ_4 specifies a salary schedule s_{kij} such that

$$\sum_{j=1}^{n} p(\theta_j|\tau_i)s_{kij} \geq \sum_{j=1}^{n} p(\theta_j|\tau_i)\left[s_{ki'j} + (U_{ki'j} - U_{kij})\right] \geq 0 \tag{17}$$

for all $k, j = \{1, 2, ..., n\}$ and $i, i' = \{k+1, ..., n\}$. The condition (17) ensures that S has no stake in misreporting her signal. More than that, it also guarantees that for all k-grand-contract the players are unable to find a profitable collective manipulation. In other words, (17) implies that

$$\{\theta_j, V_{kij}\}_{k,j=\{1,2,...,n\} \text{ and } i,i'=\{k+1,...,n\}} \in \arg \max_{\{\widehat{c}(\theta_j)\}_{j\in\{1,2,...,n\}}} V_{ki\widehat{c}(\theta_j)} + (\theta_{\widehat{c}(\theta_j)} - \theta_j)q_{ki\widehat{c}(\theta_i)}. \tag{18}$$

It follows that the outcome implemented by Γ_4 is collusion proof. Interestingly, condition (17) ensures that the information rent arising from the fact that P cannot observe S's signal is entirely captured by the latter. A obtains no extra-profit from the asymmetric information between P and S. Therefore, A is willing to self-report S information in the first stage of the game because self-reporting this information has no effect whatsoever on his payoff. Clearly, S would like to influence A's participation decision. The assumption of "no collusion in participation decision" prevent this from happening. Importantly, when A selects the k-grand-contract correspondent to the true realization of S's signal τ_k, the k-grand-contract specifies zero transfers and zero production for any message $\tau_{b\in[k+1,...,n]}$ sent by S. When the realized signal is τ_k, the only profitable manipulation for S is precisely the misreport of her signal τ_k as $\tau_{b\in[k+1,...,n]}$. But this possibility is not available when A "truthfully" selects the k-grand-contract. Given that S has no stake in misreporting her signal τ_k as $\tau_{b\in[1,...,k]}$, she can be left with no further rent. Therefore, P is able to obtain S's information without forgoing any rent to her or to A. It is as if P would directly learns S's signal. Γ_4 implements the optimal collusion-free outcome. QED.

References

[1] Abdulkadiroglu, A. and K. Chung, 2003, *Auction design with tacit collusion*, Columbia University working paper.

[2] Baliga, S. and T. Sjostrom, 1998, *Decentralization and Collusion*, Journal of Economic Theory, 83, 196-232.

[3] Burlando, A. and Motta, A., 2008a, *Can Self Reporting Reduce Corruption in Law Enforcement?*, University of Padua working paper.

[4] Burlando, A. and Motta, A., 2008b, *Conditional Delegation and Optimal Supervision*, Boston University working paper.

[5] Celik, G., 2008, *Mechanism Design with Collusive Supervision*, forthcoming, Journal of Economic Theory.

[6] Celik, G., Peters, M., 2008b, *Equilibrium Rejection of a Mechanism*, UBC Department of Economics working paper.

[7] Che, Y. and J. Kim, 2006, *Robustly collusion-proof implementation*, Econometrica, 74, 1063-1107.

[8] Che, Y. and J. Kim, 2008,*Optimal collusion-proof auctions*, Columbia University working paper.

[9] Dequiedt, V., 2006. *Ratification and Veto Constraints in Mechanism Design,* working paper.

[10] Dequiedt, V., 2007. *Efficient collusion in optimal auctions*, Journal of Economic Theory, vol. 136(1), 302-323, September.

[11] Faure-Grimaud A., Laffont J. J. and D. Martimort, 2003, *Collusion, Delegation and Supervision with Soft Information*, Review of Economic Studies, 70, 253-275.

[12] Faure-Grimaud A. and D. Martimort, 2001, *On some agency costs of intermediated contracting*, Economics Letters, 71, 75-82.

[13] Innes, Robert, 1999. *Remediation and self-reporting in optimal law enforcement*, Journal of Public Economics, Elsevier, vol. 72(3), pages 379-393, June.

[14] Innes, Robert. 2001. *Violator Avoidance Activities and Self-Reporting in Optimal Law Enforcement.* Journal of Law, Economics, and Organization v.17(1) pp. 239-56.

[15] Kaplow, Luois, and Steven Shavell. 1994. "Optimal Law Enforcement with Self Reporting of Behavior," *Journal of Political Economy*, 102(3), 583-606.

[16] Kaplow, L., and Steven Shavell. 1994. *Optimal Law Enforcement with Self Reporting of Behavior*, Journal of Political Economy, 102(3), 583-606.

[17] Khalil, F. and J. Lawarree, 2006, *Incentives for corruptible auditors in the absence of commitment*, Journal of Industrial Economics, 15(2), 457-478.

[18] Laffont, J. J. and D. Martimort, 1997, *Collusion under asymmetric information*, Econo-

metrica, 65, 875-911.

[19] Laffont, J. J. and D. Martimort, 1998, *Collusion and delegation,* Rand Journal of Economics, 29, 280-305.

[20] Laffont, J. J. and D. Martimort, 2000, *Mechanism design with collusion and correlation*, Econometrica, 68, 309-342.

[21] Laffont, J. J. and Tirole, J., 1991, *The Politics of Government Decision-Making: A Theory of Regulatory Capture*, The Quarterly Journal of Economics, Vol. 106, No. 4 (Nov., 1991), pp. 1089-1127

[22] Laffont, J. J. and Tirole, J., 1993, *A Theory of Incentives in Procurement and Delegation* (Cambridge, MA: MIT Press).

[23] Lambert-Mogilianksy, A., 1998, *On Optimality of Illegal Collusion in Contracts*, Review of Economic Design, 3, 303-348.

[24] Martimort, D., 1999., *The life cycle of regulatory agencies: Dynamic capture and transaction costs*, The Review of Economic Studies, 66, 929-947.

[25] Melumad, N., and Reichelstein, S., 1984, *Value of Communication in Agencies*, The Journal of Economic Theory, Vol. 47 (1989), 334-368.

[26] Mookherjee, D., 2006, *Decentralization, hierarchies, and incentives: A mechanism design perspective*, Journal of Economic Literature, 44(2), 367-390.

[27] Mookherjee, Dilip and Reichelstein Stefan, 1990, *Implementation via Augmented Revelation Mechanisms,* Review of Economic Studies 57, 453-475.

[28] Mookherjee, Dilip and Masatoshi Tsumagari, 2004, *The Organization of Supplier Networks: Effects of Delegation and Intermediation*, Econometrica, 72(4), 1179-1219.

[29] Pavlov, G., 2008, *Colluding on participation decisions*, Boston University working paper.

[30] Quesada, L., 2004, *Collusion as an informed principal problem*, University of Wisconsin working paper.

[31] Tirole, J., 1986, *Hierarchies and Bureaucracies*, Journal of Law, Economics, and Organization, 2, 181-214.