# Mechanism Design with Bilateral Contracting[1]

Vianney Dequiedt[2] and David Martimort[3]

This version: December 4, 2009

**Abstract:** Suppose a principal can only sign public bilateral contracts with agents who have private information on their costs of producing some outputs on his behalf. In such contexts, the principal may manipulate what he learns by contracting with an agent when dealing with others. Introducing this possibility for manipulations may simplify significantly optimal mechanisms. It restores both the continuity of the principal's and the agents' payoffs and that of the optimal mechanism with respect to the information structure. Still, correlation remains useful to better extract the agents' information rent. In private values contexts, a *Revelation Principle with bilateral contracting* identifies the set of implementable allocations by means of simple *non-manipulability constraints.* Equipped with this tool, we characterize optimal non-manipulable mechanisms in various environments. Those mechanisms trade off the marginal benefit of production against some *generalized virtual costs* that extend the standard formula found at zero correlation.

Keywords: Mechanism Design, Bilateral Contracting, Correlated Information, Robust Mechanism Design.

JEL Classification : D82.

# 1 Introduction

Over the last thirty years, mechanism design has been the most powerful tool to understand how complex organizations and institutions are shaped. By means of the Revelation

[1]We thank for comments and discussions seminar participants at University of British Columbia, Lisbon ASSET 2006, Lyon, Universidad Carlos III Madrid, Milan, Marseille LAGV 2006, Northwestern, NICO-Northwestern, Paris-PSE, Paris-Institut Poincaré, Paris Séminaire Roy, Oxford Said Business School, Erasmus-Rotterdam, Toulouse *Spring of Incentives 2009*, the Colin Clarke Lecture in Brisbane ESAM 2007, LAMES Rio 2008, Penn State and Yale. The usual disclaimer applies.

[2]CERDI, Université d'Auvergne, France. e-mail: vianney.dequiedt@u-clermont1.fr

[3]Toulouse School of Economics (IDEI-GREMAQ) and EHESS, France. e-mail: martimor@cict.fr

Principle,[4] this theory characterizes the set of implementable allocations in contexts where information is decentralized and privately known by agents of the organization. Once this first step of the analysis is performed, and a particular optimization criterion is specified at the outset, one can derive an optimal incentive feasible allocation and look for particular institutions (mechanisms) that may implement this outcome.

In the canonical framework for Bayesian collective choices,[5] the principal can agree with his privately informed agents on a grand-mechanism ruled under the aegis of a single mediator (third-party, "machine" or Court of Law). This mediator first collects messages from these agents and second send them recommendations on how to play actions as requested by the mechanism. This paper modifies this mechanism design paradigm to take into account the principal's limited ability to rely on such centralized grand-mechanism. A mechanism is now viewed as a set of separate bilateral contracts linking the principal with each of his agents, each of those contracts being ruled by a separate mediator. What the principal learns when contracting with an agent can be manipulated by the principal himself if he finds it useful in his relationships with others.

The analysis of bilateral contracting is an important step towards improving the match between the theory of mechanism design and the concerns expressed by scholars in more applied fields. Prominent economists like Aoki (1984) and Williamson (1985) have for instance repeatedly pushed the view that the firm should be better considered as a nexus of bilateral contracts rather than a single grand-contract. Those bilateral contracts link various stakeholders of the firms (creditors, shareholders, customers, workers, regulators, etc...) with its management. Thanks to his central role at the nexus of all those contractual relationships, the management finds new strategic possibilities. For instance, each worker contracts separately with the firm, but what he communicates to the management on his own performances is by and large not observed by his peers. Nevertheless, these performances of a given worker can be used strategically by the principal as a subjective measure of those of his peers to determine their compensations. These strategic manipulations might impact in turn on the agents' efforts and incentives.

To analyze such nexus of contracts in an abstract mechanism design environment, the set of incentive feasible mechanisms should account for the fact that the principal might manipulate what he learns from his relationship with a given agent when contracting with others. The broad goal assigned to this paper is precisely to understand how those manipulations affect contract design. Taking into account such manipulations simplifies significantly the characterization of incentive feasible allocations. It also allows us to reach more palatable conclusions on the design of contracts relative to those obtained when assuming that only a centralized grand-mechanism can be enforced. In more details, our

[4] Gibbard (1973) and Green and Laffont (1977) among others.
[5] Myerson (1982 and 1991, Chapter 6.4) for instance.

main results are as follows.

**Characterization of non-manipulable allocations.** In a private values setting (i.e., when the agents' private information does not enter directly into the principal's objectives), a *Revelation Principle under bilateral contracting* characterizes the set of incentive feasible allocations. For a given implementation concept characterizing the agents' behavior (Bayesian-Nash or dominant strategy) there is no loss of generality in restricting the analysis to direct revelation mechanisms which are not manipulated by the principal.

*Non-manipulability constraints* affect nevertheless contract design. To see how, consider an organization with one principal and two agents $A_1$ and $A_2$, each running a different project on the principal's behalf. Agents have private information on their costs. Assume also that there is no productive externality between projects (technologies are separable) but informational externalities do exist (costs are correlated). Private information is costless for the principal if he can design a grand-mechanism in such context.[6] However, this is no longer true when only bilateral contracts are feasible. The principal can always claim that $A_1$'s performances conflict with those of $A_2$ and punish both accordingly. To avoid such manipulations, the compensation of an agent must be less sensitive to what the principal has learned from the other. When agents work on separable projects with only informational externalities, non-manipulability is obtained with simple *sell-out contracts* that give to the principal a payoff independent of the agent's output.

**Optimal mechanisms and rent/efficiency trade-off.** Insisting on non-manipulability restores a genuine trade-off between rent extraction and efficiency even when the agents' types are correlated. Although the scope for yardstick competition is now more limited than with a centralized grand-mechanism, correlated information is still useful when writing bilateral contracts. Correlation makes it easier to extract information rents.

With separable projects, the optimal mechanism trades off the marginal efficiency of the agents' productions with virtual marginal costs that generalize those found in independent types environments. Allowing for more general production externalities between the agents' activities, we characterize non-manipulable contracts and show how they generalize the sell-out contracts found with separable projects.

**Continuity of payoffs and mechanisms.** When a grand-mechanism can be used, privately informed agents get no rent if their types are correlated whereas they do so if types are independent. This lack of continuity of the optimal mechanism with respect to the information structure is a weakness in view of the so-called *"Wilson Doctrine"* which points out that mechanisms should be robust to small perturbations of the modeling. Taking into account non-manipulability constraints restores such continuity. Not only

[6]Crémer and McLean (1985, 1988), McAfee and Reny (1992), Riordan and Sappington (1988), Johnson, Pratt and Zeckhauser (1990), d'Aspremont, Crémer and Gerard-Varet (1990), Matsushima (1991).

the principal's and the agents' payoffs vary continuously with the correlation but also the optimal mechanism keeps the same structure. To illustrate, sell-out contracts are optimal for separable projects both at zero and at a positive correlation.

**Simple bilateral contracting.** A *simple bilateral contract* uses only the corresponding agent's information and not what the principal might learn from others. Such mechanisms are non-manipulable. In Bayesian environments with separable projects, such simple bilateral contracts are dominated by non-manipulable bilateral mechanisms which use that information. By contrast, if dominant strategy and ex post participation constraints are imposed or if collusion between agents matters, simple bilateral contracts are optimal.

**Organization of the paper.** Section 2 presents the model. Section 3 develops a simple example highlighting the fact that the principal's manipulations might constrain significantly mechanisms. Section 4 proves the Revelation Principle with bilateral contracting. Equipped with this tool, we characterize optimal mechanisms for separable projects (Section 5), and general production externalities (Section 6.1). A particular attention is given to production in teams (Section 6.2.1) and multi-unit auctions (Section 6.2.2). For these two cases, the optimal non-manipulable mechanisms are then derived assuming discrete types. Section 7 analyzes various extensions allowing for dominant strategy (Section 7.1), collusion between agents (Section 7.2), secret contracts (Section 7.3) and sequential contracting (Section 7.4). Section 8 discusses the relationship of our work with the relevant literature. Section 9 proposes alleys for further research. All proofs are in an Appendix.

# 2 The Model

**Preferences.** We consider an organization with a principal ($P$) and $n$ agents ($A_i$ for $i = 1, ..., n$). Agent $A_i$ produces a good in quantity $q_i$ on the principal's behalf. Let $q = (q_1, ..., q_n)$ (resp. $t = (t_1, ..., t_n)$) denote the vector of goods (resp. transfers) which belongs to a set $\mathcal{Q} = \Pi_{i=1}^n \mathcal{Q}_i$ where $\mathcal{Q}_i \subset \mathbb{R}_+$ is compact and convex (resp. $\mathcal{T} = \Pi_{i=1}^n \mathcal{T}_i \subset \mathbb{R}^n$). By a standard convention, $A_{-i}$ denotes the set of all agents except $A_i$ and similar notations are used below for other variables.

The principal and his agents have quasi-linear utility functions defined respectively as:

$$V(q,t) = \tilde{S}(q) - \sum_{i=1}^{n} t_i \quad \text{and} \quad U_i(q,t) = t_i - \theta_i q_i.$$

The principal's surplus function $\tilde{S}(\cdot)$ is increasing in each of its arguments $q_i$ and concave in $q$. This formulation encompasses three cases of interest which will receive more attention in the sequel, specifically in organizations involving only two agents.

*Separable projects.* $\tilde{S}(\cdot)$ is separable in both $q_1$ and $q_2$, i.e., $\tilde{S}(q_1, q_2) = S(q_1) + S(q_2)$ for some function $S(\cdot)$ that is increasing and concave with the Inada condition $S'(0) = +\infty$, $S(0) = 0$ and $S'(+\infty) = 0$.

*Perfect substitutability.* $\tilde{S}(\cdot)$ depends on the total production $q_1 + q_2$ only: $\tilde{S}(q_1, q_2) = S(q_1 + q_2)$ for some increasing and concave $S(\cdot)$ which still satisfies the above conditions.

*Perfect complementarity.* $\tilde{S}(\cdot)$ can then be written as $\tilde{S}(q_1, q_2) = S(\min(q_1, q_2))$ where $S(\cdot)$ satisfies again the above conditions.

With separable projects, the only externality between agents is informational and comes from the possible correlation of their costs. Perfect substitutability arises instead in the context of multi-unit auctions. Perfect complementarity occurs in team productions.

**Information.** $A_i$ has private information on his efficiency parameter $\theta_i$. A vector of types is denoted $\theta = (\theta_1, ..., \theta_n) \in \Theta^n$. Importantly, we consider a *private values* environment, i.e., the agents' private information does not enter directly into the principal's objective.[7]

*Continuous distributions.* For most of the paper, we will consider continuous distributions and efficiency parameters belong to a set $\Theta = [\underline{\theta}, \bar{\theta}]$. Types are then jointly drawn from the common knowledge non-negative, bounded and atomless density function $\tilde{f}(\theta)$ whose support is $\Theta^n$. Assuming, for simplicity, symmetric distributions,[8] we will denote the marginal density, the cumulative distribution and the conditional density respectively as

$$f(\theta_i) = \int_{\Theta^{n-1}} \tilde{f}(\theta_i, \theta_{-i}) d\theta_{-i}, \quad F(\theta_i) = \int_{\underline{\theta}}^{\theta_i} f(\theta_i) d\theta_i \quad \text{and} \quad \tilde{f}(\theta_{-i}|\theta_i) = \frac{\tilde{f}(\theta_i, \theta_{-i})}{f(\theta_i)}.$$

*Discrete distributions.* In order to get closed form solutions in some problems below, we shall sometimes assume that each agent's type belongs to $\Theta = \{\underline{\theta}, \bar{\theta}\}$ (denote $\Delta\theta = \bar{\theta} - \underline{\theta}$). The common knowledge distribution of types is still symmetric for simplicity and, in the case of two agents, probabilities for the different type realizations are defined as

$$\tilde{p}(\underline{\theta}, \underline{\theta}) = \nu^2 + \alpha, \tilde{p}(\bar{\theta}, \underline{\theta}) = \tilde{p}(\underline{\theta}, \bar{\theta}) = \nu(1 - \nu) - \alpha, \tilde{p}(\bar{\theta}, \bar{\theta}) = (1 - \nu)^2 + \alpha.$$

The marginal distribution is $p(\underline{\theta}) = \nu$, $p(\bar{\theta}) = 1 - \nu$ and the non-negative correlation coefficient is $\tilde{p}(\underline{\theta}, \underline{\theta})\tilde{p}(\bar{\theta}, \bar{\theta}) - \tilde{p}(\bar{\theta}, \underline{\theta})\tilde{p}(\underline{\theta}, \bar{\theta}) = \alpha \in [0, \nu(1 - \nu)]$.

**Benchmark.** For separable projects, the (symmetric) first-best output requested from each agent trades off the marginal benefit of production against its marginal cost:

$$S'(q^{FB}(\theta_i)) = \theta_i, \quad i = 1, ..., n. \tag{1}$$

With correlated types, this first-best outcome can be either obtained (generically with discrete types) or arbitrarily approached (with a continuum of types) when it is possible

[7] The consequences of this assumption on some of our results will be discussed later.

[8] All our results could be straightforwardly adapted to asymmetric distributions.

to enforce a direct revelation grand-mechanism of the form $\{t_i^{FB}(\hat{\theta}_i, \hat{\theta}_{-i}), q^{FB}(\hat{\theta}_i)\}_{(\hat{\theta}_i, \hat{\theta}_{-i}) \in \Theta^n}$ that makes each agent's payment depend on the whole array of reports. In contrast with intuition, there is no efficiency/rent extraction trade-off in such correlated environments.

Instead with independent types, agents obtain costly information rents and a genuine trade-off between efficiency and rent extraction arises. The marginal benefit of production must balance its *virtual marginal cost.* With separable projects and a continuous distribution, the (symmetric) second-best output is given by the so-called Baron-Myerson outcome:[9]

$$S'(q^{BM}(\theta_i)) = \theta_i + \frac{F(\theta_i)}{f(\theta_i)}, \quad \forall \theta_i \in \Theta = [\underline{\theta}, \bar{\theta}].^{10} \tag{2}$$

**Remark 1** *The Baron-Myerson outcome is also obtained, even if types are correlated, when the principal uses with each agent a simple bilateral contract of the form* $\{t_i(\hat{\theta}_i), q_i(\hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ *that depends only on this agent's announcement on his type. The discrepancy between (1) and (2) measures then the efficiency loss incurred when a grand-mechanism is replaced by n such simple bilateral contracts.*

# 3 A Simple Example

To give some overview of the analysis we develop below, let us consider the case of separable projects with only two agents having types drawn in a discrete type distribution.

Suppose that the principal offers a grand-mechanism with the following payments

$$t_i^{FB}(\theta_i, \theta_{-i}) = S(q^{FB}(\theta_i)) - h(\theta_i, \theta_{-i}).$$

This mechanism yields payoff $h(\theta_i, \theta_{-i})$ to the principal when dealing with agent $A_i$, implements the first-best outputs and extracts all agents' surplus provided that $h(\cdot)$ satisfies the Bayesian incentive constraints (for types $\underline{\theta}$ and $\bar{\theta}$ respectively)

$$\begin{aligned}
\left(\nu + \frac{\alpha}{\nu}\right) h(\bar{\theta}, \underline{\theta}) + \left(1 - \nu - \frac{\alpha}{\nu}\right) h(\bar{\theta}, \bar{\theta}) &\geq S(q^{FB}(\bar{\theta})) - \underline{\theta} q^{FB}(\bar{\theta}), \\
\left(\nu - \frac{\alpha}{1-\nu}\right) h(\underline{\theta}, \underline{\theta}) + \left(1 - \nu + \frac{\alpha}{1-\nu}\right) h(\underline{\theta}, \bar{\theta}) &\geq S(q^{FB}(\underline{\theta})) - \bar{\theta} q^{FB}(\underline{\theta})
\end{aligned}$$

[9] Baron and Myerson (1982).

[10] Provided that the *Monotone Hazard Rate Property* holds, namely $\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right) > 0$, $q^{BM}(\theta_i)$ is indeed the optimal output. Otherwise, bunching may arise (Guesnerie and Laffont 1984, and Laffont and Martimort 2002, Chapter 3). For a discrete distribution with two types, such bunching does not arise in our framework.

and the following interim participation constraints as equalities

$$\begin{aligned}\left(\nu+\frac{\alpha}{\nu}\right)h(\underline{\theta},\underline{\theta})+\left(1-\nu-\frac{\alpha}{\nu}\right)h(\underline{\theta},\bar{\theta}) &= S(q^{FB}(\underline{\theta}))-\underline{\theta}q^{FB}(\underline{\theta}),\\ \left(\nu-\frac{\alpha}{1-\nu}\right)h(\bar{\theta},\underline{\theta})+\left(1-\nu+\frac{\alpha}{1-\nu}\right)h(\bar{\theta},\bar{\theta}) &= S(q^{FB}(\bar{\theta}))-\bar{\theta}q^{FB}(\bar{\theta}).\end{aligned}$$

Using Farkas' Lemma, it is straightforward to show that such $h(\cdot)$ exists as soon as types are correlated (i.e., $\alpha > 0$). Moreover, the above constraints imply also that

$$h(\bar{\theta},\underline{\theta})-h(\bar{\theta},\bar{\theta}) \geq \frac{\nu(1-\nu)}{\alpha}\Delta\theta q^{FB}(\bar{\theta}) > 0.^{11} \tag{3}$$

Implicit in the description of such direct revelation grand-mechanism is the fact that there exists a mediator $\mathfrak{M}$ who first collects the agents' messages $(\hat{\theta}_i, \hat{\theta}_{-i})$ and second enforces for agent $A_i$ an output $q^{FB}(\hat{\theta}_i)$ and a payment $t_i^{FB}(\hat{\theta}_i, \hat{\theta}_{-i}) = S(q^{FB}(\hat{\theta}_i)) - h(\hat{\theta}_i, \hat{\theta}_{-i})$ as requested by the mechanism.

Suppose now that the organization is no longer ruled by such a single grand-mechanism but by two bilateral contracts (thereafter called *sub-mechanisms*) $\{t_i^{FB}(\hat{\theta}_i, \tilde{\theta}_{-i}), q^{FB}(\hat{\theta}_i)\}_{(\hat{\theta}_i, \tilde{\theta}_{-i})\in\Theta^2}$ for $i = 1, 2$. Each of those bilateral contracts is under the aegis of a different mediator (or Court of Law) $\mathfrak{M}_i$ $(i = 1, 2)$. Sub-mechanisms work as follows. Agent $A_i$ first announces his type $\hat{\theta}_i$ in sub-mechanism $i$. Those reports are learned by the principal who then makes a report $\tilde{\theta}_{-i}$ back to mediator $\mathfrak{M}_i$ on what he has learned in the relationship with $A_{-i}$. Those principal's reports are of course manipulable. Given the timing of communication, the principal cannot commit not to manipulate $\tilde{\theta}_{-i}$ to reduce agent $A_i$'s payment $t_i^{FB}(\hat{\theta}_i, \tilde{\theta}_{-i}) = S(q_i(\hat{\theta}_i, \tilde{\theta}_{-i})) - h(\hat{\theta}_i, \tilde{\theta}_{-i})$. Indeed, (3) shows for instance that the principal wants to claim that a type $\bar{\theta}$ agent $A_i$ faces a type $\underline{\theta}$ agent $A_{-i}$ to reap greater revenues. With bilateral contracting, those claims by the principal towards different mediators do not conflict with each other as long as agent $A_i$ and mediator $\mathfrak{M}_i$ have no way (even an indirect one) to observe and verify agent $A_{-i}$'s own report to $\mathfrak{M}_{-i}$.

Several interesting insights can already be gleaned from this example.

1. To avoid the principal's manipulations, $A_i$'s payments could be made independent of the principal's claim $\tilde{\theta}_{-i}$, i.e.

$$h(\theta_i, \underline{\theta}) = h(\theta_i, \bar{\theta}) \qquad \forall \theta_i \in \Theta = \{\underline{\theta}, \bar{\theta}\}.$$

   Imposing a priori this *non-manipulability constraint*, and still requiring that $A_i$'s output depends only on his own type, brings us back to the traditional screening

[11] Indeed, it cannot be that $h(\bar{\theta},\underline{\theta}) = h(\bar{\theta},\bar{\theta}) = S(q^{FB}(\bar{\theta}))-\bar{\theta}q^{FB}(\bar{\theta})$. If it was so, the Bayesian incentive compatibility of a type $\underline{\theta}$ would be violated as it can be easily seen.

model for which the Baron-Myerson outcome with outputs $q^{BM}(\underline{\theta}) = q^{FB}(\underline{\theta})$ and $q^{BM}(\underline{\theta})$ given by

$$S'(q^{BM}(\bar{\theta})) = \bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta \tag{4}$$

is optimal. In the sequel, we will see that this solution is too extreme. Optimal non-manipulable contracting requires generally less stark distortions on outputs.

2. Under bilateral contracting, $A_{-i}$'s output certainly cannot be used to regulate the relationship between $A_i$ and the principal. Indeed, the fact that $A_i$ and $\mathfrak{M}_i$ cannot observe and verify $A_{-i}$'s own report to $\mathfrak{M}_{-i}$ implies also that the latter's output $q_{-i}$ is itself non-verifiable from their point of view. Otherwise, they could have indirectly inferred $A_{-i}$'s report from his realized output. In such contexts, menus of nonlinear prices of the form $\{T_i(q_i, \hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ form a quite attractive set of mechanisms to focus on.[12] With such menus, everything happens then as if $A_i$ chooses first a particular nonlinear scheme $T_i(q_i, \hat{\theta}_i)$. After having learned those reports by both agents, the principal finally chooses the outputs $(q_i(\hat{\theta}_i, \hat{\theta}_{-i}), q_{-i}(\hat{\theta}_i, \hat{\theta}_{-i}))$ to maximize his payoff.

3. This interpretation in terms of nonlinear prices makes it clear what sort of limits on commitment arises when moving from a grand-mechanism to a pair of bilateral sub-mechanisms. Suppose indeed, that such a pair $\{T_i(q_i, \hat{\theta}_i), T_{-i}(q_{-i}, \hat{\theta}_{-i})\}_{(\hat{\theta}_i, \hat{\theta}_{-i}) \in \Theta^2}$ of menus was offered under the aegis of a single mediator $\mathfrak{M}$. With such centralized mechanism, the mediator commits to implement $(q_i(\hat{\theta}_i, \hat{\theta}_{-i}), q_{-i}(\hat{\theta}_i, \hat{\theta}_{-i}))$ upon receiving reports $(\hat{\theta}_i, \hat{\theta}_{-i})$. Instead, such commitment is no longer possible under bilateral contracting. Our characterization of non-manipulable mechanisms below imposes the sequential rationality requirement that the principal optimally chooses outputs after having learned the agents' reports.[13]

# 4 Characterizing Non-Manipulable Mechanisms

## 4.1 Mechanisms

Let us describe more formally the class of mechanisms available in a bilateral contracting environment. A bilateral contract with a given agent can, in full generality, use the principal's report on any information that he may get by contracting with others. Manipulations by the principal may arise because what a given agent communicates to the principal is not observed by others who will only learn that information from the principal.

[12] We will see below in Section 4.4 that it is essentially the unique such set.

[13] The reader will recognize that this sequential rationality is also a feature of all the common agency literature. We develop this relationship in Section 8.

Formally, any general mechanism is now a pair $(g(\cdot), \mathcal{M})$ where $g(\cdot)$ is an outcome function and $\mathcal{M} = \Pi_{i=1}^n \mathcal{M}_i$ is the product space of the respective communication spaces $\mathcal{M}_i$ available to agent $A_i$ to communicate with the principal. To capture the fact that the principal plays different sub-mechanisms with each of his agents, the outcome function $g(\cdot)$ is itself decomposed into $n$ different outcome functions $g(\cdot) = (g_1(\cdot), ..., g_n(\cdot))$. Each sub-mechanism $g_i(\cdot)$ maps $\mathcal{M} = \mathcal{M}_i \times \mathcal{M}_{-i}$ into the set $\Delta(\mathcal{Q}_i \times \mathcal{T}_i)$ of (possibly random) allocations for agent $A_i$. When playing the sub-mechanism $(g_i(\cdot), \mathcal{M})$, $A_i$ sends message $m_i$ to a mediator $\mathfrak{M}_i$. Such communication is observed by $P$. Then, the principal sends a message $\hat{m}_{-i}$ to $\mathfrak{M}_i$ on whatever information he may have learned in observing the reports made by agents $A_{-i}$ in the sub-mechanisms $(g_{-i}(\cdot), \mathcal{M})$. Finally, the requested transfer $t_i(m_i, \hat{m}_{-i})$ and output $q_i(m_i, \hat{m}_{-i})$ for agent $A_i$ are implemented.[14] Because of production and/or informational externalities, $A_i$'s allocation should depend in full generality on the report $\hat{m}_{-i}$ made by the principal on the messages $m_{-i}$ sent by other agents $A_{-i}$ and observed by that principal.[15]

**Remark 2** *Standard mechanism design assumes that a unique mediator $\mathfrak{M}$ keeps one party's message secret from the other when running a centralized grand-mechanism. Instead, we suppose that, although agent $A_i$ only observes the messages $m_i$ he sends to the mediator $\mathfrak{M}_i$ ruling the sub-mechanism $g_i(\cdot)$, $P$ observes the whole array of messages $m = (m_1, ..., m_n)$ before communicating back to mediators in each sub-mechanism. This assumption is justified whenever mediators are not machine but may have their own financial objectives and may collude with the principal to share information they have gathered from the agents. Alternatively, this amounts to assuming that the only possible mediator available is the principal himself.[16] Under both interpretations, mediators make whatever information they learned from each agent available to the principal.[17] Finally, we also assume that agent $A_i$ and the mediator $\mathfrak{M}_i$ do not observe either the report $m_{-i}$ made*

[14] When allocations are random, $q_i(m_i, \hat{m}_{-i})$ and $t_i(m_i, \hat{m}_{-i})$ are viewed as distributions of outputs and transfers. With obvious notations, payoffs should be understood as expectations over those distributions.

[15] Because of bilateral contracting, the principal may a priori send different messages concerning what he learned from a third agent in two different sub-mechanisms.

[16] That there is no single benevolent mediator having access to all the agents' reports leaves to the sole principal a strategic role at a nexus of all communication channels. This "incomplete contracting" assumption is standard in the literature on vertical contracting and common agency (see for instance Martimort, 2007). We also rule out any cheap talk communication among agents that could help them replicate by themselves the existence of a missing mediator (Barany 1992, Forges 1990, Gerardi 2002).

[17] Suppose alternatively that mediators design private communication channels with each agent and keep their reports secret. There would be no scope for the principal communicating back in each sub-mechanism because he would not have observed the agents' reports in other sub-mechanisms. Only simple bilateral contracts (see Remark 3 below for their definition) are then feasible. Those simple bilateral contracts are most of the time suboptimal (as we show below) since they do not allow the principal to benefit from any informational or production externalities between agents. In other words, if the principal could choose ex ante between having separate mediators running sub-mechanisms entertaining private communication only or having sub-mechanisms making agents' reports available to him, he would choose the latter mode of bilateral contracting.

*by $A_{-i}$ into $g_{-i}(\cdot)$ or the realized trades $(q_{-i}, t_{-i})$ and infer from this (at least partially) whether the claim $\hat{m}_{-i}$ that the principal has made fits with the reports $m_{-i}$ made by $A_{-i}$.*[18]

**Remark 3** *A simple bilateral contract corresponds to an outcome function $g_i^*(\cdot)$ which only maps agent $A_i$'s communication space $\mathcal{M}_i$ into $\Delta(\mathcal{Q}_i \times \mathcal{T}_i)$. With such contract, there is no scope for using what the principal has learned from observing $A_{-i}$'s messages to improve $A_i$'s allocation. Those contracts are clearly non-manipulable.*

**Remark 4** *For minimal departure from standard mechanism design, we assume that the mechanism $(g(\cdot), \mathcal{M})$ is publicly observable by all agents. Assuming private sub-mechanisms, i.e., that $A_i$ does not observe the sub-mechanisms $(g_{-i}(\cdot), \mathcal{M})$, introduces another dimension of private information in our model: the principal being now privately informed on contractual deals made with them. We investigate this issue in Section 7.3.*

## 4.2 Timing

Summarizing, the contracting game unfolds as follows. First, agents privately learn their respective efficiency parameters. Second, the principal offers a mechanism $(g(\cdot), \mathcal{M})$ to the agents. Third, agents simultaneously accept or refuse their respective sub-mechanisms $(g_i(\cdot), \mathcal{M})$. If agent $A_i$ refuses, he gets a payoff normalized to zero. Fourth, agents simultaneously send messages $m_i$ in their respective sub-mechanisms $g_i(\cdot)$. Fifth, the principal reports in his contractual relationship with $A_i$ a message $\hat{m}_{-i}$ on what he has learned from contracting with $A_{-i}$. Finally, agent $A_i$'s outputs and transfers are implemented according to the messages $(m_i, \hat{m}_{-i})$ and the outcome function $g_i(\cdot)$.

The equilibrium concept is perfect Bayesian equilibrium (thereafter PBE).[19]

**Remark 5** *Note that contracting with each agent is simultaneous. We further discuss sequential contracting in Section 7.4.*

[18] One could think of less extreme situations where each agent may get a signal correlated with what the others are privately reporting to the principal. Of course, if this signal is public and verifiable, contingent mechanisms could be written to help circumvent the privacy problem. However, if this signal is only privately observed and can be manipulated, such contingent mechanisms lose again their force. In Section 6.2.1, we analyze a team production context with perfect complementarity in the agents' efforts. There the realized output plays the role of an ex post information on other's reports that limits the principal's set of feasible manipulations.

[19] Except in Section 7.1 where dominant strategy implementation characterizes the agents' behavior.

## 4.3 Revelation Principle

We now fully characterize the set of allocations that can be achieved as equilibria of the overall contracting game where the principal offers any possible mechanism $(g(\cdot), \mathcal{M})$ in this private values context.

For any agents' reporting strategy $m^*(\cdot) = (m_1^*(\cdot), ..., m_n^*(\cdot))$, let $\sup m_i^*(\cdot)$ denote the support of $m_i^*(\cdot)$, i.e., the set of messages $m_i$ that are sent with positive probability by $A_i$ given $m_i^*(\cdot)$. For a given mechanism with bilateral contracting $(g(\cdot), \mathcal{M})$ accepted by all types of agents, a continuation equilibrium induced by such mechanism at the communication stage is described as follows.

**Lemma 1** *Fix any arbitrary mechanism $(g(\cdot), \mathcal{M})$ accepted by all types of agents. A continuation equilibrium is a pair $\{m^*(\cdot), \hat{m}^*(\cdot)\}$ such that:*

- *The agents' strategy vector $m^*(\theta) = (m_1^*(\theta_1), ..., m_n^*(\theta_n))$ from $\Theta^n$ into $\mathcal{M} = \Pi_{i=1}^n \mathcal{M}_i$ forms a Bayesian equilibrium given the principal's optimal manipulation $\hat{m}^*(\cdot)$*

$$m_i^*(\theta_i) \in \arg\max_{m_i \in \mathcal{M}_i} E_{\theta_{-i}} \left( t_i(m_i, \hat{m}_{-i}^*(m_i, m_{-i}^*(\theta_{-i}))) - \theta_i q_i(m_i, \hat{m}_{-i}^*(m_i, m_{-i}^*(\theta_{-i}))) | \theta_i \right); \tag{5}$$

- *The principal's optimal manipulation $\hat{m}^*(\cdot) = (\hat{m}_{-1}^*(\cdot), ..., \hat{m}_{-n}^*(\cdot))$ from $\mathcal{M}$ on $\Pi_{i=1}^n \mathcal{M}_{-i}$ satisfies $\forall m = (m_1, ..., m_n) \in \mathcal{M}$*

$$(\hat{m}_{-1}^*(m), ..., \hat{m}_{-n}^*(m))$$

$$\in \arg\max_{(\hat{m}_{-1}, ..., \hat{m}_{-n}) \in \Pi_{i=1}^n \mathcal{M}_{-i}} \tilde{S}(q_1(m_1, \hat{m}_{-1}), ..., q_n(m_n, \hat{m}_{-n})) - \sum_{i=1}^n t_i(m_i, \hat{m}_{-i}). \tag{6}$$

Given a mechanism $(g(\cdot), \mathcal{M})$, a continuation equilibrium induces an allocation $a(\theta) = g(m^*(\theta), \hat{m}^*(m^*(\theta)))$ which maps $\Theta^n$ into $\Delta(\mathcal{Q} \times \mathcal{T})$. In this private values context, updated beliefs held by the principal following the agents' reports $m^*(\theta)$ do not influence his optimal manipulation. This can be seen more precisely on equation (6) which is written ex post, i.e., for each realization of the agents' reports.[20]

The following definitions are useful.

**Definition 1** *A mechanism $(g(\cdot), \mathcal{M})$ is non-manipulable if and only if $\hat{m}_{-i}^*(m) = m_{-i}$, for all $m \in \sup m^*(\cdot)$ and $i$ at a continuation equilibrium.*[21]

[20] Assuming private values simplifies significantly the analysis by avoiding any signaling issue when agents communicate their types.

[21] Note that our concept of non-manipulability is weak and that we do not impose the more stringent requirement that the mechanism is non-manipulable at all continuation equilibria.

**Definition 2** *A direct mechanism* $(\bar{g}(\cdot), \Theta^n)$ *is truthful if and only if* $m^*(\theta) = (\theta_1, \theta_2, ..., \theta_n)$, *for all* $\theta \in \Theta$ *at a continuation equilibrium.*

**Proposition 1** ***Revelation Principle with Bilateral Contracting.*** *In a private values context, any allocation* $a(\cdot)$ *achieved at a continuation equilibrium of any arbitrary mechanism* $(g(\cdot), \mathcal{M})$ *with bilateral contracting can also be implemented through a truthful and non-manipulable direct mechanism* $(\bar{g}(\cdot), \Theta^n)$.

With such direct revelation mechanisms, the agents' Bayesian incentive compatibility constraints are written as usual:

$$\underset{\theta_{-i}}{E}\left(t_i(\theta_i, \theta_{-i}) - \theta_i q_i(\theta_i, \theta_{-i})|\theta_i\right) \geq \underset{\theta_{-i}}{E}\left(t_i(\hat{\theta}_i, \theta_{-i}) - \theta_i q_i(\hat{\theta}_i, \theta_{-i})|\theta_i\right) \quad \forall(\theta_i, \hat{\theta}_i) \in \Theta^2. \quad (7)$$

The following *non-manipulability* constraints stipulate that the principal does not misrepresent what he has learned from others' reports in his relationship with any agent:

$$\tilde{S}(q(\theta)) - \sum_{i=1}^{n} t_i(\theta) \geq \tilde{S}(q_1(\theta_1, \hat{\theta}_{-1}), ..., q_n(\theta_n, \hat{\theta}_{-n})) - \sum_{i=1}^{n} t_i(\theta_i, \hat{\theta}_{-i}), \quad \forall(\theta, \hat{\theta}_{-1}, ..., \hat{\theta}_{-n}). \quad (8)$$

In the sequel, we analyze the impact of the non-manipulability constraints (8) on optimal mechanisms in different contexts involving two agents. Those constraints become

$$\tilde{S}(q(\theta)) - \sum_{i=1}^{2} t_i(\theta) \geq \tilde{S}(q_1(\theta_1, \hat{\theta}_2), q_2(\hat{\theta}_1, \theta_2)) - \sum_{i=1}^{2} t_i(\theta_i, \hat{\theta}_{-i}), \quad \forall(\theta, \hat{\theta}_1, \hat{\theta}_2). \quad (9)$$

## 4.4 Taxation Principle

We could have started with nonlinear prices as primitives of our analysis, i.e., sub-mechanisms $g_i(\cdot)$ mapping $\Theta$ into $\mathfrak{T}_i = \{T_i(\cdot) : \mathcal{Q}_i \to \mathcal{T}_i\}$. Everything happens then as if $A_i$ picks first one such nonlinear price within the family $\{T_i(q_i, \hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$, and the principal optimally chooses afterwards the particular output $q_i$ and therefore the corresponding transfers $T_i(q_i, \hat{\theta}_i)$ that this agent receives. In other words, the constraints imposed by the non-manipulability of the mechanisms are akin to assuming that the principal can commit to offer menus of nonlinear prices $\{T_i(\cdot, \hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ to his agents in the first place *but* cannot commit to a particular output schedule $\{q(\hat{\theta})\}_{\hat{\theta} \in \Theta^n}$. Outputs will be chosen after the agents' reports are known.

Focusing on sub-mechanisms of the form $\{T_i(q_i, \hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ is in fact without loss of generality under bilateral contracting. Starting indeed from any non-manipulable direct revelation mechanism, we may define the nonlinear price $T_i(q_i, \theta_i)$ as $T_i(q_i, \theta_i) = t_i(\theta_i, \theta_{-i})$ for $q_i = q_i(\theta_i, \theta_{-i})$. This definition is non-ambiguous since (8) implies that any $\theta_{-i}$ such that $q_i(\theta_i, \theta_{-i}) = q_i$ corresponds to the same transfer $t_i(\theta_i, \theta_{-i}) = t_i$. Written in terms of those nonlinear prices, the non-manipulability constraints (8) become

$$q(\theta) \in \arg\max_{q \in \mathcal{Q}} \tilde{S}(q) - \sum_{i=1}^{n} T_i(q_i, \theta_i). \tag{10}$$

# 5 Separable Projects

Let us start with the simplest setting with only two agents working each on a different project without any production externality between those projects. The principal's gross surplus function is separable and writes as $\tilde{S}(q_1, q_2) = \sum_{i=1}^{2} S(q_i)$. This case provides a useful benchmark to understand how non-manipulability constraints affect contract design when only informational externalities between agents matter.

From the non-manipulability constraints (9), we immediately get that for any pair $(\theta_{-i}, \theta'_{-i})$ we must have:

$$S(q_i(\theta_i, \theta_{-i})) - t_i(\theta_i, \theta_{-i}) \geq S(q_i(\theta_i, \theta'_{-i})) - t_i(\theta_i, \theta'_{-i})$$

and

$$S(q_i(\theta_i, \theta'_{-i})) - t_i(\theta_i, \theta'_{-i}) \geq S(q_i(\theta_i, \theta_{-i})) - t_i(\theta_i, \theta_{-i}).$$

Therefore, there exist a function $H_i(\theta_i)$ such that:

$$S(q_i(\theta_i, \theta_{-i})) - t_i(\theta_i, \theta_{-i}) = H_i(\theta_i) \tag{11}$$

This direct mechanism can be transformed into a nonlinear price $T_i(q_i, \theta_i) = t_i(\theta_i, \theta_{-i})$ for $q_i(\theta_i, \theta_{-i}) = q_i$. Such nonlinear price corresponds to a simple *sell-out contract* of the form

$$T_i(q_i, \theta_i) = S(q_i) - H_i(\theta_i). \tag{12}$$

With such scheme, agent $A_i$ pays upfront a fixed-fee $H_i(\theta_i)$ to produce on the principal's behalf. The principal, once informed on all agents' reports, chooses an output and agent $A_i$ gets the corresponding benefit $S(q_i)$ on the project he is running. The principal's payoff coming from his relationship with $A_i$ is $H_i(\theta_i)$ which does not depend on the agent's output. These fixed-fees are chosen so that the mechanism is incentive compatible and all types, even the least efficient one, participate.[22]

Let us denote by $U_i(\theta_i)$ the information rent of an agent $A_i$ with type $\theta_i$:

$$U_i(\theta_i) = \underset{\theta_{-i}}{E} \left( S(q_i(\theta_i, \theta_{-i})) - \theta_i q_i(\theta_i, \theta_{-i}) | \theta_i \right) - H_i(\theta_i). \tag{13}$$

[22]Shutting down the least efficient types is never optimal given the Inada condition $S'(0) = +\infty$.

Individual rationality implies:

$$U_i(\theta_i) \geq 0 \qquad \forall i, \quad \forall \theta_i \in \Theta. \tag{14}$$

Bayesian incentive compatibility requires:

$$U_i(\theta_i) = \arg\max_{\hat{\theta}_i \in \Theta_i} \underset{\theta_{-i}}{E}\left(S(q_i(\hat{\theta}_i, \theta_{-i})) - \theta_i q_i(\hat{\theta}_i, \theta_{-i})|\theta_i\right) - H_i(\hat{\theta}_i) \quad \forall i, \ \forall \theta_i \in \Theta. \tag{15}$$

What is remarkable here is the similarity of this formula with the Bayesian incentive constraint that would be obtained had types been independently distributed. In that case, the agent's expected payment is independent of his true type and can also be separated in the expression of the incentive constraint exactly as the function $H_i(\cdot)$ in (15). This renders the analysis of the set of non-manipulable incentive compatible allocations close to what modelers are used to do in standard mechanism design with independent types.

Assume for simplicity that $q_i(\cdot)$ is differentiable.[23] Simple revealed preferences arguments show that $H_i(\cdot)$ is also differentiable. The local first-order condition for Bayesian incentive compatibility becomes

$$\dot{H}_i(\theta_i) = \underset{\theta_{-i}}{E}\left((S'(q_i(\theta_i, \theta_{-i})) - \theta_i)\frac{\partial q_i}{\partial \theta_i}(\theta_i, \theta_{-i})|\theta_i\right) \quad \forall i, \ \forall \theta_i \in \Theta. \tag{16}$$

Consider any output schedule $q_i(\cdot)$ which is monotonically decreasing in $\theta_i$ and downward distorted below the first-best. From (16), $H_i(\cdot)$ is necessarily also decreasing in $\theta_i$. Less efficient types produce less and pay lower up-front payments. The incentive constraint (16) captures the trade-off faced by a type $\theta_i$ agent. By overreporting, this agent pays a lower up-front payment but he also produces less and enjoys a lower expected surplus.

To highlight the trade-off between efficiency and rent extraction, it is useful to rewrite incentive compatibility in terms of the agents' information rent. Equation (16) becomes:

$$\dot{U}_i(\theta_i) = -\underset{\theta_{-i}}{E}(q_i(\theta_i, \theta_{-i})|\theta_i) + \underset{\theta_{-i}}{E}\left((S(q_i(\theta_i, \theta_{-i})) - \theta_i q_i(\theta_i, \theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right). \tag{17}$$

To better understand the right-hand side of (17), consider an agent with type $\theta_i$ mimicking a less efficient type $\theta_i + d\theta_i$. By doing so, the $\theta_i$ agent produces the same amount than the $\theta_i + d\theta_i$ one but at a lower marginal cost. This gives to type $\theta_i$ a first source of information rent which is worth the first term on this right-hand side. By mimicking the $\theta_i + d\theta_i$ type, a $\theta_i$ agent $A_i$ affects also how the principal interprets the

[23]Because conditional expectations depend on $A_i$'s type, one cannot derive from revealed preferences arguments that either $q_i(\cdot)$ or $E(q_i(\cdot)|\theta_i)$ is itself monotonically decreasing in $\theta_i$. However, it is possible to use the envelope theorem in integral form (Milgrom and Segal, 2002), to characterize the rent obtained by the agents without assuming differentiability of $q_i$. The assumption of differentiability is used here only to gain better intuition on (16). See the proof of Proposition 2 for details.

other agent's report to adjust $A_i$'s own production. The corresponding marginal rent is the second term on the right-hand side of (17). It may in fact be either positive or negative. Some intuition is provided below after having derived the optimal mechanism.

Finally, the local second-order condition for incentive compatibility can be written as:

$$-\underset{\theta_{-i}}{E}\left(\frac{\partial q_i}{\partial \theta_i}(\theta_i,\theta_{-i})|\theta_i\right)+\underset{\theta_{-i}}{E}\left((S'(q_i(\theta_i,\theta_{-i}))-\theta_i)\frac{\partial q_i}{\partial \theta_i}(\theta_i,\theta_{-i})\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right)\geq 0$$
$$\forall i=1,2,\ \forall \theta_i\in\Theta. \quad (18)$$

*A regular incentive problem* is such that the agent's first-order condition (16) is both necessary and sufficient for the optimality of a truthful strategy and the right-hand side of (17) is negative so that countervailing incentives do not arise.

The optimal non-manipulable allocation $\{(q_i(\theta),U_i(\theta_i))_{i=1,2}\}$ solves:

$$(\mathcal{P}):\max_{\{(q_i(\theta),U_i(\theta_i))_{i=1,2}\}}\quad \underset{\theta}{E}\left(\sum_{i=1}^{2}S(q_i(\theta))-\theta_i q_i(\theta)-U_i(\theta_i)\right)\text{ subject to (14) to (18).}^{24}$$

To get sharp predictions on the solution, we need to generalize to environments with correlated information the well-known assumption of monotonicity of the virtual cost:

**Assumption 1** *The generalized virtual cost* $\varphi(\theta_i,\theta_{-i})=\theta_i+\frac{\frac{F(\theta_i)}{f(\theta_i)}}{1+\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\frac{F(\theta_i)}{f(\theta_i)}}$ *is strictly increasing in* $\theta_i$ *and decreasing in* $\theta_{-i}$.

The monotonicity of $\varphi(\theta_i,\theta_{-i})$ in $\theta_i$ ensures that optimal outputs are non-increasing with own types, a condition which is neither sufficient nor necessary for implementability as it can be seen from (18) but which remains a useful ingredient for it. Assumption 1 implies also the Monotone Likelihood Ratio Property (MLRP) $\frac{\partial}{\partial\theta_{-i}}\left(\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\right)\geq 0$ for all $\theta\in\Theta^2$.

Also, we assume that there is an upper bound on the possible level of correlation expressed by the following condition:

**Assumption 2**

$$\left|\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\right|\leq\min\left\{\frac{f(\theta_i)}{F(\theta_i)},\frac{q^{BM}(\theta_i)}{S(q^{FB}(\theta_i))-\theta_i q^{FB}(\theta_i)}\right\}\ \textit{for all}\ \theta\in\Theta^2,$$

[24] With correlated types, the local second-order conditions (18) are not sufficient to guarantee global incentive compatibility even if the agents' utility function satisfies a Spence-Mirrlees condition. However, this is the case if the correlation is small enough as requested by Assumption 2 below. See the proof of Proposition 2 in the Appendix for details.

*and*

$$\max_{(\theta_i,\theta_{-i})\in\Theta^2}\left|\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)\right| \leq \min_{\theta_i\in\Theta} f(\theta_i)\frac{(\min_{(\theta_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\theta_i))^2}{2\max_{(\theta_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\theta_i)}.$$

Assumption 2 ensures that the incentive problem is regular as defined above.[25]

**Proposition 2** *Assume that Assumptions 1 and 2 both hold, continuous types and projects are separable (i.e.,* $\frac{\partial^2\tilde{S}}{\partial q_1\partial q_2}=0$*). The agents' incentive problems are regular and the optimal non-manipulable Bayesian mechanism entails:*

• *A downward output distortion* $q^{SB}(\theta_i,\theta_{-i})$ *which satisfies the following "modified Baron-Myerson" formula*

$$S'(q^{SB}(\theta_i,\theta_{-i}))=\varphi(\theta_i,\theta_{-i}), \tag{19}$$

*with "no distortion at the top"* $q^{SB}(\underline{\theta},\theta_{-i})=q^{FB}(\underline{\theta},\theta_{-i}),\quad \forall\theta_{-i}\in\Theta$ *and the following monotonicity conditions*

$$\frac{\partial q^{SB}}{\partial\theta_{-i}}(\theta)\geq 0 \quad \textit{and} \quad \frac{\partial q^{SB}}{\partial\theta_i}(\theta)<0;$$

• *Agents always get a positive rent except for the least efficient ones*

$$U_i^{SB}(\theta_i)\geq 0 \quad (\textit{with } =0 \textit{ at } \theta_i=\bar{\theta}).$$

As already stressed, Bayesian incentive constraints with non-manipulability look very similar to what they are with independent types. This suggests that the trade-off between

[25] As an example, consider the bivariate normal distribution truncated on $[\theta_0-\lambda\sigma^2,\theta_0+\lambda\sigma^2]^2$ with density

$$\tilde{f}(\theta_1,\theta_2)=\frac{C(\rho,\lambda\sigma^2)}{2\pi\sigma^2(1-\rho^2)^{\frac{1}{2}}}exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(\theta_1-\theta_0)^2}{\sigma^2}+\frac{(\theta_2-\theta_0)^2}{\sigma^2}-2\frac{\rho}{\sigma^2}(\theta_1-\theta_0)(\theta_2-\theta_0)\right)\right].$$

The case $\rho=0$ corresponds to independent types. For $\rho$ small enough, we have up to terms of order at least $\rho^2$ : $C(\rho,\lambda\sigma^2)=(\Phi(\lambda)-\Phi(-\lambda))^{-2}+o(\rho^2)$ where $\Phi(x)$ is the cumulative of the standard normal distribution. Using this property, we derive successively:

$$\tilde{f}(\theta_1,\theta_2)=\frac{1}{2\pi\sigma^2\left(\Phi(\lambda)-\Phi(-\lambda)\right)^2}exp\left[-\frac{(\theta_1-\theta_0)^2}{2\sigma^2}-\frac{(\theta_2-\theta_0)^2}{2\sigma^2}\right]\left(1+\frac{\rho}{\sigma^2}(\theta_1-\theta_0)(\theta_2-\theta_0)\right)+o(\rho^2)$$

and

$$\tilde{f}(\theta_1)=\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}\left(\Phi(\lambda)-\Phi(-\lambda)\right)}exp\left(-\frac{(\theta_1-\theta_0)^2}{2\sigma^2}\right)+o(\rho^2),$$

i.e., each cost is approximatively distributed according to a truncated normal distribution. Finally, the likelihood ratio

$$\frac{\tilde{f}_{\theta_1}(\theta_2|\theta_1)}{\tilde{f}(\theta_2|\theta_1)}=\frac{\rho}{\sigma^2}(\theta_2-\theta_0)+o(\rho^2),$$

satisfies MLRP and conditions in Assumption 2 are verified when $\rho$ is small enough.

efficiency and rent extraction that occurs under independent types carries over here also even with correlation. This intuition is confirmed by equation (19) which highlights the output distortion capturing this trade-off.

With independent types, the right-hand sides of (2) and (19) are the same. The principal finds useless the report of an agent to better design the other agent's incentives. He must give up some information rent to induce information revelation anyway. Outputs are distorted downward to reduce those rents and the standard Baron-Myerson distortions follow. The optimal mechanism with separable projects and independent types can be implemented with simple bilateral contracts which are *de facto* non-manipulable by the principal. Non-manipulability has no bite in this case.

When types are instead correlated, a similar logic to that of Section 3 applies here with an added twist. Indeed, in our earlier example, because we required that $A_i$'s output depends only on $A_i$'s type, non-manipulability puts only a restriction on transfers. More generally, non-manipulability imposes only that the principal's payoff remains constant over all possible transfer-output pairs offered to an agent. This still allows the principal to link agent $A_i$'s payment to what he learns from agent $A_{-i}$'s report as long as $A_i$'s output varies accordingly. By doing so, the principal may still be able to use the benefits of correlated information. Simple bilateral contracts are not optimal.

To understand the nature of the output distortions and the role of the correlation, it is useful to compare the solution found in (19) with the standard Baron-Myerson formula (2) which corresponds also to the optimal mechanism had the principal offered (non-manipulable) simple bilateral contracts to his agents. Using (17), we observe that the second term on the right-hand side is null for a simple bilateral contract implementing the Baron-Myerson outcome $q^{BM}(\theta_i)$ since $\underset{\theta_{-i}}{E}\left(\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right) = 0$. By having $A_i$'s output depend on $\theta_{-i}$, one departs from the Baron-Myerson outcome, and the principal can use $A_{-i}$'s report to reduce $A_i$'s information rent. Suppose indeed that the principal starts from the simple bilateral Baron-Myerson contract with $A_i$ but slightly modifies it to improve rent extraction once he has learned $A_{-i}$'s type. By using $A_{-i}$'s report the principal should infer how likely it is that $A_i$ lies on his type.

From (MLRP) there exists $\theta^*_{-i}(\theta_i)$ such that $\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \geq 0$ if and only if $\theta_{-i} \geq \theta^*_{-i}(\theta_i)$. Hence, the principal's best estimate of $A_i$'s type is $\theta_i$ if he learns $\theta_{-i} = \theta^*_{-i}(\theta_i)$ from $A_{-i}$. Everything happens as if $A_{-i}$'s report did not bring more information on $A_i$'s type. The only principal's concern remains reducing the first-term on the right-hand side of (17) and the optimal output corresponds to the Baron-Myerson outcome. Think now of an observation $\theta_{-i} > \theta^*_{-i}(\theta_i)$. Such signal let the principal think that $A_i$ has not exaggerated his cost parameter and there is less need for distorting output. The distortion with respect to the first-best outcome is less than Baron-Myerson. Instead, a signal $\theta_{-i} < \theta^*_{-i}(\theta_i)$ goes against $A_i$'s report if he exaggerates his type. This requires punishing $A_i$ by increasing the output distortion beyond the Baron-Myerson solution.

# 6 General Environments

## 6.1 Characterizing Non-Manipulability

With separable projects, non-manipulability constraints are also separable and it was straightforward to derive the form of non-manipulable schemes. With production externalities, things are more complex. We now derive second-best distortions in those more general environments. For simplicity, we still focus on the case of two agents and start this analysis with the case of continuous types.

Suppose that the principal wants to implement the vector of outputs $q(\theta) = (q_1(\theta), q_2(\theta))$ in a non-manipulable way. Assume that the following properties hold for such outputs:

**Assumption 3** *$q(\theta) = (q_1(\theta), q_2(\theta))$ is continuously differentiable and satisfies:*

$$\frac{\partial^2 \tilde{S}}{\partial q_i \partial q_{-i}}(q(\theta)) \frac{\partial q_i}{\partial \theta_{-i}}(\theta) \frac{\partial q_{-i}}{\partial \theta_{-i}}(\theta) \geq 0 \quad \textit{for } i = 1, 2,\ \forall \theta \in \Theta^2. \tag{20}$$

For substitutes, Assumption 3 is satisfied when a given agent's output decreases with his own cost and increases with that of his peer. For complements, the output of an agent should decrease with both marginal costs.

As it will appear in Lemma 2 below, (20) is indeed a second-order condition ensuring that the principal's best strategy is telling the truth on whatever he has learned from the other agent. This condition is similar to those found in standard screening problems.[26]

**Assumption 4** *$q(\theta) = (q_1(\theta), q_2(\theta))$ satisfies:*

$$\left| \frac{\partial q_{-i}}{\partial \theta_{-i}}(\theta) \right| \geq \left| \frac{\partial q_i}{\partial \theta_{-i}}(\theta) \right| \ \forall \theta \in \Theta^2. \tag{21}$$

[26] In such problems, a single crossing assumption on the agent's utility function is enough to derive the almost everywhere differentiability of the screening variable. Here instead, when dealing with the non-manipulability of his report $\hat{\theta}_{-i}$ vis-à-vis agent $A_i$ computing the cross-derivative of the principal's objective can only be done once it is assumed that the screening variable $q_{-i}(\theta)$ is continuously differentiable with respect to $\theta_{-i}$. This leads us to restrict a priori to differentiable schedules instead of deriving this property from revealed preferences arguments or from the envelope theorem as in the case of separable projects. A similar trick is used in common agency literature (Stole 1991, Martimort 1992).

Assumption 4 simply means that the own-impact of an agent's cost parameter on his output is greater than its impact on the other agent's output.

Next lemma provides a local characterization of non-manipulable allocations with continuously differentiable schedules.

**Lemma 2** *Assume that $q(\theta)$ satisfies Assumptions 3 and 4. The following necessary first-order conditions for the non-manipulability constraints (8) are also locally sufficient:*

$$\frac{\partial \tilde{S}}{\partial q_i}(q(\theta))\frac{\partial q_i}{\partial \theta_{-i}}(\theta) = \frac{\partial t_i}{\partial \theta_{-i}}(\theta)\ \forall \theta \in \Theta^2. \tag{22}$$

Turning now to the issue of global optimality for the principal of non-manipulating what he has learned, we have:

**Lemma 3** *Assume that $q(\theta)$ satisfies Assumptions 3 and 4. A sufficient condition for global optimality of the principal's non-manipulating strategy is:*

$$\frac{\partial^2 \tilde{S}}{\partial q_1 \partial q_2}(q_1, q_2) = \lambda \in \mathbb{R} \quad \forall (q_1, q_2) \in \mathcal{Q}^2. \tag{23}$$

Integrating (22) immediately yields the following expressions of the transfers:

$$t_i(\theta) = \int_{\underline{\theta}}^{\theta_{-i}} \frac{\partial \tilde{S}}{\partial q_i}(q(\theta_i, x))\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, x)dx - H_i(\theta_i) \text{ for } i = 1, 2 \tag{24}$$

where $H_i(\theta_i)$ is some arbitrary function. For a given output schedule $q(\theta)$ satisfying Assumptions 3 and 4, non-manipulable transfers are thus determined up to some functions $H_i(\cdot)$. The transfers obtained in (24) generalize the sell-out contracts obtained with separable activities to the case of production externalities.

To understand the new distortions involved with a production externality, it is useful thinking of the case of a small production externality (i.e., $\left|\frac{\partial^2 \tilde{S}}{\partial q_1 \partial q_2}(q_1, q_2)\right| = |\lambda|$ small enough). The principal can still offer sell-out contracts $t_i(\theta) = \tilde{S}(q_i(\theta), 0) - H_i(\theta_i)$ with little modifications of the information rents left to the agents and little changes in allocative efficiency compared to the case without externality. However, these sell-out schemes are now manipulable. To see how, define the principal's payoff when informed on $\theta = (\theta_i, \theta_{-i})$ and choosing a manipulation $\hat{\theta} = (\hat{\theta}_i, \hat{\theta}_{-i})$ as:

$$V(\hat{\theta}, \theta) = \tilde{S}(q_i(\theta_i, \hat{\theta}_{-i}), q_{-i}(\hat{\theta}_i, \theta_{-i})) - \sum_{i=1}^{2} t_i(\theta_i, \hat{\theta}_{-i}).$$

When Assumption 3 holds, we get:

$$\frac{\partial V}{\partial \hat{\theta}_{-i}}(\hat{\theta}, \theta)|_{\hat{\theta}=\theta} = \left(\frac{\partial \tilde{S}}{\partial q_i}(q_i(\theta), q_{-i}(\theta)) - \frac{\partial \tilde{S}}{\partial q_i}(q_i(\theta), 0)\right)\frac{\partial q_i}{\partial \theta_{-i}}(\theta) = \lambda q_{-i}(\theta)\frac{\partial q_i}{\partial \theta_{-i}}(\theta) < 0 \text{ when } \lambda \neq 0.^{27} \tag{25}$$

To limit the scope for manipulating $\hat{\theta}_{-i}$, (25) shows that the principal has at his disposal roughly two strategies. The first one consists in reducing $A_{-i}$'s output. At the extreme, this would mean committing himself to always deal only with $A_i$, a non-manipulable but also highly inefficient contract when agents exert complementary activities. The second strategy consists in making $A_i$'s output $q_i$ less sensitive to $\theta_{-i}$ like in a simple bilateral contract. Which strategy is preferred depends on types realizations and the nature of the externality. We now turn to necessary conditions that must be satisfied by the optimal mechanism in such settings with production externality before giving some hints on the nature of those distortions.

**Proposition 3** *Assume that (23) is satisfied, that (20) and (21) hold for that solution and the agents' incentive problems are regular, the optimal non-manipulable output $q^{SB}(\theta)$ solves the system of partial derivative equations*

$$\text{For } i=1,2,\ \tilde{f}(\theta)\left(\left(1+\frac{F(\theta_i)}{f(\theta_i)}\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\right)\left(\frac{\partial \tilde{S}}{\partial q_i}(q^{SB}(\theta))-\theta_i\right)-\frac{F(\theta_i)}{f(\theta_i)}\right)$$

$$=\lambda\left(F(\theta_{-i})\left(\int_{\underline{\theta}}^{\theta_i}\tilde{f}_{\theta_{-i}}(x|\theta_{-i})dx\right)\frac{\partial q_{-i}^{SB}}{\partial \theta_i}(\theta)-F(\theta_i)\left(\int_{\underline{\theta}}^{\theta_{-i}}\tilde{f}_{\theta_i}(x|\theta_i)dx\right)\frac{\partial q_{-i}^{SB}}{\partial \theta_{-i}}(\theta)\right) \quad (26)$$

*with the boundary conditions*

$$\frac{\partial \tilde{S}}{\partial q_i}(q^{SB}(\underline{\theta},\theta_{-i}))=\underline{\theta} \text{ and } \frac{\partial \tilde{S}}{\partial q_{-i}}(q^{SB}(\underline{\theta},\theta_{-i}))=\varphi(\theta_{-i},\underline{\theta})\ i=1,2. \quad (27)$$

The hyperbolic system of first-order partial derivative equations (26) generalizes the Baron-Myerson formula to the case of production externalities. Finding its solutions satisfying the boundary conditions (27) which determine outputs at $\theta_1=\underline{\theta}$ and $\theta_2=\underline{\theta}$ requires numerical methods. In the Appendix, we nevertheless propose a method to approximate such solution near the boundary defined by (27) when $\lambda \neq 0$.[28] The idea is to find approximations of the characteristic curves associated to the system (26) close to the boundary $(\underline{\theta},\theta_{-i})$ to solve explicitly the system at least locally. We then check ex post that Assumptions 3 and 4 both hold for the solution when $\lambda$ is small enough.

The system (26) with the boundary conditions (27) help of course to recover the solutions we already found for separable projects. Beyond that case, non-manipulability

[27]This is so since, with substitutes $\frac{\partial q_i}{\partial \theta_{-i}}(\theta)>0$ but $\frac{\partial \tilde{S}}{\partial q_{-i}}(q_i(\theta),q_{-i}(\theta))-\frac{\partial \tilde{S}}{\partial q_{-i}}(q_i(\theta),0)<0$ when $q_{-i}(\theta)>0$ whereas it is the reverse for complements.

[28]Even when $\tilde{S}(\cdot)$ and $\tilde{f}(\cdot)$ are both real analytic, the Cauchy-Kowalevski Theorem (see John (1982) for instance) cannot be directly used to ensure that a solution to (26) exists which is real analytic close to the boundary defined by (27) since indeed the coefficients of $\frac{\partial q_{-i}^{SB}}{\partial \theta_i}(\theta)$ and $\frac{\partial q_{-i}^{SB}}{\partial \theta_{-i}}(\theta)$ into (26) are both zero simultaneously when the right-hand side of (26) vanishes.

constraints force now the principal to take into account any impact of his output choice for agent $A_i$ on the transfer he gives to $A_{-i}$ and this introduces the new terms on the right-hand side of (26). However, as in the case of separable projects, non-manipulability does not matter for independent types. The optimal solution is then the second-best outcome not taking into account the possibility of manipulations. It corresponds to outputs distortions given by the familiar Baron-Myerson conditions:

$$\frac{\partial \tilde{S}}{\partial q_i}(q^{SB}(\theta)) = \theta_i + \frac{F(\theta_i)}{f(\theta_i)}.$$

In order to give more insights on the nature of these distortions, assume that $\tilde{S}(\cdot)$ is quadratic and writes as $\tilde{S}(q) = \mu(q_1 + q_2) - \frac{1}{2}(q_1^2 + q_2^2) + \lambda q_1 q_2$, where $|\lambda| < 1$ to ensure strict concavity of $\tilde{S}(\cdot)$. Denote $l = \frac{\tilde{f}_{\theta_i}(\underline{\theta}|\underline{\theta})}{\tilde{f}(\underline{\theta}|\underline{\theta})}$ the likelihood ratio at $(\underline{\theta}, \underline{\theta})$. This can be viewed as an index of the correlation across types. Assuming strict (MLRP), we have $l < 0$. Any real analytic solution to (26)-(27) close to $(\underline{\theta}, \underline{\theta})$ (which lies on the boundary surfaces defined in (27)) can be approximated locally as follows.

**Corollary 1** *Assume that $\tilde{S}(\cdot)$ is quadratic as above and $f(\cdot)$ is real analytic with $-\frac{f'(\underline{\theta})}{2f(\underline{\theta})} = m$. Locally around $(\underline{\theta}, \underline{\theta})$, any symmetric real analytic solution to (26)-(27) admits the following approximation:*

$$q_i^{SB}(\theta_i, \theta_{-i}) - q^{FB}(\underline{\theta}, \underline{\theta}) = -\frac{1}{1-\lambda^2}\left((\theta_i - \underline{\theta}) + (l - m)(\theta_i - \underline{\theta})^2\right)$$

$$+\frac{\lambda}{1-\lambda^2}\left(-2(\theta_{-i} - \underline{\theta}) + (l - m)(\theta_{-i} - \underline{\theta})^2\right) - \frac{2\lambda l}{1-\lambda^2}(\theta_i - \underline{\theta})(\theta_{-i} - \underline{\theta}) + o(||\theta - \underline{\theta}||^2) \quad (28)$$

*where* $lim_{||\theta-\underline{\theta}||\to 0}\frac{o(||\theta-\underline{\theta}||^2)}{||\theta-\underline{\theta}||^2} = 0.$

From (28), Assumptions 3 and 4 hold for the optimal output $q^{SB}(\theta)$ at least locally around $(\underline{\theta}, \underline{\theta})$. To understand the nature of the output distortions away from the first-best, it is necessary to decompose it into three elements. First, there is the *generalized virtual cost effect* that comes on the first bracketed term on the right-hand side of (28). This term survives when there is no production externality and is only due, as in Section 5, to the fact that costs are replaced by generalized virtual costs in evaluating the rent/efficiency trade-off under non-manipulability. Within that bracket, the negative term $-\frac{(l-m)}{1-\lambda^2}(\theta_i - \underline{\theta})^2$ captures how correlation affects optimal outputs. Because learning from agent $A_{-i}$ that his type is close to $\underline{\theta}$ can only be bad news when $A_i$ reports himself a type $\theta_i$ above $\underline{\theta}$, this first effect leads to an exacerbated downward distortion of $q_i^{SB}(\theta_i, \theta_{-i})$. This term is reduced in absolute value when correlation diminishes. The next bracketed term captures the impact of the production externality that would arise if the right-hand side of (26) was set at zero. This *indirect effect of production externality* comes from the fact that, as the *generalized* *virtual cost effect* distorts the output of a given agent, substitutability or complementarity imply further distortion of the other agent's output. For instance, with substitutes the downward distortion of $A_{-i}$'s output due to the generalized virtual cost effect leads to raise $A_i$'s output and that all the more that the correlation increases. Finally, the last term on the right-hand side of (28) captures the impact of the production externality on the principal's incentives to manipulate: *a direct effect of production externality.* It represents the extra distortions needed to move sell-out contracts towards being non-manipulable. This last term increases output distortions around $(\underline{\theta}, \underline{\theta})$ for substitutes. Indeed, manipulations vis-à-vis $A_i$ are better fought by making $A_i$'s output less sensitive to $A_{-i}$'s cost. Distortions are instead reduced with complements because manipulations are then better fought by making outputs less sensitive to $A_{-i}$'s cost. Note finally that, as the correlation increases (in the sense of having $|l|$ bigger), the direct effect of production externality on output distortions is magnified.

## 6.2 Specific Results with Discrete Types

The complexity of finding solutions to (26)-(27) suggests to investigate now the nature of optimal non-manipulable mechanisms in a discrete types environment where full-fledged solutions could be found. To analyze those incentives for manipulation and their consequences on contract design, we study two polar cases of interest: perfect complements and perfect substitutes. As in the case of separable projects, our goal here is to first derive implications of the non-manipulability constraints in such simple environments when characterizing the set of incentive-feasible allocations and then to find the corresponding optimal mechanism.

### 6.2.1 Teams

Consider a team where agents exert efforts $q_1$ and $q_2$ which are perfect complements in the production process. Denote by $q = \min(q_1, q_2)$ this output and by $S(q)$ the principal's surplus ($S'(0) = +\infty$, $S' > 0$, $S'' < 0$ with $S(0) = 0$).[29] Any symmetric mechanism is characterized by an output schedule with three possible elements $\{q(\bar{\theta}, \bar{\theta}), q(\bar{\theta}, \underline{\theta}) = q(\underline{\theta}, \bar{\theta}), q(\underline{\theta}, \underline{\theta})\}$ and a four-uple of transfers $\{t(\bar{\theta}, \bar{\theta}), t(\bar{\theta}, \underline{\theta}), t(\underline{\theta}, \bar{\theta}), t(\underline{\theta}, \underline{\theta})\}$.

Given the constraints on output observability, some manipulations are not possible. For instance, pretending that $(\hat{\theta}_1, \hat{\theta}_2) = (\underline{\theta}, \bar{\theta})$ when $(\theta_1, \theta_2) = (\underline{\theta}, \underline{\theta})$ is not feasible given that such report would require implementing $q(\underline{\theta}, \bar{\theta})$ vis-à-vis $A_1$ and a different output

[29] The Inada condition again ensures that it is worth always contracting with both agents so that the issue of "shutting-down" the worst types again does not arise. Now the ability of the principal to manipulate reports towards both agents is constrained by the fact that, assuming there is no waste of their individual inputs, both agents observe the same final output on which contracts can be written.

$q(\underline{\theta}, \underline{\theta})$ vis-à-vis $A_2$. The only two relevant non-manipulability constraints (8) are:

$$S(q(\underline{\theta}, \underline{\theta})) - 2t(\underline{\theta}, \underline{\theta}) \geq S(q(\underline{\theta}, \bar{\theta})) - 2t(\underline{\theta}, \bar{\theta}) \tag{29}$$

$$S(q(\bar{\theta}, \bar{\theta})) - 2t(\bar{\theta}, \bar{\theta}) \geq S(q(\bar{\theta}, \underline{\theta})) - 2t(\bar{\theta}, \underline{\theta}). \tag{30}$$

Constraint (29) comes from the fact that the principal can always report to an efficient agent that the other is not even when both are. Constraint (30) captures the fact that the principal can always report to an inefficient agent that the other is efficient again even when indeed both are inefficient. Since the principal can only lie to both agents at the same time, it is worth noticing that those constraints correspond to global deviations.

**Proposition 4** *Assume that the level of correlation $\alpha$ is small enough. The optimal symmetric non-manipulable team production mechanism is such that (30) is binding. Optimal outputs are given by:*

$$S'(q^{SB}(\underline{\theta}, \underline{\theta})) = 2\underline{\theta},$$

$$S'(q^{SB}(\underline{\theta}, \bar{\theta})) = \underline{\theta} + \bar{\theta} + \frac{\nu}{1-\nu}\left(\frac{1 + \alpha\frac{3-2\nu}{\nu^2}}{1 - \alpha\frac{1-2\nu}{\nu(1-\nu)}}\right)\Delta\theta, S'(q^{SB}(\bar{\theta}, \bar{\theta})) = 2\bar{\theta} + \frac{2\nu}{1-\nu}\left(\frac{1 - \frac{\alpha}{\nu(1-\nu)}}{1 + \alpha\frac{2\nu}{(1-\nu)^3}}\right)\Delta\theta.$$

Observe that, in the limit of zero correlation, the optimal outputs above again converge towards the Baron-Myerson outcome where the marginal efficiency of production is equal to the sum of the agents' virtual costs, i.e.,

$$S'(q_0^{SB}(\underline{\theta}, \bar{\theta})) = \underline{\theta} + \bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta, S'(q_0^{SB}(\bar{\theta}, \bar{\theta})) = 2\bar{\theta} + \frac{2\nu}{1-\nu}\Delta\theta.$$

Because of perfect complementarity, these Baron-Myerson outputs entail thus a double distortion $\frac{2\nu}{1-\nu}\Delta\theta$ when both agents are inefficient whereas there is a simple distortion $\frac{\nu}{1-\nu}\Delta\theta$ when only one agent is inefficient.

Starting from this benchmark, the principal would like to use the types correlation to reduce the rent of an efficient type. This can be done by reducing the payment $t(\bar{\theta}, \underline{\theta})$ that this agent could get by lying on his type when facing an efficient agent since such event is rather unlikely with a positive correlation. At the same time, satisfying the participation constraint of an inefficient agent requires also to raise $t(\bar{\theta}, \bar{\theta})$. Altogether those changes in payments makes it attractive to manipulate reports so that (30) necessarily binds and payments satisfy

$$S(q(\bar{\theta}, \bar{\theta})) - S(q(\bar{\theta}, \underline{\theta})) = 2t(\bar{\theta}, \bar{\theta}) - 2t(\bar{\theta}, \underline{\theta}).$$

Relaxing this constraint requires moving $q^{SB}(\bar{\theta}, \bar{\theta})$ up closer to the first-best and $q^{SB}(\underline{\theta}, \bar{\theta})$ down closer to zero.

### 6.2.2 Multi-Unit Auctions

Consider an auction context with agents producing perfect substitutes in quantities $q_1$ and $q_2$ on the principal's behalf. The principal's surplus from consuming $q = q_1 + q_2$ units of the good is $S(q)$, with the Inada conditions $S(0) = S'(+\infty) = 0$, $S'(0) = +\infty$, $S'' < 0$. The first-best outcome is such that each agent produces the quantities defined as follows:

$$S'(2q^{FB}(\underline{\theta}, \underline{\theta})) = S'(q^{FB}(\underline{\theta}, \bar{\theta})) = \underline{\theta}, q^{FB}(\bar{\theta}, \underline{\theta}) = 0, S'(2q^{FB}(\bar{\theta}, \bar{\theta})) = \bar{\theta}.$$

Under asymmetric information, any symmetric mechanism is now characterized by a four-uple of non-negative quantities for each individual agent $\{q(\bar{\theta}, \bar{\theta}), q(\bar{\theta}, \underline{\theta}), q(\underline{\theta}, \bar{\theta}), q(\underline{\theta}, \underline{\theta})\}$ and a four-uple of corresponding transfers $\{t(\bar{\theta}, \bar{\theta}), t(\bar{\theta}, \underline{\theta}), t(\underline{\theta}, \bar{\theta}), t(\underline{\theta}, \underline{\theta})\}$.

To extract the agents' information rent, the principal would like again to punish the inefficient agent in case of conflicting reports by setting $t(\bar{\theta}, \underline{\theta})$ sufficiently low. However, non-manipulability in state $(\bar{\theta}, \bar{\theta})$ imposes first the following *local non-manipulability constraint* corresponding to the case where the principal only manipulates report towards one of the two agents

$$S(2q(\bar{\theta}, \bar{\theta})) - 2t(\bar{\theta}, \bar{\theta}) \geq S(q(\bar{\theta}, \bar{\theta}) + q(\bar{\theta}, \underline{\theta})) - t(\bar{\theta}, \bar{\theta}) - t(\bar{\theta}, \underline{\theta}). \quad (31)$$

Second, it imposes also the *global non-manipulability constraint* corresponding instead to the case where the principal manipulates simultaneously his reports towards both agents

$$S(2q(\bar{\theta}, \bar{\theta})) - 2t(\bar{\theta}, \bar{\theta}) \geq S(2q(\bar{\theta}, \underline{\theta})) - 2t(\bar{\theta}, \underline{\theta}) \quad (32)$$

Because the surplus function $S(\cdot)$ is concave and we expect $q(\bar{\theta}, \underline{\theta}) \leq q(\bar{\theta}, \bar{\theta})$, (32) is automatically satisfied when (31) already holds. In contrast with the case of perfect complements (Section 6.2.1), the relevant non-manipulability constraint is now local.

Intuition built by looking at the form taken by the efficient allocation suggests a kind of "winner-takes-it-all" solution such that the good is entirely produced by the efficient agent only when the other reports being inefficient, i.e. $q(\bar{\theta}, \underline{\theta}) = 0$. When (31) is binding, transfers must then satisfy the following condition that, again, generalizes the expression of sell-out contracts to the auction context

$$S(2q(\bar{\theta}, \bar{\theta})) - S(q(\bar{\theta}, \bar{\theta})) = t(\bar{\theta}, \bar{\theta}) - t(\bar{\theta}, \underline{\theta}).$$

**Proposition 5** *Assume that the level of correlation $\alpha$ is small enough. The optimal symmetric non-manipulable multi-unit auction is such that (31) is binding. Optimal outputs are given by:*

$$S'(2q^{SB}(\underline{\theta}, \underline{\theta})) = S'(q^{SB}(\underline{\theta}, \bar{\theta})) = \underline{\theta}, q^{SB}(\bar{\theta}, \underline{\theta}) = 0,$$

$$S'(2q^{SB}(\bar{\theta},\bar{\theta})) = \bar{\theta} + \frac{\nu}{1-\nu}\left(\frac{1-\frac{\alpha}{\nu(1-\nu)}}{1+\frac{\alpha(2-\nu)}{(1-\nu)^3}}\right)\Delta\theta$$

$$+\frac{\alpha}{(1-\nu)((1-\nu)^2+\alpha)+\alpha}(S'(q^{SB}(\bar{\theta},\bar{\theta})) - S'(2q^{SB}(\bar{\theta},\bar{\theta}))) \quad (33)$$

In the limiting case of zero correlation, the optimal non-manipulable auction above converges towards the optimal auction with independent types and no manipulability constraints which implements

$$S'(2q_0^{SB}(\bar{\theta},\bar{\theta})) = \bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta.$$

Depending on the exact shape of the surplus $S(\cdot)$, the optimal quantity $q^{SB}(\bar{\theta},\bar{\theta})$ is either greater or smaller than $q_0^{SB}(\bar{\theta},\bar{\theta})$. Correlation might be best used either to extract more rent or to increase efficiency.[30]

# 7 Extensions

This Section discusses the robustness of our findings to alternative assumptions on contracting possibilities.

## 7.1 Dominant Strategy and Simple Bilateral Contracting

Let us come back to the case of continuous types. Section 5 showed that, in a Bayesian setting, any information learned by the principal when contracting with a given agent is used to regulate another agent when types are correlated. We now strengthen the implementation concept and require that agents play dominant strategies in the mechanism offered by the principal. We ask whether it makes optimal non-manipulable mechanisms look more like a set of simple bilateral contracts: an extreme case of non-manipulability.

Notice first that the notion of non-manipulability is independent of the implementation concept used to describe the agents' behavior. Our framework can be easily adapted to dominant strategy implementation.

**Proposition 6** ***Revelation Principle for Dominant Strategy Implementation with Bilateral Contracting.*** *In a private values context, any allocation $a(\cdot)$ achieved*

[30] Suppose for instance that the surplus is given by $S(q) = q^\beta$, with $0 < \beta < 1$. Straightforward computations show that $q^{SB}(\bar{\theta},\bar{\theta}) \geq q_0^{SB}(\bar{\theta},\bar{\theta})$ if and only if $\beta \geq 1 - \frac{\log(2-\frac{\theta}{\bar{\theta}+\frac{\nu}{1-\nu}\Delta\theta})}{\log 2}$.

*at a dominant strategy equilibrium of any arbitrary mechanism $(g(\cdot), \mathcal{M})$ with bilateral contracting can alternatively be implemented as a truthful and non-manipulable dominant strategy equilibrium of a direct mechanism $(\bar{g}(\cdot), \Theta^n)$.*

Non-manipulability being independent of the implementation concept, we still have the following expression of the transfers when projects are separable:

$$t_i(\theta_i, \theta_{-i}) = S(q_i(\theta_i, \theta_{-i})) - H_i(\theta_i).$$

Strengthening also the participation condition, we obtain:

**Proposition 7** *Assume that projects are separable (i.e., $\frac{\partial^2 \tilde{S}}{\partial q_1 \partial q_2} = 0$), the mechanism is implemented in dominant strategy and satisfy ex post participation constraints. The optimal non-manipulable such mechanism can be achieved with a pair of simple bilateral contracts $(t_i^{BM}(\theta_i), q_i^{BM}(\theta_i))$ implementing the Baron-Myerson output for each agent.*

Non-manipulability and dominant strategy implementation are two different concepts with different implications. These restrictions justify simple bilateral contracts only when taken in tandem. With dominant strategy and non-manipulability, informational externalities can no longer be exploited and the principal cannot do better than offering simple bilateral contracts. Therefore, the Baron-Myerson outcome is optimal even with correlated types.

**Remark 6** *Simple bilateral contracts are suboptimal if we do not impose non-manipulability even under dominant strategy implementation and ex post participation. Insisting only on dominant strategy and ex post participation, the optimal quantities are given by Baron-Myerson formulae taking into account the fact that the principal uses the correlation of types to update his beliefs accordingly. We get:*

$$S'(q_i(\theta_i, \theta_{-i})) = \theta_i + \frac{\tilde{F}(\theta_i \mid \theta_{-i})}{\tilde{f}(\theta_i \mid \theta_{-i})}.$$

*The optimal mechanism without the non-manipulability constraint yields a strictly higher payoff than a pair of bilateral contracts when types are correlated.*

## 7.2 Horizontal Collusion and Simple Bilateral Contracting

We now investigate the possibility that agents may collude and how such collusion may interact with non-manipulability to "simplify" contracts.[31] Again, we focus on the case

[31] We do not tackle the important issue (left for future research) of knowing whether bilateral contracting might facilitate or hinder collusion. This would require a theory showing how the transaction costs of collusive behavior between asymmetrically informed agents are affected by bilateral contracting. Building such theory is beyond the scope of this paper.

of separable projects and, following Laffont and Maskin (1980), suppose that, when colluding, agents learn each other's types.

Given the form that (symmetric) non-manipulable mechanisms take in this environment (see equation (12)), coalition incentive compatibility requires that agents jointly tell the truth when having the option to deviate as a coalition:

$$(\theta_1, \theta_2) \in \arg \max_{(\hat{\theta}_1, \hat{\theta}_2) \in \Theta^2} \sum_{i=1}^{2} S(q_i(\hat{\theta}_i, \hat{\theta}_{-i})) - \theta_i q_i(\hat{\theta}_i, \hat{\theta}_{-i}) - H(\hat{\theta}_i).$$

This yields the necessary first-order conditions:

$$-H'(\theta_k) + \sum_{i=1}^{2} \left(S'(q_i(\theta_k, \theta_{-k})) - \theta_i\right) \frac{\partial q_i}{\partial \theta_k}(\theta_k, \theta_{-k}) = 0 \quad \text{for } k = 1, 2. \tag{34}$$

These collusion-proofness conditions are helpful to show the following result.

**Proposition 8** *Assume that agents work on separable projects and can collude:*

- *The optimal mechanism described in Proposition 2 is not collusion-proof;*

- *The only differentiable output schedules $q_i(\theta_i, \theta_{-i})$ which are such that $q_i(\theta_i, \theta_{-i}) \leq q^{FB}(\theta_i)$ (with equality at $\theta_i = \underline{\theta}$ only) and $\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, \theta_{-i}) \geq 0$ and implementable by a collusion-proof non-manipulable mechanism are such that $\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, \theta_{-i}) = 0$. The optimal mechanism within this class is a pair of simple bilateral contracts implementing the Baron-Myerson outcome $q^{BM}(\theta_i)$.*

The optimal mechanism characterized in Proposition 2 when agents do not collude makes the output of any given agent depend also on the report of the other. From the coalition's viewpoint, reporting truthfully is not optimal however. Indeed, given that agent $A_i$ produces below the first-best for that mechanism, the coalition would like that he overstates his report because revealing such information has a positive effect on $A_{-i}$'s payoff. This points at the difficulty in reconciling non-manipulability and collusion-proofness unless the principal gives up any attempt to make the contract of an agent depend on information he learns from the other. Under some weak conditions, the only possibility left is to offer simple bilateral contracts.

## 7.3 Secret Contracts

Our analysis so far focused on the case where the mechanisms offered by the principal are public. This assumption allowed us to focus on the role of privacy in communication only. An extra degree of privacy arises when the principal offers secret mechanisms to each agent. In this case, not only the particular choice of $A_{-i}$ within the menu he receives but this menu itself is not observed by $A_i$.

In the case of separable projects, avoiding the principal's manipulation on whatever messages he receives from the other agents still imposes that

$$S(q_i^S(\theta_i,\theta_{-i})) - t_i^S(\theta_i,\theta_{-i}) = H_i^S(\theta_i),$$

for some functions $H_i^S(\cdot)$ where $(t_i^S(\theta_i,\theta_{-i}), q_i^S(\theta_i,\theta_{-i}))$ is the direct mechanism[32] offered to $A_i$ in the game with secret bilateral contracts. It should be noted at this stage that with separable projects, the offer $(t_{-i}^S(\cdot), q_{-i}^S(\cdot))$ made to agent $A_{-i}$ does not influence how the principal manipulates the report he makes to agent $A_i$. As a consequence, whether the offers are public or secret does not change the incentives faced by agent $A_i$: the conjectures about the offer made to $A_{-i}$ following any unexpected deviation that the principal may envision $\{t_i(\theta), q_i(\theta)\}_{\theta\in\Theta^2}$ do not intervene in the reasoning of agent $A_i$. Therefore, we can immediately replicate the analysis made in Section 5.

**Proposition 9** *When projects are separable (i.e., $\frac{\partial^2 \tilde{S}}{\partial q_1 \partial q_2} = 0$), the equilibrium outcomes of the game with public contracts and of the game with secret contracts coincide.*

## 7.4 Sequential Contracting

The model of bilateral contracting we developed so far has the principal making simultaneous offers to the agents. One important question is whether the principal could gain by instead contracting sequentially with each agent in turn. To simplify, let us focus on the case of separable projects with two agents. One potential benefit of sequential bilateral contracting is to relax some of the non-manipulability constraints faced by the principal. Intuitively, when the principal contracts with the second agent, he can condition his offer on all information already obtained from the first one. This replaces the problem of ex post manipulation by a problem of interim manipulation vis-à-vis the second agent.

The analysis of sequential bilateral contracting nevertheless raises a number of issues linked to the various subtleties that may arise in terms of information leakages depending on the exact timing chosen for such games. In general dynamic contracting environments, it is for instance no longer warranted that the usual form of the Revelation Principle applies so that the first agent $A_1$ finds it optimal to tell the truth.[33] Moreover, the contracting sub-game with $A_2$ becomes now an informed principal problem. From the *Unscrutability Principle*,[34] we know that such informed principal would like to offer a menu

[32] In this setting, the restriction to direct mechanisms is without loss of generality, see the proof of Proposition 9 in the Appendix.
[33] Bester and Strausz (2001).
[34] Myerson (1983).

of sub-mechanisms $\{t_2(\tilde{\theta}_1, \hat{\theta}_2), q_2(\tilde{\theta}_1, \hat{\theta}_2)\}_{(\tilde{\theta}_1, \hat{\theta}_2)\in\Theta^2}$ delaying all information revelation till after $A_2$'s acceptance. This simultaneous revelation of the principal and $A_2$'s information stands in contrast with what is feasible with simultaneous bilateral contracting where the principal reports always after both agents. The contracting possibilities come definitively closer to what is feasible under a centralized mechanism. Alternatively, if we were instead keeping a more symmetric approach by still allowing the principal to report to $A_2$ what he has learned from $A_1$ only after $A_2$'s own report, we would be obviously back to the same non-manipulability constraints than under simultaneous contracting and our previous analysis would carry over.

To make what we believe is a more relevant comparison between the simultaneous and sequential contracting games, we are thus led to study the following timing. First agents privately learn their respective efficiency parameters. Second the principal offers to $A_1$ a sub-mechanism $\{t_1(\hat{\theta}_1, \tilde{\theta}_2), q_1(\hat{\theta}_1, \tilde{\theta}_2)\}_{(\hat{\theta}_1, \tilde{\theta}_2)\in\Theta^2}$. Third, $A_1$ reports $\hat{\theta}_1$ in this sub-mechanism; this message is observed by the principal. Fourth, the principal offers to $A_2$ a sub-mechanism $\{t_2(\hat{\theta}_2), q_2(\hat{\theta}_2)\}_{\hat{\theta}_2\in\Theta}$. [35] The equilibrium offers may depend of course on the information the principal learns from $A_1$ if those offers are separating. Fifth, $A_2$ reports $\hat{\theta}_2$ in this sub-mechanism. Sixth, the principal uses the information learned from $A_2$ to report $\tilde{\theta}_2$ in $A_1$'s sub-mechanism. Finally, production and transfers occur.

To establish our next result, we need to ensure that the virtual type of $A_2$, computed once the realization of $\theta_1$ is known, is increasing in $\theta_2$ to guarantee that the incentive problem with respect to $A_2$ is well-behaved.

**Assumption 5** *The function $\theta_2 + \frac{\tilde{F}(\theta_2|\theta_1)}{\tilde{f}(\theta_2|\theta_1)}$ is increasing in $\theta_2$ for all $\theta_1$.*

In this setting, we can prove that the unique equilibrium is fully revealing, i.e., such that both agents report truthfully in their sub-mechanisms, the principal does not manipulate his report on $A_2$ in $A_1$'s sub-mechanism and reveals his information on $A_1$'s report through his contract to $A_2$. Moreover, we have:

**Proposition 10** *When projects are separable and Assumption 5 holds, the principal's expected profit in the sequential bilateral contracting game is lower than in the simultaneous contracting game.*

This proposition justifies our focus on a game with simultaneous bilateral contracting.

[35] Equivalently, the principal offers to $A_2$ a non linear price $\{t_2(q_2)\}$ and let $A_2$ choose the quantity $q_2$. In this sequential game, the strategy space available to $P$ in the relationship with $A_2$ is therefore different from what is available in the simultaneous game.

# 8 Relationships with the Literature

This paper is linked to several trends of the mechanism design literature reviewed below.

**Partial Commitment.** Our modeling of the principal's limited inability to commit to a grand-mechanism leads to a tractable characterization of non-manipulable mechanisms with bilateral contracting by means of a simple Revelation Principle. More generally, models with partial commitment require giving up such simple approach and might involve partially revealing strategies (Bester and Strausz 2000 and 2001, Krishna and Morgan 2006). This difficulty is avoided in our context because we focus on private values environments where the principal's utility function does not directly depend on the agents' types. Hence, agents do not manipulate their reports to the principal to affect his beliefs about their types and influence his optimal manipulations. Whatever information is learned by the principal with an agent, non-manipulability requires that it is truthfully revealed to others. Non-manipulability constraints can thus be interpreted as incentive compatibility constraints with respect to the endogenous private information that the principal learns from the agents. This is reminiscent of the *posterior implementability* concept developed by Green and Laffont (1987) in which agents' equilibrium strategies are best-responses to each other even after they learned information revealed by the play of the mechanism itself. However, non-manipulability concerns the principal's behavior.

Finally, Baliga, Corchon and Sjostrom (1997) investigate implementation when the mediator himself is a player and reacts to whatever information privately informed agents may report by choosing a decision. Formally, the mechanism design game is transformed into a signaling game. We are less extreme in modeling the principal's lack of commitment and still allow some commitment to bilateral contracts.

**Mechanism design in correlated environments.** Results on the irrelevance of private information in correlated information environments (Crémer and McLean 1985, 1988, Riordan and Sappington 1988, Johnson, Pratt and Zeckhauser 1990, d'Aspremont, Crémer and Gerard-Varet 1990, Matsushima 1991 and McAfee and Reny 1992) have already been attacked on various fronts. A first approach is to introduce risk-aversion and wealth effects (Robert 1991, Eso 2004), limited liability (Demougin and Garvie 1991), ex post participation constraints (Demski and Sappington 1988, Dana 1993), or limited enforceability (Compte and Jehiel 2009). In our paper instead, the benefits of using correlated information is undermined by non-manipulability constraints on the principal's side.

A second approach argues that correlation may not be as generic as suggested by the earlier literature. Neeman (2004) points out that the type of an agent should not simultaneously determine his beliefs on others and be payoff-relevant. Bergemann and Morris (2005) model higher order beliefs and show that robust implementation may amount to ex post implementation. Chung and Ely (2007) show that a maxmin principal may want to rely on dominant strategy implementation. These approaches yield somewhat extreme results since Bayesian mechanisms are given up. Bilateral contracting between the principal and his agents also relaxes the common knowledge requirements assumed in standard mechanism design but it does so in a simple and tractable way that preserves the properties of Bayesian implemtation.[36] As a result, Bayesian implementation keeps much of its force. Resolution techniques to derive optimal mechanisms are also similar to those already well-known for independent types.

Full surplus extraction in correlated environments may also be limited when agents collude (Laffont and Martimort 2000). Key to this collusion possibility is the fact that agents can coordinate their strategies. This coordination is certainly harder when agents contract separately with their principal. Our focus on bilateral contracting points at another polar case which leaves less scope for horizontal collusion between agents but introduces the possibility of manipulations by the principal.[37]

**Subjective evaluations.** There is a literature on the design of incentive contracts between a principal and his agents in moral hazard contexts where the principal's evaluation of the agents' performance is subjective, i.e., private information of the principal himself (McLeod 2003, Fuchs 2007). One possible solution to restore double-edged incentives is "burning money." Another solution suggested by Rahman and Obara (2007), is to use correlated strategy as the implementation concept. In our model, both the option of "burning money" and that of appointing another mediator making secret recommendations rely on the ability of the principal not to collude with a third-party. With such collusion, those solutions lose their bites and we are back to the same analysis as that undertaken in this paper. In this respect, our paper, by allowing for several agents, informational and production externalities, and continuous types goes beyond the findings in Strausz (2006) who analyzes also the principal's incentives to manipulate an informative signal vis-à-vis a unique agent.[38]

**Contractual externalities with bilateral contracts.** The IO literature on bilateral contracting (O'Brien and Shaffer 1992, McAfee and Schwartz 1994, Segal 1999) analyzes complete information environments where a manufacturer (principal) can contract with his retailers (agents) only through simple bilateral contracts. The focus is on the principal's opportunistic behavior that arises when he strikes each of those bilateral deals independently and the retailers' payoffs depend on each others' contracting variables with

[36]Readers accustomed with the moral hazard literature know that correlation between the agents' performances may be used to better design incentives without of course voiding the agency problem of its interest. Our results have the same flavor.

[37]In a moral hazard context, Gromb and Martimort (2007) study the consequences for organizational design of vertical collusion between the principal and each of his agents.

[38]We thank R. Strausz for making us aware of this paper after earlier drafts of ours were completed.

the principal. Although, we share with this literature some concerns in studying the opportunistic behavior of a principal, this is in a different context. Our paper deals instead with informational externalities across agents. Since non-manipulability constraints depend only on the principal's payoff, introducing payoff externalities between agents would not change our analysis. Moreover, for most of our analysis above, mechanisms are public and, the principal's opportunistic behavior comes from his possibilities to manipulate communication and not to sign independent secret deals.

**Common agency.** Our modeling of the principal's lack of commitment is reminiscent of the common agency literature.[39] This should come at no surprise. In our framework, the key issue is to prevent the principal's opportunistic behavior vis-à-vis each of his agents. Under common agency, the same kind of opportunistic behavior occurs, with the common agent reacting to his principals' offers. Beside the allocation of bargaining powers between parties, there is another difference between common agency and the environment described in this paper. The principal has more commitment power here since he can design this mechanism in a first stage and restrict the choice of the uninformed agents. Although a priori minor, this latter difference simplifies the analysis. This instilled minimal level of commitment allows us to maintain much of the optimization techniques available in standard mechanism design without falling into the difficulties faced when characterizing Nash equilibria in the context of multi-contracting mechanism design, in particular, the possible multiplicity of equilibria. Martimort (2007) argues that one should look for minimal departures of the centralized mechanism design framework which go towards modeling multi-contracting settings but avoid this difficulty. The non-manipulability constraints we developed here can be viewed as such departure.

# 9 Conclusion

Considering bilateral contracts paves the way to a theory which responds to some of the most often heard criticisms addressed to the mechanism design methodology. Even in correlated information environments, considering non-manipulable mechanisms restores a genuine trade-off between efficiency and rent extraction. This leads to a more standard second-best analysis that stresses the role of virtual generalized costs even in correlated environments. In several contexts of interest (separable projects, auctions, team productions, more general production externalities), we analyzed this trade-off and characterized optimal non-manipulable mechanisms.

Several extensions would be worth pursuing.

[39] Bernheim and Whinston (1986), Stole (1991), Martimort (1992, 2007), Mezzetti (1997), Martimort and Stole (2002, 2009), Peters (2001). Most often private information is modeled on the common agent's side in this literature, an exception being Martimort and Moreira (2009).

First, each of these particular settings studied above certainly deserves further analysis either by modifying information structures and preferences or by focusing on organizational problems coming from the analysis of real world institutions in particular contexts (political economy, regulation, vertical restraints in a IO context, etc..).

Second, introducing a bias in the principal's preferences towards either agent could also raise interesting issues. First by making the principal's objective function somewhat congruent with that of one of the agents, one goes towards a simple modeling of vertical collusion and favoritism. Second, this congruence may introduce interesting aspects related to the common values aspect that arises in such environment and that have been set aside by our focus on a private values setting.

Third, it would be worth investigating further what is the scope for horizontal collusion between agents in the environments depicted in this paper. Considering collusion may justify the constraint on bilateral contracting in the first place. Indeed, bilateral contracting introduces private communication between agents which may make it difficult for agents to enforce any collusive agreement compared to the case of a grand-mechanism making all players' strategies public information. Such analysis could lead to an interesting trade-off between the cost of the principal's opportunism under bilateral contracting and the fact that collusion is easier with more centralized procedures.

Fourth, it might be worth extending our analysis of sequential contracting games to the case of production externalities. Such general investigation might lead to interesting insights on optimal contracting modes in environments with limited commitment.

In practice, the degree of transparency of communication in an organization may be intermediate between what arises either with bilateral contracting or with a more centralized grand-mechanism. Reputation on the principal's side might help preventing manipulability but the extent by which it is so remains unknown.

# References


Aoki, M. (1984): *The Cooperative Game Theory of the Firm*, London: Oxford University Press.

Aspremont, C. d', Crémer, J., and L.A. Gérard-Varet (1990): "Incentives and the Existence of Pareto-Optimal Revelation Mechanisms," *Journal of Economic Theory*, 51, 233-254.

Baliga, S., L. Corchon and T. Sjostrom (1997): "The Theory of Implementation when the Planner is a Player," *Journal of Economic Theory*, 77, 15-33.

Barany, I., (1992): "Fair Distribution Protocols or How the Players Replace Fortune," *Mathematics of Operations Research*, 17: 327-340.

Baron, D. and R. Myerson (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50, 911-930.

Bergemann, D. and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73, 1771-1814.

Bernheim, D. and M. Whinston (1986): "Menu Auctions, Resource Allocations and Economic Influence," *Quarterly Journal of Economics*, 101, 1-31.

Bester, H. and R. Strausz (2000): "Imperfect Commitment and the Revelation Principle: The Multi-Agent Case," *Economics Letters*, 69, 165-171.

Bester, H. and R. Strausz (2001): "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica* 69, 1077-1098.

Biglaiser, G. and C. Mezzetti (1993): "Principals Competing for an Agent in the Presence of Adverse Selection and Moral Hazard," *Journal of Economic Theory*, 61, 302-330.

Chung, K.S. and J. Ely (2007): "Foundations for Dominant Strategy Mechanisms," *Review of Economic Studies*, 74, 447-476.

Compte, O. and P. Jehiel (2009): "Veto Constraints in Mechanism Design: Inefficiency with Correlated Types," *American Economic Journal Microeconomics*, 1, 182-207.

Crémer, J., and R. McLean (1985): "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, 53, 345-362.

Crémer, J., and R. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica* , 56, 1247-1257.

Dana, J. (1993): "The Organization and Scope of Agents: Regulating Multiproduct Industries," *Journal of Economic Theory*, 59, 288-310.

Demougin, D. and D. Garvie (1991): "Contractual Design with Correlated Information under Limited Liability," *RAND Journal of Economics*, 22, 477-489.

Demski, J. and D. Sappington (1989): "Incentive Schemes with Multiple Agents and Bankruptcy Constraints," *Journal of Economic Theory*, 44, 156-167.

Esö, P. (2004): "An Optimal Auction with Correlated Values and Risk-Aversion," *Journal of Economic Theory*, 125, 78-89.

Forges, F., (1990): "Universal Mechanisms", *Econometrica,* 58: 1341-1364.

Fuchs, W. (2007): "Contracting with Repeated Moral Hazard and Subjective Evaluations," *American Economic Review*, 97, 1432-1448.

Gelfand, I. and S. Fomin, (2000), *Calculus of Variations,* Dover: Mineola.

Gerardi, D., (2002): "Unmediated Communication in Games with Complete and Incomplete Information", mimeo Cowles Foundation.

Gibbard, A. (1973): "Manipulation for Voting Schemes," *Econometrica*, 41, 617-631.

Green, J. and J.-J. Laffont (1977): "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, 45, 427-438.

Green, J. and J.-J. Laffont (1987): "Posterior Implementability in a Two-Person Decision Problem," *Econometrica*, 55, 69-94.

Gromb, D. and D. Martimort (2007): "Collusion and the Organization of Delegated Expertise," *Journal of Economic Theory*, 137, 271-299.

Guesnerie, R. and J.-J. Laffont (1984): "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329-369.

John, F. (1982): *Partial Differential Equations*, Berlin: Springer.

Johnson, S., J. Pratt and R. Zeckhauser (1990): "Efficiency Despite Mutually Payoff-Relevant Information: The Finite Case," *Econometrica*, 58, 873-900.

Krishna, V. and J. Morgan (2006): "Contracting for Information under Imperfect Commitment," *mimeo*, Penn State University.

Laffont, J.-J. and D. Martimort (1997): "Collusion under Asymmetric Information," *Econometrica*, 65, 875-912.

— (2000): "Mechanism Design with Collusion and Correlation," *Econometrica*, 68, 309-342.

— (2002): *The Theory of Incentives: The Principal-Agent Model*, Princeton: Princeton University Press.

Laffont, J.-J. and E. Maskin (1980): "A Differential Approach to Dominant Strategy Mechanisms," *Econometrica*, 48, 1507-1520.

Lewis, T. and D. Sappington (1989): "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, 49, 294-313.

Martimort, D. (1992): "Multi-Principaux avec Anti-Selection," *Annales d'Economie et de Statistiques*, 28, 1-38.

— (2007): "Multicontracting Mechanism Design," in *Advances in Economic Theory , Proceedings of the 2005 World Congress of the Econometric Society.*

Martimort, D. and H. Moreira (2009): "Common Agency with Informed Principals," forthcoming *Theoretical Economics.*

Martimort, D. and L. Stole (2002): "The Revelation and Delegation Principles in Common Agency Games," *Econometrica*, 70, 1659-1674.

— (2009): "Selecting Equilibria in Common Agency Games," *Journal of Economic Theory.*, 144: 604-644.

Maskin, E. and J. Tirole (1990): "The Principal-Agent Relationship with an Informed Principal, 1: Private Values," *Econometrica*, 58, 379-410.

Matsushima, H., (1991): "Incentive Compatible Mechanisms with Full Transferability," *Journal of Economic Theory*, 54, 198-203.

McAfee, P. and M. Schwartz (1994): "Opportunism in Multilateral Vertical Contracting: Non-Discrimination, Exclusivity and Uniformity," *American Economic Review*, 84, 210-230.

McAfee, P. and P. Reny (1992): "Correlated Information and Mechanism Design," *Econometrica*, 60, 395-421.

McLeod, B. (2003): "Optimal Contracting with Subjective Evaluations," *American Economic Review*, 93, 216-240.

Mezzetti, C. (1997): "Common Agency with Horizontally Differentiated Principals," *RAND Journal of Economics*, 28, 323-345.

Milgrom, P. and I. Segal, (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70(2), 583-601.

Myerson, R., (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Models", *Journal of Mathematical Economics,* 10: 67-81.

— (1983), "Mechanism Design with an Informed Principal", *Econometrica*, 51: 1767-1797.

— (1991): *Game Theory: Analysis of Conflict*, Harvard: Harvard University Press.

O'Brien, D. and G. Shaffer (1992): "Vertical Control with Bilateral Contracts," *RAND Journal of Economics,* 23, 299-308.

Peters, M. (2001): "Common Agency and the Revelation Principle," *Econometrica*, 69, 1349-1372.

Neeman, Z. (2004): "The Relevance of Private Information in Mechanism Design," *Journal of Economic Theory*, 117, 55-77.

Rahman, D. and I. Obara (2007): "Secret Contracts for Efficient Partnerships," mimeo University Minnesota.

Riordan, M. and D. Sappington (1988): "Optimal Contracts with Public Ex Post Information," *Journal of Economic Theory*, 45, 189-199.

Robert, J. (1991): "Continuity in Auction Design," *Journal of Economic Theory*, 55, 169-179.

Segal, I. (1999): "Contracting with Externalities," *Quarterly Journal of Economics*, 104, 337-388.

Stole, L. (1991): "Mechanism Design under Common Agency," mimeo University of Chicago.

Strausz, R. (2006): "Interim Information in Long-Term Contracts," *Journal of Economics and Management Startegy*, 15, 1041-1067.

Williamson, O. (1985): *The Economic Institutions of Capitalism*, New-York: The Free Press.

# Appendix

**Proof of Lemma 1.** Take any arbitrary mechanism $(g(\cdot), \mathcal{M}) = ((g_1(\cdot), \mathcal{M}), ..., (g_n(\cdot), \mathcal{M}))$ for any arbitrary communication space $\mathcal{M} = \Pi_{i=1}^n \mathcal{M}_i$. Consider also a perfect Bayesian continuation equilibrium of the overall contractual game induced by $(g(\cdot), \mathcal{M})$. Such continuation $PBE$ is a triplet $\{m^*(\cdot), \hat{m}^*(\cdot), d\mu(\theta|m)\}$ that satisfies:

• Agent $A_i$ with type $\theta_i$ reports a private message $m_i^*(\theta_i)$ to the principal. The strategy $m^*(\theta) = (m_1^*(\theta_1), ..., m_n^*(\theta_n))$ forms a Bayesian-Nash equilibrium among the agents. The corresponding equilibrium conditions are stated in (5) .

• $P$ updates his beliefs on the agents' types following Bayes' rule whenever possible, i.e, when $m \in \text{supp} \quad m^*(\cdot)$. Otherwise, beliefs are arbitrary. Let denote $d\mu(\theta|m)$ the updated beliefs following the observation of a vector of messages $m$.

• Given any such vector $m$ (either on or out of the equilibrium path) and the corresponding posterior beliefs, the principal reports the messages $(\hat{m}_{-1}^*(m), ..., \hat{m}_{-n}^*(m))$ which maximizes his expected payoff, i.e.,

$$(\hat{m}_{-1}^*(m), ..., \hat{m}_{-n}^*(m))$$

$$\in \arg \max_{(\hat{m}_{-1},\dots,\hat{m}_{-n})\in\Pi_{i=1}^{n}\mathcal{M}_{-i}} \int_{\Theta^n} \left\{ \tilde{S}(q_1(m_1,\hat{m}_{-1})),\dots,q_n(m_n,\hat{m}_{-n})) - \sum_{i=1}^{n} t_i(m_i,\hat{m}_{-i}) \right\} d\mu(\theta|m). \tag{A.1}$$

Because we are in a private values context where the agents' types do not enter directly into the principal's utility function, expectations do not matter and (A.1) can be rewritten more simply as (6).

**Proof of Proposition 1.** Consider the agents' Bayesian incentive compatibility conditions that must be satisfied by $m^*(\cdot)$. For $A_i$, we have for instance

$$m_i^*(\theta_i) \in \arg \max_{\tilde{m}_i\in\mathcal{M}_i} \mathop{E}_{\theta_{-i}} \left( t_i\left(\tilde{m}_i,\hat{m}^*_{-i}(\tilde{m}_i,m^*_{-i}(\theta_{-i}))\right) - \theta_i q_i\left(\tilde{m}_i,\hat{m}^*_{-i}(\tilde{m}_i,m^*_{-i}(\theta_{-i}))\right)|\theta_i\right).$$

The proof of a Revelation Principle will now proceed in two steps. In the first one, we replace the mechanism $(g(\cdot),\mathcal{M})$ by another mechanism $(\tilde{g}(\cdot),\mathcal{M})$ which is not manipulable by the principal. In the second step, we replace $(\tilde{g}(\cdot),\mathcal{M})$ by a direct, truthful and still non-manipulable mechanism $(\bar{g}(\cdot),\Theta^n)$.

**Step 1.** Consider the new mechanism $(\tilde{g}(\cdot),\mathcal{M})$ defined as:

$$\tilde{t}_i(m_i,m_{-i}) = t_i(m_i,\hat{m}_i^*(m_i,m_{-i})) \text{ and } \tilde{q}_i(m_i,m_{-i}) = q_i(m_i,\hat{m}^*_{-i}(m_i,m_{-i})) \text{ for } i=1,\dots,n. \tag{A.2}$$

**Lemma 4** $(\tilde{g}(\cdot),\mathcal{M})$ *is not manipulable by the principal, i.e.,* $\hat{m}^*_{-i}(m) = m \quad \forall m \in \mathcal{M}$ *given that* $\tilde{g}(\cdot)$ *is offered.*

**Proof.** Fix any $m=(m_1,\dots,m_n)\in\mathcal{M}$. By (6), we have:

$$\tilde{S}(q_1(m_1,\hat{m}^*_{-1}(m)),..,q_n(m_n,\hat{m}^*_{-n}(m))) - \sum_{i=1}^{n} t_i(m_i,\hat{m}^*_{-i}(m))$$

$$\geq \tilde{S}(q_1(m_1,\tilde{m}_{-1}),\dots,q_n(m_n,\tilde{m}_{-n})) - \sum_{i=1}^{n} t_i(m_i,\tilde{m}_{-i}) \quad \forall(\tilde{m}_{-1},\dots,\tilde{m}_{-n}).$$

In particular, we get:

$$\tilde{S}(q_1(m_1,\hat{m}^*_{-1}(m)),..,q_n(m_n,\hat{m}^*_{-n}(m))) - \sum_{i=1}^{n} t_i(m_i,\hat{m}^*_{-i}(m))$$

$$\geq \tilde{S}(q_1(m_1,\hat{m}^*_{-1}(m_1,m'_{-1})),\dots,q_n(m_n,\hat{m}^*_{-n}(m_n,m'_{-n}))) - \sum_{i=1}^{n} t_i(m_i,\hat{m}^*_{-i}(m_i,m'_{-i})). \tag{A.3}$$

Then, using the definition of $\tilde{g}(\cdot)$ given in (A.2), (A.3) ensures that $\forall(m'_{-1},..m'_{-n})$:

$$\tilde{S}(\tilde{q}(m)) - \sum_{i=1}^{n} \tilde{t}_i(m) \geq \tilde{S}(\tilde{q}_1(m_1,m'_{-1}),..,\tilde{q}_n(m_n,m'_n)) - \sum_{i=1}^{n} \tilde{t}_i(m_i,m'_{-i}). \tag{A.4}$$

Given that $\tilde{g}(\cdot)$ is played, the best manipulation made by the principal is $\hat{m}^*_{-i}(m) = m$ for all $m$. $\tilde{g}(\cdot)$ is not manipulable by the principal. ■

It is straightforward to check that the new mechanism $\tilde{g}(\cdot)$ still induces an equilibrium strategy vector $m^*(\theta) = (m_1^*(\theta_1), ..., m_n^*(\theta_n))$ for the agents. Indeed, $m^*(\cdot)$ satisfies by definition the following Bayesian-Nash constraints:

$$m_i^*(\theta_i) \in \arg\max_{m_i \in \mathcal{M}_i} \underset{\theta_{-i}}{E} \left( t_i(m_i, \hat{m}^*_{-i}(m_i, m^*_{-i}(\theta_{-i}))) - \theta_i q_i(m_i, \hat{m}^*_{-i}(m_i, m^*_{-i}(\theta_{-i}))) | \theta_i \right)$$

which can be rewritten as:

$$m_i^*(\theta_i) \in \arg\max_{m_i \in \mathcal{M}_i} \underset{\theta_{-i}}{E} \left( \tilde{t}_i(m_i, m^*_{-i}(\theta_{-i})) - \theta_i \tilde{q}_i(m_i, m^*_{-i}(\theta_{-i})) | \theta_i \right). \qquad \text{(A.5)}$$

Hence, $m^*(\cdot)$ is a Bayesian-Nash equilibrium of the new mechanism $\tilde{g}(\cdot)$.

**Step 2.** Consider now the direct revelation mechanism $(\bar{g}(\cdot), \Theta^n)$ defined as:

$$\bar{t}_i(\theta) = \tilde{t}_i(m^*(\theta)) \text{ and } \bar{q}_i(\theta) = \tilde{q}_i(m^*(\theta)) \quad \text{for } i = 1, ..., n. \qquad \text{(A.6)}$$

**Lemma 5** $\bar{g}(\cdot)$ *is truthful in Bayesian incentive compatibility and not manipulable.*

**Proof.** First consider the non-manipulability of the mechanism $\bar{g}(\cdot)$. From (A.4), we get:

$$\tilde{S}\left(\bar{q}(\theta)\right) - \sum_{i=1}^{n} \bar{t}_i(\theta) \geq$$

$$\tilde{S}\left(\tilde{q}_1(m_1^*(\theta_1), m'_{-1}), .., \tilde{q}_n(m_n^*(\theta_n), m'_n)\right) - \sum_{i=1}^{n} \tilde{t}_i(m_i^*(\theta_i), m'_{-i}) \quad \forall m'_{-i} \in \mathcal{M}_{-i}. \qquad \text{(A.7)}$$

Taking $m'_{-i} = m^*_{-i}(\theta'_{-i})$, (A.7) becomes

$$\tilde{S}\left(\bar{q}(\theta)\right) - \sum_{i=1}^{n} \bar{t}_i(\theta) \geq$$

$$\tilde{S}(\bar{q}_1(\theta_1, \theta'_{-1}), .., \bar{q}_n(\theta_n, \theta'_{-n})) - \sum_{i=1}^{n} \bar{t}_i(\theta_i, \theta'_{-i}) \quad \forall (\theta'_{-1}, ..., \theta'_{-n}). \qquad \text{(A.8)}$$

Hence, $\bar{g}(\cdot)$ is non-manipulable.

Turning to (A.5), it is immediate to check that the agents' Bayesian incentive constraints can be written as:

$$\theta_i \in \arg\max_{\hat{\theta}_i \in \Theta} \underset{\theta_{-i}}{E} \left( \bar{t}_i(\hat{\theta}_i, \theta_{-i}) - \theta_i \bar{q}_i(\hat{\theta}_i, \theta_{-i}) | \theta_i \right). \qquad \text{(A.9)}$$

■

**Proof of Proposition 2.** Let us define

$$\tilde{U}_i(\hat{\theta}_i,\theta_i)=\underset{\theta_{-i}}{E}\left(S(q_i(\hat{\theta}_i,\theta_{-i}))-\theta_i q_i(\hat{\theta}_i,\theta_{-i})|\theta_i\right)-H_i(\hat{\theta}_i).$$

$\tilde{U}_i(\hat{\theta}_i,\cdot)$ is differentiable for all $\hat{\theta}_i$. Without loss of generality we can restrict attention to quantity schedules that are bounded above by $\bar{q}$ large enough. Therefore there exists an integrable function $b(\theta_i)$ such that

$$\left|\frac{\partial\tilde{U}_i}{\partial\theta_i}(\hat{\theta}_i,\theta_i)\right|=\left|\underset{\theta_{-i}}{E}\left(q_i(\hat{\theta}_i,\theta_{-i})-(S(q_i(\hat{\theta}_i,\theta_{-i}))-\theta_i q_i(\hat{\theta}_i,\theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\Big|\theta_i\right)\right|\le b(\theta_i),$$

for all $\hat{\theta}_i$ and almost all $\theta_i$. We can now apply Theorem 2 in Milgrom and Segal (2002, p. 586) to ensure that

$$U_i(\theta_i)=U_i(\bar{\theta})+\int_{\theta_i}^{\bar{\theta}}\underset{\theta_{-i}}{E}\left(q_i(x,\theta_{-i})-(S(q_i(x,\theta_{-i}))-xq_i(x,\theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|x)}{\tilde{f}(\theta_{-i}|x)}\Big|x\right)dx.$$

Therefore, we obtain:

$$\underset{\theta_i}{E}(U_i(\theta_i))=U_i(\bar{\theta})$$
$$+\int_{\underline{\theta}}^{\bar{\theta}}f(\theta_i)\left(\int_{\theta_i}^{\bar{\theta}}\underset{\theta_{-i}}{E}\left(q_i(x,\theta_{-i})-(S(q_i(x,\theta_{-i}))-xq_i(x,\theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|x)}{\tilde{f}(\theta_{-i}|x)}\Big|x\right)dx\right)d\theta_i.$$

Integrating by parts yields

$$\underset{\theta_i}{E}(U_i(\theta_i))=U_i(\bar{\theta})+\underset{\theta}{E}\left(\frac{F(\theta_i)}{f(\theta_i)}\left(q_i(\theta)-(S(q_i(\theta))-\theta_i q_i(\theta))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\right)\right). \tag{A.10}$$

First, let us suppose that (14) is binding only at $\theta_i=\bar{\theta}$. Of course minimizing the agents' information rent requires to set $U_i(\bar{\theta})=0$ when the right-hand side in (17) is negative; something that will be checked later. Inserting (A.10) into the principal's objective function and optimizing pointwise yields (19).

*Monotonicity conditions.* Assumption 1 and strict concavity of $S(\cdot)$ immediately imply that $\frac{\partial q^{SB}}{\partial\theta_{-i}}(\theta_i,\theta_{-i})\ge 0$ and $\frac{\partial q^{SB}}{\partial\theta_i}(\theta_i,\theta_{-i})<0$.

*Monotonicity of $U^{SB}(\theta_i)$.* This monotonicity is ensured whenever the following sufficient condition holds

$$q^{SB}(\theta_i,\theta_{-i})\ge\left(S(q^{SB}(\theta_i,\theta_{-i}))-\theta_i q^{SB}(\theta_i,\theta_{-i})\right)\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \tag{A.11}$$

since then integrating over $\theta_{-i}$ yields that the right-hand side of (17) is negative and thus $U_i(\theta_i)$ is non-increasing as supposed. Note that (A.11) holds when $\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\le 0$. When instead $\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} > 0$, we have $q^{FB}(\theta_i) > q^{SB}(\theta_i, \theta_{-i}) > q^{BM}(\theta_i)$. Therefore, a sufficient condition for (A.11) is

$$\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \leq \frac{q^{BM}(\theta_i)}{S(q^{FB}(\theta_i)) - \theta_i q^{FB}(\theta_i)}$$

as requested in Assumption 2.

*Second-order conditions.* For $q^{SB}(\theta_i, \theta_{-i})$ the local second-order condition (18) becomes

$$\underset{\theta_{-i}}{E}\left(\frac{\frac{\partial q^{SB}}{\partial \theta_i}(\theta_i, \theta_{-i})}{1 + \frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\frac{F(\theta_i)}{f(\theta_i)}}\bigg|\theta_i\right) \geq 0$$

which holds since $\frac{\partial q^{SB}}{\partial \theta_i}(\theta_i, \theta_{-i}) \leq 0$ from Assumption 1 and $1 + \frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\frac{F(\theta_i)}{f(\theta_i)} > 0$ from Assumption 2.

*Global incentive compatibility.* Observe that

$$\frac{\partial U^{SB}}{\partial \hat{\theta}_i}(\hat{\theta}_i, \theta_i) = \underset{\theta_{-i}}{E}\left((S'(q^{SB}(\hat{\theta}_i, \theta_{-i})) - \theta_i)\frac{\partial q^{SB}}{\partial \hat{\theta}_i}(\hat{\theta}_i, \theta_{-i})|\theta_i\right) - \dot{H}^{SB}(\hat{\theta}_i).$$

Taking into account the first-order condition (16), we get:

$$\frac{\partial U^{SB}}{\partial \hat{\theta}_i}(\hat{\theta}_i, \theta_i)$$

$$= \underset{\theta_{-i}}{E}\left((S'(q^{SB}(\hat{\theta}_i, \theta_{-i})) - \theta_i)\frac{\partial q^{SB}}{\partial \hat{\theta}_i}(\hat{\theta}_i, \theta_{-i})|\theta_i\right) - \underset{\theta_{-i}}{E}\left((S'(q^{SB}(\hat{\theta}_i, \theta_{-i})) - \hat{\theta}_i)\frac{\partial q^{SB}}{\partial \hat{\theta}_i}(\hat{\theta}_i, \theta_{-i})|\hat{\theta}_i\right)$$

$$= \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial q^{SB}}{\partial \hat{\theta}_i}(\hat{\theta}_i, \theta_{-i})\left(\left(\hat{\theta}_i - \theta_i + \frac{\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}}{1 + \frac{\tilde{f}_{\theta_i}(\theta_{-i}|\hat{\theta}_i)}{\tilde{f}(\theta_{-i}|\hat{\theta}_i)}\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}}\right)\tilde{f}(\theta_{-i}|\theta_i) - \frac{\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}}{1 + \frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\hat{\theta}_i)}{\tilde{f}(\theta_{-i}|\hat{\theta}_i)}}\tilde{f}(\theta_{-i}|\hat{\theta}_i)\right)d\theta_{-i}$$

$$= \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial q^{SB}}{\partial \hat{\theta}_i}(\hat{\theta}_i, \theta_{-i})\psi(\hat{\theta}_i, \theta_i, \theta_{-i})d\theta_{-i}$$

where

$$\psi(\hat{\theta}_i, \theta_i, \theta_{-i}) = \left(\hat{\theta}_i - \theta_i + \frac{\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}}{1 + \frac{\tilde{f}_{\theta_i}(\theta_{-i}|\hat{\theta}_i)}{\tilde{f}(\theta_{-i}|\hat{\theta}_i)}\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}}\right)\tilde{f}(\theta_{-i}|\theta_i) - \frac{\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}}{1 + \frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\hat{\theta}_i)}{\tilde{f}(\theta_{-i}|\hat{\theta}_i)}}\tilde{f}(\theta_{-i}|\hat{\theta}_i).$$

Observe first that $\psi(\theta_i, \theta_i, \theta_{-i}) = 0$. By the Intermediate Values Theorem, we have $\tilde{f}(\theta_{-i}|\hat{\theta}_i) - \tilde{f}(\theta_{-i}|\theta_i) = \tilde{f}_{\theta_i}(\theta_{-i}|\xi)(\hat{\theta}_i - \theta_i)$, for some $\xi \in ]\theta_i, \hat{\theta}_i[$. Denote $M = \max_{(\theta_i, \theta_{-i}) \in \Theta^2} \tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)$, we have $\tilde{f}(\theta_{-i}|\hat{\theta}_i) - \tilde{f}(\theta_{-i}|\theta_i) \leq M(\hat{\theta}_i - \theta_i)$ for $\hat{\theta}_i \geq \theta_i$ and $\psi(\hat{\theta}_i, \theta_i, \theta_{-i}) \geq 0$ for $\hat{\theta}_i \geq \theta_i$ when

$$M \leq \tilde{f}(\theta_{-i}|\theta_i)\left(\frac{1 + \frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\hat{\theta}_i)}{\tilde{f}(\theta_{-i}|\hat{\theta}_i)}}{\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}}\right). \quad \text{(A.12)}$$

Since $\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\hat{\theta}_i)}{\tilde{f}(\theta_{-i}|\hat{\theta}_i)} \geq -\frac{M}{\min_{(\hat{\theta}_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\hat{\theta}_i)}$, (A.12) holds when

$$M \leq \frac{f(\hat{\theta}_i)}{F(\hat{\theta}_i)} \frac{\tilde{f}(\theta_{-i}|\theta_i)}{1+\frac{\tilde{f}(\theta_{-i}|\theta_i)}{\min_{(\hat{\theta}_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\hat{\theta}_i)}} \quad \forall\hat{\theta}_i \geq \theta_i, \forall\theta_{-i}.$$

A sufficient condition for this is

$$M \leq \min_{\theta_i\in\Theta} f(\theta_i)\frac{(\min_{(\theta_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\theta_i))^2}{2\max_{(\theta_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\theta_i)}$$

as requested by Assumption 2 since then

$$\min_{\theta_i\in\Theta} f(\theta_i)\frac{(\min_{(\theta_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\theta_i))^2}{2\max_{(\theta_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\theta_i)} \leq \min_{\theta_i\in\Theta} f(\theta_i)\frac{\tilde{f}(\theta_{-i}|\theta_i)}{1+\frac{\tilde{f}(\theta_{-i}|\theta_i)}{\min_{(\hat{\theta}_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\hat{\theta}_i)}}$$

$$\leq \frac{f(\hat{\theta}_i)}{F(\hat{\theta}_i)}\frac{\tilde{f}(\theta_{-i}|\theta_i)}{1+\frac{\tilde{f}(\theta_{-i}|\theta_i)}{\min_{(\hat{\theta}_i,\theta_{-i})\in\Theta^2}\tilde{f}(\theta_{-i}|\hat{\theta}_i)}}.$$

Turning now to the case $\hat{\theta}_i < \theta_i$. Note that we have then $\tilde{f}(\theta_{-i}|\hat{\theta}_i) - \tilde{f}(\theta_{-i}|\theta_i) \geq M(\hat{\theta}_i - \theta_i)$ for $\hat{\theta}_i \leq \theta_i$. Therefore, we get:

$$\psi(\hat{\theta}_i, \theta_i, \theta_{-i}) \leq (\hat{\theta}_i - \theta_i)\left(\tilde{f}(\theta_{-i}|\theta_i) + M\frac{\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}}{1+\frac{F(\hat{\theta}_i)}{f(\hat{\theta}_i)}\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\hat{\theta}_i)}{\tilde{f}(\theta_{-i}|\hat{\theta}_i)}}\right) \leq 0 \text{ for } \hat{\theta}_i < \theta_i$$

when Assumption 2 holds. ■

**Proof of Lemma 2.** Starting from (8) and writing a first-order condition yields (22). To prove that those conditions are also locally sufficient, denote the principal's ex post profit as:

$$V(\hat{\theta}, \theta) = \tilde{S}(q_1(\theta_1, \hat{\theta}_2), q_2(\hat{\theta}_1, \theta_2)) - t_1(\theta_1, \hat{\theta}_2) - t_2(\hat{\theta}_1, \theta_2).$$

We have:

$$\frac{\partial^2 V}{\partial\hat{\theta}_2^2}(\hat{\theta}, \theta) = \frac{\partial^2\tilde{S}}{\partial q_1^2}(q_1(\theta_1, \hat{\theta}_2), q_2(\hat{\theta}_1, \theta_2))\left(\frac{\partial q_1}{\partial\theta_2}(\theta_1, \hat{\theta}_2)\right)^2$$

$$+\frac{\partial\tilde{S}}{\partial q_1}(q_1(\theta_1, \hat{\theta}_2), q_2(\hat{\theta}_1, \theta_2))\frac{\partial^2 q_1}{\partial\theta_2^2}(\theta_1, \hat{\theta}_2) - \frac{\partial^2 t_1}{\partial\theta_2^2}(\theta_1, \hat{\theta}_2).$$

Taking into account (22) and differentiating with respect to $\theta_2$ yields:

$$\frac{\partial^2\tilde{S}}{\partial q_1^2}(q(\theta))\left(\frac{\partial q_1}{\partial\theta_2}\right)^2 + \frac{\partial\tilde{S}}{\partial q_1}(q(\theta))\frac{\partial^2 q_1}{\partial\theta_2^2}(\theta) - \frac{\partial^2 t_1}{\partial\theta_2^2}(\theta) = -\frac{\partial^2\tilde{S}}{\partial q_1\partial q_2}(q(\theta))\frac{\partial q_1}{\partial\theta_2}(\theta)\frac{\partial q_2}{\partial\theta_2}(\theta).$$

Finally, we get:

$$\frac{\partial^2 V}{\partial\hat{\theta}_2^2}(\theta, \theta) = -\frac{\partial^2\tilde{S}}{\partial q_1\partial q_2}(q(\theta))\frac{\partial q_1}{\partial\theta_2}(\theta)\frac{\partial q_2}{\partial\theta_2}(\theta) \leq 0$$

when Assumption 3 holds.

Direct computations yields also $\frac{\partial^2 V}{\partial\hat{\theta}_1\partial\hat{\theta}_2}(\theta,\theta) = \frac{\partial^2\tilde{S}}{\partial q_1\partial q_2}(q(\theta))\frac{\partial q_1}{\partial\theta_2}(\theta)\frac{\partial q_2}{\partial\theta_1}(\theta)$. Finally, we have:

$$\frac{\partial^2 V}{\partial\hat{\theta}_2^2}(\theta,\theta)\frac{\partial^2 V}{\partial\hat{\theta}_1^2}(\theta,\theta)-\left(\frac{\partial^2 V}{\partial\hat{\theta}_1\partial\hat{\theta}_2}(\theta,\theta)\right)^2$$
$$=\left(\frac{\partial^2\tilde{S}}{\partial q_1\partial q_2}(q(\theta))\right)^2\frac{\partial q_1}{\partial\theta_2}(\theta)\frac{\partial q_2}{\partial\theta_1}(\theta)\left(\frac{\partial q_1}{\partial\theta_1}(\theta)\frac{\partial q_2}{\partial\theta_2}(\theta)-\frac{\partial q_1}{\partial\theta_2}(\theta)\frac{\partial q_2}{\partial\theta_1}(\theta)\right)\geq 0$$

which ensures concavity of the principal's problem at $\hat{\theta}=\theta$ when Assumption 4 holds. ■

**Proof of Lemma 3.** To prove global optimality of a non-manipulable strategy, it turns out that an approach in terms of nonlinear prices helps. Define thus $T_i(q_i,\theta_i)=t_i(\theta_i,\theta_{-i})$ for $\theta_{-i}$ such that $q_i(\theta_i,\theta_{-i})$. From (8), this definition is without any ambiguity because all type $\theta_{-i}$ which corresponds to the same output $q_i(\theta_i,\theta_{-i})$ must also correspond to the same transfer $t_i(\theta_i,\theta_{-i})$ otherwise a valuable manipulation would be feasible. Assume now that $\frac{\partial q_i}{\partial\theta_{-i}}(\theta_i,\theta_{-i})\neq 0$ so that $q_i(\theta_i,\theta_{-i})$ is invertible with respect to $\theta_{-i}$. Denote $\Theta_{-i}(q_i,\theta_i)$ the inverse function.

The non-manipulability constraints can be rewritten as:

$$q(\theta)=(q_1(\theta),q_2(\theta))\in\arg\max_q\tilde{S}(q_1,q_2)-\sum_{i=1}^{2}T_i(q_i,\theta_i)=J(\theta,q). \tag{A.13}$$

It can be checked that:

$$\frac{\partial T_i}{\partial q_i}(q_i,\theta_i)=\frac{\partial\tilde{S}}{\partial q_i}(q_i,q_{-i}(\theta_i,\Theta_{-i}(q_i,\theta_i))).$$

So that the first-order conditions for (A.13) defines $q(\theta)$. The local analysis above also proves that second-order conditions are always locally satisfied.

We turn next to the global concavity of $J(\theta,q)$. Observe that:

$$\frac{\partial^2 T_i}{\partial q_i^2}(q_i,\theta_i)=\frac{\partial^2\tilde{S}^2}{\partial q_i^2}(q_i,q_{-i}(\theta_i,\Theta_{-i}(q_i,\theta_i)))+\frac{\partial^2\tilde{S}^2}{\partial q_i\partial q_{-i}}(q_i,q_{-i}(\theta_i,\Theta_{-i}(q_i,\theta_i)))\frac{\frac{\partial q_{-i}}{\partial\theta_{-i}}(\theta_i,\Theta_{-i}(q_i,\theta_i))}{\frac{\partial q_i}{\partial\theta_{-i}}(\theta_i,\Theta_{-i}(q_i,\theta_i))}.$$

Assume now that $\frac{\partial^2\tilde{S}^2}{\partial q_i\partial q_{-i}}(q_i,q_{-i})=\lambda$, then $\frac{\partial^3\tilde{S}^2}{\partial q_i^2\partial q_{-i}}(q_i,q_{-i})=0$ so that $\frac{\partial^2\tilde{S}^2}{\partial q_i^2}(q_i,q_{-i}(\theta_i,\Theta_{-i}(q_i,\theta_i)))=\frac{\partial^2\tilde{S}^2}{\partial q_i^2}(q_i,q_{-i})$ for any $q_{-i}$. From Assumption 3 and $\frac{\partial q_i}{\partial\theta_{-i}}(\theta_i,\theta_{-i})\neq 0$, we finally get:

$$\frac{\partial^2 J}{\partial q_i^2}(\theta,q)=-\frac{\partial^2 T_i}{\partial q_i^2}(q_i,\theta_i)+\frac{\partial^2\tilde{S}^2}{\partial q_i^2}(q_i,q_{-i})=\lambda\frac{\frac{\partial q_{-i}}{\partial\theta_{-i}}(\theta_i,\Theta_{-i}(q_i,\theta_i))}{\frac{\partial q_i}{\partial\theta_{-i}}(\theta_i,\Theta_{-i}(q_i,\theta_i))}\leq 0.$$

Similarly, we have:

$$\left(\frac{\partial^2 J}{\partial q_1^2}\frac{\partial^2 J}{\partial q_2^2}-\left(\frac{\partial^2 J}{\partial q_1\partial q_2}\right)^2\right)(\theta,q)=\lambda^2\left(\frac{\frac{\partial q_2}{\partial\theta_2}(\theta_1,\Theta_2(q_1,\theta_1))}{\frac{\partial q_1}{\partial\theta_2}(\theta_1,\Theta_2(q_1,\theta_1))}\frac{\frac{\partial q_1}{\partial\theta_1}(\Theta_1(q_2,\theta_2),\theta_2)}{\frac{\partial q_2}{\partial\theta_1}(\Theta_1(q_2,\theta_2),\theta_2)}-1\right)\geq 0$$

when Assumption 4 holds. ■

**Proof of Proposition 3.** First, observe that condition (24) allows us to express the agent's incentive compatibility constraint as:

$$U_i(\theta_i) = \arg\max_{\hat{\theta}_i \in \Theta} \underset{\theta_{-i}}{E}\left(\int_{\underline{\theta}}^{\theta_{-i}} \frac{\partial \tilde{S}}{\partial q_i}(q(\hat{\theta}_i, x))\frac{\partial q_i}{\partial \theta_{-i}}(\hat{\theta}_i, x)dx - \theta_i q_i(\hat{\theta}_i, \theta_{-i})|\theta_i\right) - H_i(\hat{\theta}_i). \quad \text{(A.14)}$$

Using the Envelope Theorem yields the expression of the derivative of agent's information rent for a given differentiable output vector $q(\theta)$:

$$\dot{U}_i(\theta_i) = -\underset{\theta_{-i}}{E}\left(q_i(\theta_i, \theta_{-i})|\theta_i\right) + \underset{\theta_{-i}}{E}\left(\left(\int_{\underline{\theta}}^{\theta_{-i}} \frac{\partial \tilde{S}}{\partial q_i}(q(\theta_i, x))\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, x)dx - \theta_i q_i(\theta_i, \theta_{-i})\right)\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right). \quad \text{(A.15)}$$

Integrating by parts the second term of (A.15) and taking into account that $\int_{\underline{\theta}}^{\bar{\theta}} \tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)d\theta_{-i} = 0$ yields a new expression of $\dot{U}_i(\theta_i)$ as

$$\dot{U}_i(\theta_i) = -\underset{\theta_{-i}}{E}\left(q_i(\theta_i, \theta_{-i})|\theta_i\right) - \int_{\underline{\theta}}^{\bar{\theta}}\left(\int_{\underline{\theta}}^{\theta_{-i}} \tilde{f}_{\theta_i}(x|\theta_i)dx\right)\left(\frac{\partial \tilde{S}}{\partial q_i}(q(\theta_i, \theta_{-i}))dx - \theta_i\right)\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, \theta_{-i})d\theta_{-i}. \quad \text{(A.16)}$$

From (A.16), agent $A_i$'s information rent is decreasing when Assumption 2 holds and thus (14) is binding at $\bar{\theta}$. This yields the following expression of $A_i$'s expected rent:

$$\underset{\theta_i}{E}\left(U_i(\theta_i)\right) = \underset{\theta}{E}\left(\frac{F(\theta_i)}{f(\theta_i)}q_i(\theta)\right)$$

$$+\underset{\theta}{E}\left(\frac{F(\theta_i)}{\tilde{f}(\theta_i, \theta_{-i})}\left(\int_{\underline{\theta}}^{\theta_{-i}} \tilde{f}_{\theta_i}(x|\theta_i)dx\right)\left(\frac{\partial \tilde{S}}{\partial q_i}(q(\theta_i, \theta_{-i}))dx - \theta_i\right)\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, \theta_{-i})\right).$$

Inserting these expected rents into the principal's objective function yields the following calculus of variations problem:

$$(\mathcal{P}) : \max_{\{q(\cdot)\}} \int_{\Theta^2} \Phi\left(\theta, q(\theta), \nabla q(\theta)\right) d\theta$$

where admissible arcs $q(\cdot)$ are in $\mathcal{C}^1$, and

$$\Phi\left(\theta, q(\theta), \nabla q(\theta)\right) = \tilde{f}(\theta)\left(\tilde{S}(q(\theta)) - \sum_{i=1}^{2}\left(\theta_i + \frac{F(\theta_i)}{f(\theta_i)}\right)q_i(\theta)\right)$$

$$-\sum_{i=1}^{2} F(\theta_i)\left(\int_{\underline{\theta}}^{\theta_{-i}} \tilde{f}_{\theta_i}(x|\theta_i)dx\right)\left(\frac{\partial \tilde{S}}{\partial q_i}(q(\theta_i, \theta_{-i})) - \theta_i\right)\frac{\partial q_i}{\partial \theta_{-i}}(\theta).$$

Given that Assumption 2 holds, $\Phi\left(\theta, s, v\right)$ is concave in $(s, v)$, the necessary conditions for optimality are also sufficient. The first necessary conditions are the Euler-Lagrange conditions[40] satisfied by $q^{SB}(\theta)$. They are obtained by looking at variations of the functional

[40]Gelfand and Fomin (2000, p. 153).

inside the square $\Theta^2$ and can be written as:

$$\Phi_{q_i} = \sum_{k=1}^{2} \frac{\partial \Phi_{q_{i\theta_k}}}{\partial \theta_k} \quad \text{for } i = 1, 2. \tag{A.17}$$

Simplifying yields (26).

The second set of necessary conditions for optimality is obtained by looking at variations of the functional on the boundary $\Gamma$ of $\Theta^2$. They can be written as:

$$\int_\Gamma \psi(\theta) \overrightarrow{G}_i \cdot d\overrightarrow{n} = 0 \quad \text{for } i = 1, 2. \tag{A.18}$$

for any function $\psi(\theta) \in \mathcal{C}^1$ where $d\overrightarrow{n}$ is the normal outward to $\Gamma$ and $\overrightarrow{G}_i = (\Phi_{q_{i\theta_1}}, \Phi_{q_{i\theta_2}})$. These conditions are obviously satisfied since terms of the form $F(\theta_i)\left(\int_{\underline{\theta}}^{\theta_{-i}} \tilde{f}_{\theta_i}(x|\theta_i)dx\right)$ are zero on the boundary. The boundary conditions (27) come from taking $\theta_i = \underline{\theta}$ into (26).

**•Using characteristics to approximate solutions close to the boundary surfaces:** When $\frac{\partial^2 \tilde{S}}{\partial q_1 \partial q_2} = \lambda > 0$, we can rewrite the system of partial differential equations (26) as:

$$a(\theta_1, \theta_2)\frac{\partial q_1^{SB}}{\partial \theta_1}(\theta) - a(\theta_2, \theta_1)\frac{\partial q_1^{SB}}{\partial \theta_2}(\theta) =$$

$$-\frac{\tilde{f}(\theta)}{\lambda}\left(\left(1 + \frac{F(\theta_2)}{f(\theta_2)}\frac{\tilde{f}_{\theta_2}(\theta_1|\theta_2)}{\tilde{f}(\theta_1|\theta_2)}\right)\left(\frac{\partial \tilde{S}}{\partial q_2}(q^{SB}(\theta)) - \theta_2\right) - \frac{F(\theta_2)}{f(\theta_2)}\right), \tag{A.19}$$

$$a(\theta_1, \theta_2)\frac{\partial q_2^{SB}}{\partial \theta_1}(\theta) - a(\theta_2, \theta_1)\frac{\partial q_2^{SB}}{\partial \theta_2}(\theta) =$$

$$\frac{\tilde{f}(\theta)}{\lambda}\left(\left(1 + \frac{F(\theta_1)}{f(\theta_1)}\frac{\tilde{f}_{\theta_1}(\theta_2|\theta_1)}{\tilde{f}(\theta_2|\theta_1)}\right)\left(\frac{\partial \tilde{S}}{\partial q_1}(q^{SB}(\theta)) - \theta_1\right) - \frac{F(\theta_1)}{f(\theta_1)}\right) \tag{A.20}$$

where $a(\theta_1, \theta_2) = F(\theta_2)\left(\int_{\underline{\theta}}^{\theta_1} \tilde{f}_{\theta_2}(x|\theta_2)dx\right) \leq 0$.

Let introduce the variable $z \in \mathbb{R}$ to parameterize characteristic curves which are tangent at each point to the surfaces $q_i = q_i^{SB}(\theta)$ defined through the system (A.19)-(A.20). We set:

$$\frac{d\theta_1}{dz}(z) = a(\theta_1, \theta_2) \text{ and } \frac{d\theta_2}{dz}(z) = -a(\theta_2, \theta_1). \tag{A.21}$$

Let $Q(z) = (Q_1(z), Q_2(z)) = q^{SB}(\theta(z))$. Equations (A.19) and (A.20) define a system of differential equations such that for $i = 1, 2$:

$$\dot{Q}_i(z) = \frac{(-1)^i \tilde{f}(\theta(z))}{\lambda}\left(\left(1 + \frac{F(\theta_{-i}(z))}{f(\theta_{-i}(z))}\frac{\tilde{f}_{\theta_{-i}}(\theta_i(z)|\theta_{-i}(z))}{\tilde{f}(\theta_i(z)|\theta_{-i}(z))}\right)\left(\frac{\partial \tilde{S}}{\partial q_{-i}}(Q(z)) - \theta_{-i}(z)\right) - \frac{F(\theta_{-i}(z))}{f(\theta_{-i}(z))}\right). \tag{A.22}$$

At this stage the difficulty in using the standard method of characteristics (as in John (1982) for instance) comes from the fact that the boundary conditions (27) correspond to characteristic curves. Nevertheless, with a little bit more work, we can prove existence (locally around the boundary) and provide an approximation for a solution to (26)-(27).

Let choose the initial values for $z = 0$ as

$$\theta_1(0) = \theta_2(0) = \theta_0 \in (\underline{\theta}, \bar{\theta}). \tag{A.23}$$

Since $a(\cdot)$ is continuous and satisfies a Lipschitz condition, the Uniqueness Theorem for ordinary differential equations ensures that the system (A.21) with these initial conditions has a unique solution. It can be easily checked that $\theta_1(z)$ (resp. $\theta_2(z)$) is strictly decreasing (resp. increasing). Moreover, in the $(\theta_1, \theta_2)$ space the curve corresponding to the solution of (A.21)-(A.23) cannot reach the boundary $\theta_2 = \bar{\theta}$ for some $z_0$ such that $\theta_1(z_0) > \underline{\theta}$ because the unique solution to (A.21) such that, for some finite $z_0$ we have $\theta_2(z_0) = \bar{\theta}$ and $\theta_1(z_0) > \underline{\theta}$, is such that $\theta_2(z) = \bar{\theta}$ for all $z$ by the Uniqueness Theorem for ordinary differential equations. This would contradict the initial conditions (A.23). Moreover, because $\theta_1(z)$ (resp. $\theta_2(z)$) is decreasing and thus bounded below by $\underline{\theta}$ (resp. increasing) (bounded above by $\bar{\theta}$), it has a limit when $z \to +\infty$ and this limit has to be $\underline{\theta}$ (resp. some $\theta_2^*$ such that $\theta_2^* < \bar{\theta}$). (Note that the limit is not reached. Indeed, by the Uniqueness Theorem, there exists a unique solution to (A.21) with the conditions $\theta_1(z_1) = \underline{\theta}$ and $\theta_2(z_1) = \theta_2^*$ for some $z_1 < +\infty$ and this limit is such that $\theta_1(z) = \underline{\theta}$ and $\theta_2(z) = \theta_2^*$ for all $z$ contradicting (A.23).) Note for each initial condition $\theta_0$, the corresponding value of $\theta_2^*$ as $\theta_2^*(\theta_0)$. This function is weakly increasing (otherwise, there would be a contradiction with the Uniqueness Theorem for differential equations), obviously continuous in $\theta_0$ and such that first, since $\theta_2^*(\theta_0) \geq \theta_0$ we have $lim_{\theta_0 \to \bar{\theta}} \theta_2^*(\theta) = \bar{\theta}$, and second $\theta_2^*(\underline{\theta}) = \underline{\theta}$. Hence, any $\theta_2^* \in \Theta$ is the limit of a schedule $\theta_2(z)$ for some initial condition $\theta_0$.

To understand how the system $(\theta_1(z), \theta_2(z))$ behaves as $z \to +\infty$, observes that (A.21) can be approximated as:

$$\dot{\theta}_1(z) = F(\theta_2^*) \tilde{f}_\theta(\underline{\theta}|\theta_2^*)(\theta_1(z) - \underline{\theta}) \text{ and } \dot{\theta}_2(z) = -f(\underline{\theta}) \left( \int_{\underline{\theta}}^{\theta_2^*} \tilde{f}_\theta(x|\underline{\theta}) dx \right) (\theta_1(z) - \underline{\theta}). \tag{A.24}$$

Integrating yields:

$$\theta_1(z) = \underline{\theta} + \mu exp(F(\theta_2^*) \tilde{f}_\theta(\underline{\theta}|\theta_2^*) z) \text{ and } \theta_2(z) = \theta_2^* - \mu \frac{f(\underline{\theta}) \left( \int_{\underline{\theta}}^{\theta_2^*} \tilde{f}_\theta(x|\underline{\theta}) dx \right)}{F(\theta_2^*) \tilde{f}_\theta(\underline{\theta}|\theta_2^*)} exp(F(\theta_2^*) f_\theta(\underline{\theta}|\theta_2^*) z) \tag{A.25}$$

for one constant $\mu \in \mathbb{R}$ which depends on the initial condition $\theta_0$. Changing variables, we set $y = \mu exp(F(\theta_2^*) \tilde{f}_\theta(\underline{\theta}|\theta_2^*) z)$ so that $F(\theta_2^*) \tilde{f}_\theta(\underline{\theta}|\theta_2^*) dz = \frac{dy}{y}$. Slightly abusing notations, we get:

$$\theta_1(y) = \underline{\theta} + y, \text{ and } \theta_2(y) = \theta_2^* - \frac{f(\underline{\theta}) \left( \int_{\underline{\theta}}^{\theta_2^*} \tilde{f}_\theta(x|\underline{\theta}) dx \right)}{F(\theta_2^*) \tilde{f}_\theta(\underline{\theta}|\theta_2^*)} y \tag{A.26}$$

and (A.22) becomes:

$$\dot{Q}_i(y) = \frac{(-1)^i \tilde{f}(\theta(y))}{yF(\theta_2^*)\tilde{f}_\theta(\underline{\theta}|\theta_2^*)\lambda}\times$$

$$\left(\left(1 + \frac{F(\theta_{-i}(y))}{f(\theta_{-i}(y))}\frac{\tilde{f}_{\theta_{-i}}(\theta_i(y)|\theta_{-i}(y))}{\tilde{f}(\theta_i(y)|\theta_{-i}(y))}\right)\left(\frac{\partial \tilde{S}}{\partial q_{-i}}(Q(y))) - \theta_{-i}(y)\right) - \frac{F(\theta_{-i}(y))}{f(\theta_{-i}(y))}\right) \quad \text{(A.27)}$$

with the initial data $Q(0) = q^{SB}(\underline{\theta}, \theta_2^*)$ we obtain thereby a solution $Q(y, \theta_2^*)$. These ordinary differential equations (A.27) have singularities at $y = 0$ since the numerator and denominator on the right-hand side of (A.27) are both equal to zero at that point.

However, using Lhospital's rule, the system (A.27) gives us the derivatives at 0, namely $(\dot{Q}_1(0, \theta_2^*), \dot{Q}_2(0, \theta_2^*))$, of that solutions as the solutions to (A.28)-(A.29) below:

$$\dot{Q}_1(0, \theta_2^*) = -\frac{\tilde{f}(\underline{\theta}, \theta_2^*)}{F(\theta_2^*)\tilde{f}_\theta(\underline{\theta}|\theta_2^*)\lambda}\left(\left(1 + \frac{F(\theta_2^*)}{f(\theta_2^*)}\frac{\tilde{f}_{\theta_2}(\underline{\theta}|\theta_2^*)}{\tilde{f}(\underline{\theta}|\theta_2^*)}\right)\left(\lambda \dot{Q}_1(0, \theta_2^*) + S_{22}\dot{Q}_2(0, \theta_2^*) - \dot{\theta}_2(0)\right)\right.$$

$$\left. -\dot{\theta}_2(0)\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right)|_{\theta_2^*} + \frac{d}{dy}\left(\frac{F(\theta_2(y))}{f(\theta_2(y))}\frac{\tilde{f}_{\theta_2}(\theta_1(y)|\theta_2(y))}{\tilde{f}(\theta_1(y)|\theta_2(y))}\right)|_{y=0}(\varphi(\underline{\theta}, \theta_2^*) - \theta_2^*)\right) \quad \text{(A.28)}$$

$$\dot{Q}_2(0, \theta_2^*) = \frac{\tilde{f}(\underline{\theta}, \theta_2^*)}{F(\theta_2^*)\tilde{f}_\theta(\underline{\theta}|\theta_2^*)\lambda}\left(S_{11}\dot{Q}_1(0, \theta_2^*) + \lambda \dot{Q}_2(0, \theta_2^*) - 2\dot{\theta}_1(0)\right) \quad \text{(A.29)}$$

where $S_{11} = \frac{\partial^2 \tilde{S}}{\partial q_1^2}(Q(0, \theta_2^*))$, $S_{22} = \frac{\partial^2 \tilde{S}}{\partial q_2^2}(Q(0, \theta_2^*))$.

This system admits a unique solution in $(\dot{Q}_1(0, \theta_2^*), \dot{Q}_2(0, \theta_2^*))$, which proves local existence, since

$$\begin{vmatrix} -1-\gamma & -\gamma\frac{S_{22}}{\lambda} \\ \epsilon\frac{S_{11}}{\lambda} & \epsilon - 1 \end{vmatrix} = 2 + \gamma\epsilon\left(\frac{S_{11}S_{22}}{\lambda^2} - 1\right) \neq 0$$

where $\gamma = 1 + \frac{\tilde{f}(\underline{\theta}, \theta_2^*)}{F(\theta_2^*)\tilde{f}_\theta(\underline{\theta}|\theta_2^*)} < 0$ when the correlation is small enough and $\epsilon = \gamma - 1 = \frac{\tilde{f}(\underline{\theta}, \theta_2^*)}{F(\theta_2^*)\tilde{f}_\theta(\underline{\theta}|\theta_2^*)} < 0$. This defines the derivative and the local behavior of at least a solution $(Q_1(y, \theta_2^*), Q_2(y, \theta_2^*))$ as $(Q_1(y, \theta_2^*) = Q_1(0, \theta_2^*) + \dot{Q}_1(0, \theta_2^*)y, Q_2(y, \theta_2^*) = Q_2(0, \theta_2^*) + \dot{Q}_1(0, \theta_2^*)y)$.

Now, solving the system $\theta = \theta(y, \theta_2^*)$ for $y$ small enough yields $(y, \theta_2^*) = (Y(\theta), \Theta_2^*(\theta))$. Using (A.26 ), we get:

$$y = \theta_1 - \underline{\theta} \text{ and } \theta_2 - \theta_2^* = \beta(\theta_2^*)(\theta_1 - \underline{\theta})$$

where $\beta(\theta_2^*) = -\frac{f(\underline{\theta})\left(\int_{\underline{\theta}}^{\theta_2^*} \tilde{f}_\theta(x|\underline{\theta})dx\right)}{F(\theta_2^*)\tilde{f}_\theta(\underline{\theta}|\theta_2^*)}$. This system can be uniquely inverted for $\theta_1$ close enough to $\underline{\theta}$ since the derivative w.r.t. $\theta_2^*$ of the right-hand side of the second equation is non-zero for $\theta_2$ close enough to $\theta_2^*$. Finally, locally around $(\underline{\theta}, \theta_2^*)$, we get $q^{SB}(\theta) = Q(Y(\theta), \Theta_2^*(\theta))$ for a solution to (26) such that $Q(0, \theta_2^*) = q^{SB}(\underline{\theta}, \theta_2^*)$.

Finally, tedious but straightforward computations show that the derivatives $\frac{\partial q_i}{\partial \theta_1}(\theta)$ and $\frac{\partial q_i}{\partial \theta_2}(\theta)$ for $i = 1, 2$ satisfy Assumptions 3 and 4 provided $\lambda$ is small enough and $\frac{\partial \varphi}{\partial \theta_i}(\theta_i, \theta_{-i}) \geq |\frac{\partial \varphi}{\partial \theta_{-i}}(\theta_i, \theta_{-i})|$. ■

**Proof of Corollary 1.** Denote $q^{FB}(\underline{\theta}, \underline{\theta}) = \frac{\mu - \underline{\theta}}{1-\lambda}$ the first-best output at $(\underline{\theta}, \underline{\theta})$, and the new variables $y_i(x_i, x_{-i}) = q_i^{SB}(\theta_i, \theta_{-i}) - q^{FB}(\underline{\theta}, \underline{\theta})$ and $x_i = \theta_i - \underline{\theta}$. We are looking for an analytic solution to (26) in the neighborhood of $(\underline{\theta}, \underline{\theta})$. Up to terms of order more than 2, (26) (for $i = 1$) can be rewritten in the neighborhood of $(\underline{\theta}, \underline{\theta})$ as:

$$(1+l(\theta_1-\underline{\theta}))(\mu-q_1^{SB}+\lambda q_2^{SB}-\theta_1)-(\theta_1-\underline{\theta})-m(\theta_1-\underline{\theta})^2 = l\lambda(\theta_1-\underline{\theta})(\theta_2-\underline{\theta})\left(\frac{\partial q_2^{SB}}{\partial \theta_1} - \frac{\partial q_2^{SB}}{\partial \theta_2}\right)$$

which yields with the new variables

$$(1 + lx_1)(-y_1 + \lambda y_2 - x_1) - x_1 - mx_1^2 = l\lambda x_1 x_2 \left(\frac{\partial y_2}{\partial x_1} - \frac{\partial y_2}{\partial x_2}\right) \quad \text{(A.30)}$$

and a similar equation is obtained by permuting indices.

We look for a symmetric analytic solution of the form:

$$y_i(x_i, x_{-i}) = a_1 x_i + a_2 x_{-i} + b_1 x_i^2 + b_2 x_{-i}^2 + b_3 x_i x_{-i} + o_i(||x||^2),$$

where $o_i(||x||^2)$ $(i = 1, 2)$ is of order more than 2. Inserting this expression into (A.30) and identifying the coefficients yields:

$$a_1 = -\frac{2}{1-\lambda^2}, \quad a_2 = -\frac{2\lambda}{1-\lambda^2}, \quad b_1 = \frac{l-m}{1-\lambda^2}, \quad b_2 = \frac{\lambda(l-m)}{1-\lambda^2}, \quad b_3 = -\frac{2l\lambda}{1-\lambda^2}.$$

This yields the expression of the solution in the text. Assumption 3 is easily checked. Note that $|a_1| > |a_2|$ so that Assumption 4 holds. ■

**Proof of Proposition 4.** Let us write the principal's objective function in this team context as:

$$W(q(\cdot), t(\cdot)) = (\nu^2 + \alpha)(S(q(\underline{\theta}, \underline{\theta})) - 2t(\underline{\theta}, \underline{\theta}))$$
$$+2(\nu(1-\nu) - \alpha)(S(q(\underline{\theta}, \bar{\theta})) - t(\bar{\theta}, \underline{\theta})) - t(\underline{\theta}, \bar{\theta})) + ((1-\nu)^2 + \alpha)(S(q(\bar{\theta}, \bar{\theta})) - 2t(\bar{\theta}, \bar{\theta})).$$

For $\alpha$ small enough, intuition suggests that the relevant Bayesian incentive constraint is that of an efficient agent and the relevant participation constraint that of an inefficient one. Those constraints can be written respectively as:

$$U(\underline{\theta}) = \left(\nu + \frac{\alpha}{\nu}\right)(t(\underline{\theta}, \underline{\theta}) - \underline{\theta} q(\underline{\theta}, \underline{\theta})) + \left(1 - \nu - \frac{\alpha}{1-\nu}\right)(t(\underline{\theta}, \bar{\theta}) - \underline{\theta} q(\underline{\theta}, \bar{\theta}))$$

$$\geq \left(\nu + \frac{\alpha}{\nu}\right)(t(\bar{\theta}, \underline{\theta}) - \underline{\theta} q(\bar{\theta}, \underline{\theta})) + \left(1 - \nu - \frac{\alpha}{\nu}\right)(t(\bar{\theta}, \bar{\theta}) - \underline{\theta} q(\bar{\theta}, \bar{\theta})). \quad \text{(A.31)}$$

$$U(\bar{\theta}) = \left(\nu - \frac{\alpha}{1-\nu}\right)(t(\bar{\theta}, \underline{\theta}) - \bar{\theta} q(\bar{\theta}, \underline{\theta})) + \left(1 - \nu + \frac{\alpha}{1-\nu}\right)(t(\bar{\theta}, \bar{\theta}) - \bar{\theta} q(\bar{\theta}, \bar{\theta})) \geq 0. \quad \text{(A.32)}$$

Neglecting the Bayesian incentive constraint of an inefficient agent and the participation constraint of an inefficient one, the principal's problem so relaxed becomes thus:

$$(\mathcal{P}): \max_{\{q(\cdot),t(\cdot)\}} W(q(\cdot),t(\cdot)) \text{ subject to (29), (30), (A.31) and (A.32).}$$

These constraints define a convex set with non-empty interior so that constraint qualification holds. Denoting respectively by $\beta$, $\gamma$, $\lambda$ and $\mu$ the non-negative multipliers of those constraints, forming the Lagrangean and optimizing with respect to transfers yields the following Karush-Kuhn-Tucker conditions:

$$-2(\nu^2+\alpha)+\frac{\lambda}{\nu}(\nu^2+\alpha)-2\beta=0$$

$$-2(\nu(1-\nu)-\alpha)+\frac{\lambda}{\nu}(\nu(1-\nu)-\alpha)+2\beta=0,$$

$$-2(\nu(1-\nu)-\alpha)-\frac{\lambda}{\nu}(\nu^2+\alpha)+2\gamma+\frac{\mu}{1-\nu}(\nu(1-\nu)-\alpha)=0,$$

$$-2((1-\nu)^2+\alpha)-\frac{\lambda}{\nu}(\nu(1-\nu)-\alpha)-2\gamma+\frac{\mu}{1-\nu}((1-\nu)^2+\alpha)=0.$$

Solving this system yields,

$$\beta=0, \gamma=\alpha>0, \lambda=2\nu>0, \mu=2. \tag{A.33}$$

From which we deduce that (29) is slack and (30), (A.31) and (A.32) are all binding at the optimum. Using (30) and (A.32) binding, yields

$$t(\bar{\theta},\bar{\theta})-\bar{\theta}q(\bar{\theta},\bar{\theta})=\frac{\nu(1-\nu)-\alpha}{2(1-\nu)}\left[S(q)-2\bar{\theta}q\right]_{q(\underline{\theta},\bar{\theta})}^{q(\bar{\theta},\bar{\theta})}, t(\bar{\theta},\underline{\theta})-\bar{\theta}q(\underline{\theta},\bar{\theta})=-\frac{(1-\nu)^2+\alpha}{2(1-\nu)}\left[S(q)-2\bar{\theta}q\right]_{q(\underline{\theta},\bar{\theta})}^{q(\bar{\theta},\bar{\theta})}$$

and the expressions of the agent's rents as

$$U(\underline{\theta})=\Delta\theta\left(\left(\nu+\frac{\alpha}{\nu}\right)q(\underline{\theta},\bar{\theta})+\left(1-\nu-\frac{\alpha}{\nu}\right)q(\bar{\theta},\bar{\theta})\right)+\frac{\alpha}{\nu(1-\nu)}\left[S(q)-2\bar{\theta}q\right]_{q(\bar{\theta},\bar{\theta})}^{q(\underline{\theta},\bar{\theta})}$$

$$>U(\bar{\theta})=0.$$

Inserting those values into $W(q(\cdot),t(\cdot))$ and optimizing gives the expression of the optimal outputs in the proposition.

The expression above for an efficient agent's rent already shows that, for $\alpha$ not too large, an efficient agent's rent is strictly positive given that outputs are so that the latter's participation constraint is slack. We have:

$$U^{SB}(\bar{\theta})=0<U^{SB}(\underline{\theta})=\Delta\theta\left(\left(\nu+\frac{\alpha}{\nu}\right)q^{SB}(\underline{\theta},\bar{\theta})+\left(1-\nu-\frac{\alpha}{\nu}\right)q^{SB}(\bar{\theta},\bar{\theta})\right)$$

$$-\frac{\alpha}{\nu(1-\nu)}\left(S(q^{SB}(\bar{\theta},\bar{\theta}))-2\bar{\theta}q^{SB}(\bar{\theta},\bar{\theta})-(S(q^{SB}(\underline{\theta},\bar{\theta}))-2\bar{\theta}q^{SB}(\underline{\theta},\bar{\theta}))\right).$$

We now check that an inefficient agent's incentive constraint is slack, at least for $\alpha$ small enough. This amounts to verify

$$0 > (\nu(1-\nu)-\alpha)(t^{SB}(\underline{\theta},\underline{\theta}) - \bar{\theta} q^{SB}(\underline{\theta},\underline{\theta})) + ((1-\nu)^2+\alpha)(t^{SB}(\underline{\theta},\bar{\theta}) - \bar{\theta} q^{SB}(\underline{\theta},\bar{\theta})).$$

However this inequality holds strictly for $\alpha = 0$ since then

$$\Delta(\nu q_0^{SB}(\underline{\theta},\underline{\theta}) + (1-\nu) q_0^{SB}(\underline{\theta},\bar{\theta})) > \Delta(\nu q_0^{SB}(\bar{\theta},\underline{\theta}) + (1-\nu) q_0^{SB}(\bar{\theta},\bar{\theta}))$$

where $S'(q_0^{SB}(\underline{\theta},\underline{\theta})) = 2\underline{\theta}$, $S'(q_0^{SB}(\underline{\theta},\bar{\theta})) = \underline{\theta} + \bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta$, and $S'(q_0^{SB}(\bar{\theta},\bar{\theta})) = 2\bar{\theta} + \frac{2\nu}{1-\nu}\Delta\theta$ and, by continuity, it holds also for $\alpha$ small enough. ∎

**Proof of Proposition 5.** Let us write the principal's objective function in this multiunit auction context as:

$$W(q(\cdot),t(\cdot)) = (\nu^2+\alpha)(S(2q(\underline{\theta},\underline{\theta})) - 2t(\underline{\theta},\underline{\theta}))$$

$$+2(\nu(1-\nu)-\alpha)(S(q(\underline{\theta},\bar{\theta})+q(\bar{\theta},\underline{\theta}))-t(\underline{\theta},\underline{\theta}))-t(\underline{\theta},\bar{\theta}))+((1-\nu)^2+\alpha)(S(2q(\bar{\theta},\bar{\theta}))-2t(\bar{\theta},\bar{\theta})).$$

For $\alpha$ small enough, intuition suggests again that the relevant Bayesian incentive constraint is that of an efficient agent and the relevant participation constraint that of an inefficient one which can be written still as (A.31) and (A.32). We first neglect the Bayesian incentive constraint of an inefficient agent and the participation constraint of an inefficient one. Non-manipulability imposes (31) in state $(\bar{\theta},\bar{\theta})$. In state $(\underline{\theta},\underline{\theta})$, it requires

$$S(2q(\underline{\theta},\underline{\theta})) - 2t(\underline{\theta},\underline{\theta}) \geq S(q(\underline{\theta},\underline{\theta}) + q(\underline{\theta},\bar{\theta})) - t(\underline{\theta},\underline{\theta}) - t(\underline{\theta},\bar{\theta}). \quad \text{(A.34)}$$

The global non-manipulability constraint that prevents simultaneous deviations towards both agents can be written as

$$S(2q(\underline{\theta},\underline{\theta})) - 2t(\underline{\theta},\underline{\theta}) \geq S(2q(\underline{\theta},\bar{\theta})) - 2t(\underline{\theta},\bar{\theta}). \quad \text{(A.35)}$$

Due to the strict concavity of $S(\cdot)$, (A.34) is more stringent than (A.35).

In state $(\underline{\theta},\bar{\theta})$, non-manipulability imposes also two local and one global constraints that can be written in compact form as

$$S(q(\underline{\theta},\bar{\theta}) + q(\bar{\theta},\underline{\theta})) - t(\underline{\theta},\bar{\theta}) - t(\bar{\theta},\underline{\theta}) \geq$$

$$\max\{S(q(\underline{\theta},\bar{\theta}) + q(\bar{\theta},\bar{\theta})) - t(\underline{\theta},\bar{\theta}) - t(\bar{\theta},\bar{\theta}); S(q(\underline{\theta},\underline{\theta}) + q(\bar{\theta},\underline{\theta})) - t(\underline{\theta},\underline{\theta}) - t(\bar{\theta},\underline{\theta})$$

$$; S(q(\underline{\theta},\underline{\theta}) + q(\bar{\theta},\bar{\theta})) - t(\underline{\theta},\underline{\theta}) - t(\bar{\theta},\bar{\theta})\} \quad \text{(A.36)}$$

We first neglect constraints (A.34) and (A.36) and check ex post that these constraints can be satisfied at no additional cost for the principal. Consider thus the principal's so relaxed problem

$$(\mathcal{P}): \max_{\{q(\cdot),t(\cdot)\}} W(q(\cdot),t(\cdot)) \text{ subject to (31), (A.31) and (A.32).}$$

At the optimum of $(\mathcal{P})$, (31), (A.31) and (A.32) are all binding. In particular, (31) binding implies

$$t(\bar{\theta},\underline{\theta}) = t(\bar{\theta},\bar{\theta}) + S(q(\bar{\theta},\underline{\theta}) + q(\bar{\theta},\bar{\theta})) - S(2q(\bar{\theta},\bar{\theta})).$$

Inserting into the binding participation constraint (A.32) gives the values of the transfers

$$t(\bar{\theta},\bar{\theta}) - \bar{\theta}q(\bar{\theta},\bar{\theta}) = -\left(\nu - \frac{\alpha}{1-\nu}\right)\left[S(q(\bar{\theta},\bar{\theta})+q) - \bar{\theta}q\right]_{q(\bar{\theta},\bar{\theta})}^{q(\bar{\theta},\underline{\theta})},$$

$$t(\bar{\theta},\underline{\theta}) - \bar{\theta}q(\bar{\theta},\underline{\theta}) = -\left(\nu - \frac{\alpha}{1-\nu}\right)\left[S(q(\bar{\theta},\bar{\theta})+q) - \bar{\theta}q\right]_{q(\bar{\theta},\bar{\theta})}^{q(\bar{\theta},\underline{\theta})}.$$

We can then use these expressions to obtain the information rents left to one agent as

$$U(\underline{\theta}) = \Delta\theta\left(\left(\nu + \frac{\alpha}{\nu}\right)q(\underline{\theta},\bar{\theta}) + \left(1 - \nu - \frac{\alpha}{\nu}\right)q(\bar{\theta},\bar{\theta})\right) + \frac{\alpha}{\nu(1-\nu)}\left[S(q(\bar{\theta},\bar{\theta})+q) - \bar{\theta}q\right]_{q(\bar{\theta},\bar{\theta})}^{q(\bar{\theta},\underline{\theta})}$$

$$> U(\bar{\theta}) = 0.$$

Inserting these values into $W(q(\cdot),t(\cdot))$ and optimizing for $\alpha$ small gives the expression of the optimal quantities in the proposition.

An argument similar to that in the Proof of Proposition 4 shows that, for $\alpha$ small enough, an inefficient agent's incentive constraint is slack.

Let us check that the non-manipulability constraints (A.34) and (A.36) both hold. In the optimal mechanism, identified so far, there is one degree of freedom. The expected transfer obtained by an efficient agent is fixed but the choice of $t(\underline{\theta},\underline{\theta})$ and $t(\underline{\theta},\bar{\theta})$ is possible in a wide set (because of correlation it is nevertheless constrained by the incentive constraint of an inefficient agent). Constraint (A.34) can be expressed as

$$t^{SB}(\underline{\theta},\bar{\theta}) - t^{SB}(\underline{\theta},\underline{\theta}) \geq S(q^{SB}(\underline{\theta},\bar{\theta}) + q^{SB}(\underline{\theta},\underline{\theta})) - S(2q^{SB}(\underline{\theta},\underline{\theta})) \qquad (A.37)$$

When (31) is binding, the first inequality in constraint (A.36) amounts to

$$S(q^{SB}(\underline{\theta},\bar{\theta})) - S(q^{SB}(\bar{\theta},\bar{\theta})) \geq S(q^{SB}(\underline{\theta},\bar{\theta}) + q^{SB}(\bar{\theta},\bar{\theta})) - S(2q^{SB}(\bar{\theta},\bar{\theta})),$$

which holds since $q^{SB}(\underline{\theta},\bar{\theta}) \geq q^{SB}(\bar{\theta},\bar{\theta})$, and $S(\cdot)$ is concave. Still when (31) is binding, the second inequality in (A.36) is more stringent than the third one as long as $q^{SB}(\underline{\theta},\underline{\theta}) \geq q^{SB}(\bar{\theta},\bar{\theta})$ by the same concavity argument as above. This second inequality can be written as

$$t^{SB}(\underline{\theta},\bar{\theta}) - t^{SB}(\underline{\theta},\underline{\theta}) \leq S(q^{SB}(\underline{\theta},\bar{\theta})) - S(q^{SB}(\underline{\theta},\underline{\theta})). \qquad (A.38)$$

It is possible to satisfy all the non-manipulability constraints at no additional cost for the principal if one can find $t^{SB}(\underline{\theta},\underline{\theta})$ and $t^{SB}(\underline{\theta},\bar{\theta})$ such that (A.37) and (A.38) are both satisfied. But observe that, when (A.38) is taken as an equality, (A.37) holds because

$S(\cdot)$ is concave. Finally, for $\alpha$ sufficiently small, such transfers do not violate the incentive constraint of an inefficient agent. ■

**Proof of Proposition 6.** The proof is straightforwardly adapted from that of Proposition 1 by replacing Bayesian incentive constraint with dominant strategy incentive compatibility ones. ■

**Proof of Proposition 7.** Denoting $u_i(\theta_i, \theta_{-i}) = t_i(\theta_i, \theta_{-i}) - \theta_i q_i(\theta_i, \theta_{-i})$ the ex post rent of an agent $A_i$ with type $\theta_i$ when $A_{-i}$ reports $\theta_{-i}$, dominant strategy incentive compatibility implies that $q_i(\theta_i, \theta_{-i})$ is weakly decreasing in $\theta_i$, for all $\theta_{-i}$, and

$$u_i(\theta_i, \theta_{-i}) = u_i(\bar{\theta}, \theta_{-i}) + \int_{\theta_i}^{\bar{\theta}} q_i(u, \theta_{-i}) du. \tag{A.39}$$

Imposing ex post participation constraints which hold irrespectively of the agents' beliefs on each other types, we must have:

$$u_i(\theta_i, \theta_{-i}) \geq 0, \quad \forall (\theta_i, \theta_{-i}) \in \Theta^2.$$

Consider the simple bilateral contracts $\{t_i^{BM}(\hat{\theta}_i), q^{BM}(\hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ where $t_i^{BM}(\hat{\theta}_i) = \hat{\theta}_i q_i^{BM}(\hat{\theta}_i) + \int_{\hat{\theta}_i}^{\bar{\theta}} q_i^{BM}(u) du$. These simple bilateral contracts are such that only the inefficient agents' participation constraints are binding, namely $u_i(\bar{\theta}, \theta_{-i}) = 0$ for all $\theta_{-i} \in \Theta$. These contracts satisfy also incentive compatibility. Moreover, they implement the optimal Baron-Myerson quantities. They thus maximize the principal's expected payoff within the set of simple bilateral contracts.

We must check that more complex bilateral mechanisms cannot achieve a greater payoff. Notice that non-manipulability and dominant strategy incentive compatibility imply that there exists functions $H_i(\cdot)$ $(i = 1, 2)$ such that

$$H_i(\theta_i) = S(q_i(\theta_i, \theta_{-i})) - \theta_i q_i(\theta_i, \theta_{-i}) - u_i(\bar{\theta}, \theta_{-i}) - \int_{\theta_i}^{\bar{\theta}} q_i(x, \theta_{-i}) dx \quad \forall \theta_{-i}. \tag{A.40}$$

The principal's problem can thus be written

$$(\mathcal{P}) : \max_{\{q(\cdot), H(\cdot)\}} \sum_{i=1}^{2} E_{\theta_i}(H_i(\theta_i))$$

$$\text{subject to (A.40), } q_i(., \theta_{-i}) \text{ non-increasing and } u_i(\bar{\theta}, \theta_{-i}) \geq 0 \ \forall \theta_{-i} \in \Theta.$$

This last constraint is obviously binding at the optimum.

For any acceptable non-manipulable and dominant strategy mechanism which implements a quantity schedule $q_i(\theta_i, \theta_{-i})$, (A.40) implies that the principal can get the same payoff with a non-manipulable mechanism that implements the schedule $q_i(\theta_i) = q_i(\theta_i, \bar{\theta})$.

The optimal such output is then $q^{BM}(\theta_i)$. Moreover, this outcome can be implemented with a set of simple bilateral contracts with corresponding transfers $t_i(\theta_i) = t_i(\theta_i, \bar{\theta})$ and outputs $q_i(\theta_i) = q_i(\theta_i, \bar{\theta})$ which depend only on $A_i$'s type. ■

**Proof of Proposition 8.** A non-manipulable and collusion-proof mechanism that is Bayesian incentive compatible must satisfy (34) and (16). This yields:

$$E_{\theta_i}\left((S'(q_i(\theta_i, \theta_{-i})) - \theta_i)\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, \theta_{-i})|\theta_{-i}\right) = 0. \tag{A.41}$$

Clearly, the second-best $q^{SB}(\theta)$ does not satisfy this condition. An output schedule such that $q_i(\theta_i, \theta_{-i}) \leq q^{FB}(\theta_i)$ with equality only at $\theta_i = \underline{\theta}$ and $\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, \theta_{-i}) \geq 0$ can only satisfy (A.41) when

$$\frac{\partial q_i}{\partial \theta_{-i}}(\theta_i, \theta_{-i}) = 0.$$

Non-manipulable and collusion-proof mechanisms are thus necessarily simple bilateral mechanisms that implement the Baron-Myerson outcome $q^{BM}(\theta_i)$. ■

**Proof of Proposition 9.** It should be clear that offering the same contracts as in the case of public offers is an optimal equilibrium strategy for the principal within the class of mechanisms where he is a priori restricted to report to $A_i$ the message he receives from agent $A_{-i}$. The proof follows the same steps as the proof of Proposition 2 and is omitted. The only new question to investigate is whether the principal could deviate to a larger class of mechanisms to communicate with $A_i$ possibly the endogenous information he has on whatever private offers he makes to agent $A_{-i}$. Denote thus by $\mathcal{P}_i$ any arbitrary compact message space available to the principal to communicate with $A_i$ on top of the type space available to report on $A_{-i}$'s type, and by $\{\tilde{t}_i(\theta_i, \theta_{-i}, p_i), \tilde{q}_i(\theta_i, \theta_{-i}, p_i)\}_{\{\hat{\theta}_i \in \Theta, \hat{\theta}_{-i} \in \Theta, p_i \in \mathcal{P}_i\}}$ a menu of so extended direct mechanisms which is assumed to be lower-semi continuous in $p_i$. Finally, denote by $p(\theta) = (p_1(\theta), p_2(\theta))$ an array of best-responses for the principal, a priori this is a correspondence but slightly abusing notations we will denote similarly any selection within such correspondence. Optimality of the principal's behavior at the last stage of the game requires:

$$(\theta, p(\theta)) \in \arg \max_{\{\hat{\theta} \in \Theta^2, p \in \prod_{i=1}^2 \mathcal{P}_i\}} \sum_{i=1}^{2} S(\tilde{q}_i(\theta_i, \hat{\theta}_{-i}, p_i)) - \tilde{t}_i(\theta_i, \hat{\theta}_{-i}, p_i) \tag{A.42}$$

where the maximum above is achieved by compactness of $\mathcal{P}_i$ and lower-semi continuity in $p_i$. Let define the new mechanism $(t_i^S(\theta), q_i^S(\theta)) = (\tilde{t}_i(\theta, p_i(\theta)), \tilde{q}_i(\theta, p_i(\theta)))$. Such mechanism does not use "extended" reports from the principal. The optimality condition (A.42) can be rewritten as:

$$\theta \in \arg \max_{\hat{\theta} \in \Theta^2} \sum_{i=1}^{2} S(q_i^S(\theta_i, \hat{\theta}_{-i})) - t_i^S(\theta_i, \hat{\theta}_{-i}). \tag{A.43}$$

so that the new mechanism $(t_i^S(\theta), q_i^S(\theta))$ is non-manipulable. This shows that there is no point in enlarging the set of mechanisms available to the principal. ■

**Proof of Proposition 10.** The proof is in several steps.

**$A_1$'s sub-mechanism and reporting strategy.** In any equilibrium of the sequential bilateral contracting game, the principal offers to $A_1$ a sell-out contract of the form

$$t_1(\theta_1, \theta_2) = S(q_1(\theta_1, \theta_2)) - H_1(\theta_1) \tag{A.44}$$

or alternatively the nonlinear price

$$T_1(q_1, \theta_1) = S(q_1) - H_1(\theta_1).$$

Indeed, exactly as in the simultaneous timing, such contract prevents manipulations by $P$ of $A_2$'s report towards $A_1$ once $P$ will have learned this report.

A priori, in this model with sequential contracting, we might allow for possible (non-truthful) mixed strategy for $A_1$, with different types maybe sending the same message $\hat{\theta}_1$. Let denote by $d\sigma(\hat{\theta}_1|\theta_1)$ the conditional density corresponding to such mixed-strategy. (We will see below that a simple argument justifies that such strategy has indeed to be truthful.) When $A_1$ plays a non-truthful strategy in equilibrium and reports $\hat{\theta}_1$, it is possible to use Bayes rule to compute $d\sigma(\theta_1|\hat{\theta}_1)$ and the conditional distribution of $\theta_2$ conditional on having learned $\hat{\theta}_1$ which we denote (slightly abusing notations) $\tilde{f}(\theta_2|\hat{\theta}_1)$ with cumulative distribution $\tilde{F}(\theta_2|\hat{\theta}_1)$.

**$A_2$'s optimal sub-mechanism.** Let us turn now to the bilateral contract between $P$ and $A_2$. Contracting with $A_2$ takes place after that $P$ has privately learned the report $\hat{\theta}_1$ from $A_1$. Following this report $\hat{\theta}_1$, the principal can use any simple sub-mechanism of the form $\{t_2(\hat{\theta}_2), q_2(\hat{\theta}_2)\}_{\hat{\theta}_2 \in \Theta}$ in his relationship with $A_2$. Of course, such mechanism must satisfy the following constraints

$$t_2(\theta_2) - \theta_2 q_2(\theta_2) \geq t_2(\tilde{\theta}_2) - \theta_2 q_2(\tilde{\theta}_2), \quad \forall(\theta_2, \tilde{\theta}_2) \tag{A.45}$$

$$t_2(\theta_2) - \theta_2 q_2(\theta_2) \geq 0 \quad \forall \theta_2. \tag{A.46}$$

Those constraints are respectively the dominant strategy incentive and ex post participation constraints for $A_2$. Note that the set of such incentive feasible allocations does not depend directly neither on the principal's information $\hat{\theta}_1$ nor on $A_1$'s type $\theta_1$ (or more exactly, $A_2$'s expectations over that type given what he learns from observing $P$'s offer). In particular, the set of such incentive feasible allocations is the same as if $\hat{\theta}_1$ was known also by $A_2$.

The principal's problem when dealing with $A_2$ consists in maximizing his expected profit conditionally on having received report $\hat{\theta}_1$ from $A_1$ subject to (A.45) and (A.46).

A relaxed version of this problem is obtained when the monotonicity condition ($q_2(\theta_2|\hat{\theta}_1)$ non-increasing) implied by (A.45) is omitted. For this relaxed problem, the participation constraint (A.46) binds at $\theta_2 = \bar{\theta}$. Let thus denote by $\{t_2^D(\hat{\theta}_2|\hat{\theta}_1), q_2^D(\hat{\theta}_2|\hat{\theta}_1)\}_{\hat{\theta}_2\in\Theta}$ the solution to this relaxed problem. This mechanism is computed as if $\hat{\theta}_1$ was known also by $A_2$. Such optimal mechanism implements the following output schedule

$$S'(q_2^D(\theta_2|\hat{\theta}_1)) = \theta_2 + \frac{\tilde{F}(\theta_2|\hat{\theta}_1)}{\tilde{f}(\theta_2|\hat{\theta}_1)}, \quad \forall(\hat{\theta}_1, \theta_2) \tag{A.47}$$

with the dominant strategy transfers

$$t_2^D(\theta_2|\hat{\theta}_1) = \theta_2 q_2^D(\theta_2|\hat{\theta}_1) + \int_{\theta_2}^{\bar{\theta}} q_2^D(u|\hat{\theta}_1)du. \tag{A.48}$$

Note that the output $q_2^D(\theta_2|\hat{\theta}_1)$ given in (A.47) may not necessarily be non-increasing. Moreover, only the ex post participation constraints of a $\bar{\theta}$-agent $A_2$ is binding following $P$'s learning of $\hat{\theta}_1$ with the transfers given in (A.48).

Clearly, this is a best strategy for $P$ to solve the relaxed problem by offering such sub-mechanism $\{t_2^D(\hat{\theta}_2|\hat{\theta}_1), q_2^D(\hat{\theta}_2|\hat{\theta}_1)\}_{\hat{\theta}_2\in\Theta}$ for any $\hat{\theta}_1$ he may have learned even if, actually, $\hat{\theta}_1$ is not known by $A_2$.[41] Moreover, the out-of equilibrium beliefs that $A_2$ may held following any deviation away from this offer do not affect his play and are thus arbitrary.

Moreover, by definition of the optimal sub-mechanism $\{t_2^D(\hat{\theta}_2|\hat{\theta}_1), q_2^D(\hat{\theta}_2|\hat{\theta}_1)\}_{\hat{\theta}_2\in\Theta}$, it must be that

$$E_{\theta_2}(S(q_2^D(\theta_2|\hat{\theta}_1)) - t_2^D(\theta_2|\hat{\theta}_1)|\hat{\theta}_1) \geq E_{\theta_2}(S(q_2^D(\theta_2|\tilde{\theta}_1)) - t_2^D(\theta_2|\tilde{\theta}_1)|\hat{\theta}_1) \quad \forall(\hat{\theta}_1, \tilde{\theta}_1) \tag{A.49}$$

where $\tilde{\theta}_1$ is in the support of the mixed reporting strategy of at least one type of $A_1$. Compared to the simultaneous timing, the sequential contracting game replaces the set of ex post non-manipulability constraints of $P$ when dealing with $A_2$ by an interim non-manipulability condition which is readily verified by definition of the optimality of $\{t_2^D(\hat{\theta}_2|\hat{\theta}_1), q_2^D(\hat{\theta}_2|\hat{\theta}_1)\}_{\hat{\theta}_2\in\Theta}$. However, doing so is at the cost of replacing the interim incentive and participation constraints of $A_2$ that prevail under simultaneous contracting by dominant strategy incentive constraints and an ex post participation ones.

$A_1$**'s reporting strategy.** Let us turn now to the first-period reporting strategy by $A_1$ that $P$ would like to induce. Note first that $P$'s expected payoff in his relationship with $A_2$ can be written as

$$E_{\theta_1}\left(\int_{\Theta} E_{\theta_2}(S(q_2^D(\theta_2|\hat{\theta}_1)) - t_2^D(\theta_2|\hat{\theta}_1)|\hat{\theta}_1)d\sigma(\hat{\theta}_1|\theta_1)\right).$$

[41] This standard argument is due to the private values context ($\hat{\theta}_1$ does not affect directly $A_2$'s utility). See Martimort and Moreira (2009) for a similar argument.

Now observe that, by definition of $\{t_2^D(\hat{\theta}_2|\theta_1), q_2^D(\hat{\theta}_2|\theta_1)\}_{\hat{\theta}_2\in\Theta}$, we have for any mixed-strategy $d\sigma(\cdot|\theta_1)$ that may have been played by $A_1$

$$E_{\theta_2}(S(q_2^D(\theta_2|\theta_1)) - t_2^D(\theta_2|\theta_1)|\theta_1) \geq E_{\theta_2}\left(\int_\Theta \left(S(q_2^D(\theta_2|\hat{\theta}_1)) - t_2^D(\theta_2|\hat{\theta}_1)\right) d\sigma(\hat{\theta}_1|\theta_1)|\theta_1\right).$$

Moreover, given Assumption 5, $q_2^D(\theta_2|\theta_1)$ satisfies the monotonicity condition ($q_2^D(\theta_2|\hat{\theta}_1)$ non-increasing) and the left-hand side above is the principal's payoff with $A_2$ in the continuation equilibrium following a truthful report by $A_1$. Taking expectations yields

$$E_{\theta_1}\left(E_{\theta_2}(S(q_2^D(\theta_2|\theta_1)) - t_2^D(\theta_2|\theta_1)|\theta_1)\right) \geq E_{\theta_1}\left(E_{\theta_2}\left(\int_\Theta \left(S(q_2^D(\theta_2|\hat{\theta}_1)) - t_2^D(\theta_2|\hat{\theta}_1)\right) d\sigma(\hat{\theta}_1|\theta_1)|\theta_1\right)\right)$$

which immediately gives us that truthtelling is the best reporting strategy that $P$ may induce from $A_1$ to maximize his expected payoff with $A_2$.

On the other hand, the principal's expected payoff with $A_1$ obtained by offering any sell-out contract as in (A.44) and having any reporting strategy $d\sigma(\cdot|\theta_1)$ is $\int_\Theta H_1(\hat{\theta}_1)d\sigma(\hat{\theta}_1|\theta_1)$. This payoff can also be achieved with a sell-out contract having $H_1'(\theta_1) = \int_\Theta H_1(\hat{\theta}_1)d\sigma(\hat{\theta}_1|\theta_1)$ and an output $q_1'(\theta_1, \theta_2)$ such that $S(q_1'(\theta_1, \theta_2)) = \int_\Theta S(q_1(\hat{\theta}_1, \theta_2))d\sigma(\hat{\theta}_1|\theta_1)$ without perturbating $A_1$'s incentive and participation constraints. This shows finally that the best reporting strategy that $P$ can induce from $A_1$ is truthtelling.

**Optimality of simultaneous contracting.** Consider now the function

$$H_2(\theta_2) = E_{\theta_1}(S(q_2^D(\theta_2|\theta_1)) - t_2^D(\theta_2|\theta_1)|\theta_2),$$

and suppose that, in the game with simultaneous bilateral contracting, $P$ offers to $A_2$, the following sell-out contract

$$\tilde{t}_2(\theta_1, \theta_2) = S(q_2^D(\theta_2|\theta_1)) - H_2(\theta_2).$$

This mechanism is clearly non-manipulable in the simultaneous contracting game. By construction, it is Bayesian incentive compatible for $A_2$ and satisfies the interim participation constraints of all types of this agent. Moreover, this mechanism gives the same ex ante payoff to the principal as the optimal sequential contract he can offer to $A_2$. This allows to conclude that the principal could always obtain a higher payoff in the simultaneous bilateral contracting game than in a fully revealing equilibrium of the sequential bilateral contracting game. ■