Mechanism Design with Private Communication[1]

Vianney Dequiedt[2] and David Martimort[3]

This version: October 1, 2006

Abstract: This paper relaxes an important assumption of the mechanism design literature: That communication between the principal and each of his agents is *public*. Doing so yields two important results. First, it simplifies significantly mechanisms and institutions and shows the major role played by *sell-out contracts*, nonlinear prices and all-pay auctions. Second, it restores continuity with respect to the information structure but still maintains the useful role of correlation as a means to better extract the agent's information rents. We first prove a *Revelation Principle with private communication* that characterizes the whole set of implementable allocations which cannot be manipulated by the principal by means of simple *non-manipulability constraints.* Equipped with this tool, we investigate optimal non-manipulable mechanisms in various environments of increasing complexity (unrelated projects, auctions, team production, more general production externality). We also demonstrate a *Taxation Principle with private communication* and draw some links between our framework and the common agency literature.

Keywords: Mechanism Design, Private Communication.

JEL Classification : D82.

[1]We thank seminar participants in Marseille LAGV 2006 Conference, Paris-PSE, Paris-Institut Poincaré, Erasmus-Rotterdam, Toulouse, and also Olivier Compte, Philippe Jehiel, Frédéric Koessler, Andy Postlewaite, Thomas Palfrey, Jérôme Pouyet and Jean Tirole for helpful comments. The usual disclaimer applies.

[2]INRA-GAEL, University of Grenoble, France. e-mail: dequiedt@grenoble.inra.fr

[3]University of Toulouse, Institut Universitaire de France and EHESS, France. e-mail: martimor@cict.fr

1 Introduction

Over the last thirty years or so, the theory of mechanism design has been viewed as the most powerful tool to understand how complex organizations and institutions are shaped. By means of the Revelation Principle,[1] this theory offers a full characterization of the set of implementable allocations in contexts where information is decentralized and privately known by agents at the periphery of the organization. Once this first step of the analysis is performed and once a particular welfare criterion is specified at the outset, one can find an optimal incentive feasible allocation and look for practical mechanisms that could implement this outcome.

Although this methodology has been successful to understand auction design, regulatory environments, optimal organizations of the firm, etc... it has also faced severe critiques coming from various fronts. The first line of critiques followed the works of Riordan and Sappington (1988), Crémer and McLean (1985, 1988), Johnson, Pratt and Zeckhauser (1990), D'Aspremont, Crémer and Gerard-Varet (1990), Matsushima (1991) and McAfee and Reny (1992). In various contexts, those authors all argue that private information may not entail any cost for an organization. As long as the agents' types are correlated, a clever mechanism designer can design complex lotteries to induce costless information revelation and fully extract the agents' surplus if needed. Without correlation, private information provides information rents to the agents and optimal mechanisms must generally reach a genuine trade-off between rent extraction and allocative efficiency which disappears when types are (even slightly) correlated. This lack of continuity of the optimal mechanism with respect to the information structure is clearly troublesome and a significant impediment to what has been recently known as the *"Wilson Doctrine"*. This doctrine suggests that mechanisms should be robust to small perturbations of the game modelling and in particular of the extent of common knowledge required. Clearly, the received theory of mechanism design fails this test.

Although related to the first critique above, the second source of scepticism on the relevance of the received theory points out that mechanisms are in practice much simpler than the complex monetary lotteries predicted by the theory in correlated environments to fully extract the agents' information rent. In real-world organizations, the scope for yardstick competition and relative performance evaluations seems more limited than predicted by the theory. Agents hardly receive contracts which are so dependent on what their peers might claim. Multilateral contracting in complex organizations seems closer to a superposition of simple bilateral contracts between the principal and each of his agents, although how it differs has to a large extent not yet been explored theoretically.[2]

[1] Gibbard (1973) and Green and Laffont (1977) among others.

[2] Payments on financial and electricity markets depend on how much an agent wants to buy from an

Finally, an often heard criticism of the mechanism design literature points out that communication between the principal and his agents may not be as transparent as assumed. For instance, in the canonical framework for Bayesian collective choices under asymmetric information due to Myerson (1991, Chapter 6.4), all communication between the principal and his agents is public. This facilitates the implementation of the allocation recommended by the mechanism and makes credible that the principal sticks to the complex rewards and punishments predicted by the theory.[3] The flip-side of opaque institutions is that the principal may act opportunistically and manipulate himself the agents' messages if he finds it beneficial to do so. Lack of transparency and opportunistic behavior on the principal's side go hands in hands.

The model developed below responds to those criticisms and goes towards modelling weaker institutions than currently assumed in standard mechanism design theory. To do so, we relax the assumption that communication between the principal and each of his agents is *public*. Instead, we assume that communication is private. Doing so buys us two important things. First, it simplifies significantly mechanisms and shows the major role played by *sell-out contracts* and more generally nonlinear prices in such environments. Second, it restores continuity with respect to the information structure but still maintains the useful role of correlation as a means to better (but not fully) extract the agents' information rent.

• **Simplicity of mechanisms and institutions:** When communication between the principal and his agents is private, the former might have strong incentives to manipulate what he has learned from one agent in order to punish arbitrarily others and reap the corresponding punishment. With *private communication*, the set of incentive feasible mechanisms which cannot be manipulated by the principal is thus severely restricted.

We first prove a *Revelation Principle with private communication.* For a given implementation concept (Bayesian or dominant strategy) characterizing the agents' behavior there is no loss of generality in restricting the analysis to mechanisms that cannot be manipulated by the principal. This characterizes the set of mechanisms that can be implemented with private communication by means of simple *non-manipulability constraints.*

Equipped with this tool, we investigate the form of optimal non-manipulable mechanisms in various environments of increasing complexity.

We start with the simple case where agents run different projects on behalf of the

asset and, of course, of the equilibrium price of that asset but rarely on the whole vector of quantities requested by others. Similarly, incentive payments within firms do not exhibit the randomness predicted by the theory.

[3]This should be contrasted with the case of moral hazard where agents are first asked to report confidentially their types to the principal who then recommends to them some actions which depend only on their own announced types. See Myerson (1982).

principal. The only externality between them is an informational one: Their costs are correlated. In this context, non-manipulability constraints have strong implications on the form of feasible contracts. To avoid manipulations, the principal makes the agent's residual claimant for the return of his own project through a *sell-out contract* whose entry fee depends on the agent's report on his type only. By doing so, the principal commits himself to be indifferent between all possible outputs that a given agent may produce.

In the case of auctions design (both unit and multi-unit), there is also a negative externality between competing bidders on top of the informational one. The greater the probability for an agent of being chosen to produce a good, the lower that probability for another one. Even though such environments are more complex, the same kind of sell-out mechanisms works well. The buyer (principal) selects first the most efficient seller, i.e., the one who pays the highest entry fee, and then offers him a sell-out contract. Again, the principal is indifferent between all possible outputs that this winning agent could produce and, on top of that, the principal does not want to manipulate the identity of who gets to produce the good.

Finally, we consider a team production context where agents exert efforts which are perfect complements. The externality between agents is then positive; the higher the input of one agent is, the higher the input of the other should be. Focusing on incentive mechanisms dependent only on the final output, *sell-out contracts* are now more complex: Each agent only gets a fraction of the overall return of the team activity but still pays an entry fee contingent only on what he reports on his type.

To avoid manipulations by the principal, non-manipulable mechanisms must limit the informational role of what has been learned from others in determining the compensation and output of a given agent. In that context, *personalized nonlinear prices* play a significant role and a *Taxation Principle with private communication* holds. Taking into account the non-manipulability constraints is actually equivalent to imposing that the principal proposes menus of nonlinear prices to the agents and then picks his most desired quantity vector. Complex organizations are then run by contracts which look very much like bilateral ones. Nevertheless, in Bayesian environments, the optimal mechanism still *strictly* dominates the simple superposition of bilateral contracts.

Equipped with this Taxation Principle and the intuition developed in previous examples, we uncover some general techniques to characterize non-manipulable mechanisms in more complex environments than those described earlier on. The key observation is that, under private communication, the variables available for contracting between the principal and any of the agents are not observable by others. In other words, non-manipulability constraints can also be understood as incentive constraints on the principal's side preventing him from lying on what he has learned from contracting with others. With this approach in mind, we use standard mechanism design techniques to recover the shape of nonlinear prices predicted by the Taxation Principle.

• **Continuity of the optimal mechanism:** In all those environments and even when the agents' types are correlated, there always exists a genuine trade-off between rent extraction and efficiency at the optimum. Of course, how this trade-off affects contract design depends on the level of correlation but it does so in an intuitive way. Correlation makes it easier to extract the agents' information rent. When correlation diminishes, the optimal mechanism implements an allocation that comes close to that obtained for independent types *but* without the non-manipulability constraint. Non-manipulability constraints do not bind in the limit of no correlation. With independent types, there always exists an implementation of the second-best which is non-manipulable by the principal. Continuity of the optimal mechanism with respect to the information structure is thus restored when non-manipulability constraints are taken into account. More generally, standard techniques used to perform second-best analysis in settings with independent types can be rather straightforwardly adapted to the case of correlation. In particular, a *generalized virtual cost* that takes into account the correlation of types can be defined and plays the same role as in the independent type case.

Section 2 discusses the relevant literature. Section 3 presents our general model and exposes a few polar cases of interest for the rest of the analysis. In Section 4, we develop a very simple example highlighting the role of private communication in constraining mechanisms. Section 5 proves the Revelation and Taxation Principles with private communication. Equipped with these tools, we characterize optimal mechanisms respectively for the case of unrelated projects in Section 6, unit auctions in Section 7, and, finally, in environments with more general production externalities in Section 8. In Section 9, we analyze the case of team production in a context which has some specific features. Finally, Section 10 concludes. All proofs are relegated to Appendix A. Appendix B contains the analysis of a discrete type model in the case of weak and strong correlation.

2 Literature Review

The strong results on the benefits of correlated information pushed forward by Riordan and Sappington (1988), Crémer and McLean (1985, 1988), Johnson, Pratt and Zeckhauser (1990), D'Aspremont, Crémer and Gerard-Varet (1990), Matsushima (1991) and McAfee and Reny (1992) has already been attacked on various fronts. A first approach is to introduce exogenous limits or costs on feasible punishments by means of risk-aversion and wealth effects (Robert (1991), Eso (2004)), limited liability (Demougin and Garvie (1991)), ex post participation constraints (Dana (1993), Demski and Sappington (1988)), or limited enforceability and veto constraints (Compte and Jehiel (2006)). Here instead, the benefits of using correlated information is undermined by incentive constraints on the principal's side.

A second approach points out that correlated information may not be as generic as suggested by the earlier literature. Enriching the information structure may actually lead to a significant simplification of mechanisms. Neeman (2004) argues that the type of an agent should not simultaneously determine his beliefs on others and be payoff-relevant. Such extension of the type space might reinstall some sort of conditional independence and avoids full extraction.[4] Bergeman and Morris (2005) argue that modelling higher order beliefs leads to ex post implementation whereas Chung and Ely (2005) show that a maxmin principal may want to rely on dominant strategy. Although important, these approaches lead also to somewhat extreme results since Bayesian mechanisms have to be given up.[5,6] Our approach is less extreme. It stills relaxes the common knowledge requirement assumed in standard mechanism design but private communication does so in a simple and tractable way. As a result, optimal mechanisms keep much of the features found in the case of independent types and Bayesian implementation keeps some of its force.

A last approach to avoid the full surplus extraction in correlated environments consists in considering collusive behavior. Laffont and Martimort (2000) showed that mechanisms extracting entirely all the agents' surplus are not robust to horizontal collusion between the agents.[7] Key to this horizontal collusion possibility is the fact that the agents can coordinate their strategies in any grand-mechanism offered by the designer. This coordination is facilitated when communication is public. Hence, our focus on private communication points at another polar case which leaves less scope for such horizontal collusion. Finally, Gromb and Martimort (2005) propose a specific model of expertise involving both moral hazard in information gathering and adverse selection and show that private communication between the principal and each of his experts opens the possibility for some vertical collusion which is harmful for the organization.

Our characterization of non-manipulable mechanisms by means of a simple Taxation Principle is reminiscent of the common agency literature which has already forcefully

[4]Heifetz and Neeman (2006) exhibit conditions under which this conditional independence is generic.

[5]This might appear as too extreme in view of the recent (mostly) negative results pushed forward by the ex post implementation literature in interdependent value environments (see Dasgupta and Maskin (2000), Perry and Reny (2002) and Jehiel and al. (2006))

[6]Also, if the aim of study is to model long-run institutions, it is not clear that agents remain in such high degree of ignorance on each other unless they are also boundedly rational and cannot learn about types distributions from observing past performances.

[7]Their model has only two agents. With more than two agents and in the absence of sub-coalitional behavior, Che and Kim (2006) showed that correlation can still be used to the principal's benefits.

stressed the role of nonlinear prices as means of describing feasible allocations.[8] This resemblance comes at no surprise. Under private communication and centralized mechanism design, 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 the principals' offers. However, and in sharp contrast, there is still a bit of commitment in the game we are analyzing here in the sense that the principal first chooses the menu of nonlinear prices available to the informed agents. In a true common agency game, the informed agents would be offering mechanisms first and there would be a priori no restriction in their possible deviations.[9] Although minor a priori, this difference between our model and the common agency framework will significantly simplify 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.[10] Martimort (2005) discusses this point and argues that one should look for minimal departures of the centralized mechanism design framework which go towards modelling multi-contracting settings. The non-manipulability constraint modeled below can precisely be viewed as such a minimal departure. Once this step is performed, one gets also an important justification for what can be mostly viewed as an ad hoc assumption generally made under common agency: In a complete information environment, Bernheim and Whinston (1986) suggested indeed that principals should offer the so-called *truthful contributions* which are basically the *"sell-out"* contracts that are used below.

The last branch of the literature related to our work is the IO literature on bilateral contracting (Hart and Tirole (1990), O'Brien and Shaffer (1992), McAfee and Schwartz (1994), Segal (1999) and Segal and Whinston (2003) among others). Those papers analyze complete information environments with secret bilateral contracting between a principal (manufacturer) and his agents (retailers). They also focus on some form of opportunism on the principal's side coming from the fact that bilateral contracts with agents are secret. Our framework differs along several lines from this literature but has also some common concerns. First, asymmetric information is a key ingredient of our framework. Second, in our context and except in Section 9, the mechanism offered by the principal itself is publicly observable and the issue of beliefs on other agents' offers does not arise.

[8]Bernheim and Whinston (1986), Stole (1991), Martimort (1992 and 2005), Mezzetti (1997), Martimort and Stole (2002, 2003, 2005), Peters (2001 and 2003). Most of the time private information is modeled on the common agent's side in this literature (an exception is Martimort and Moreira (2005)).

[9]This case is analyzed by Martimort and Moreira (2005) who characterize Bayesian equilibria.

[10]The most noticeable difficulty being of course the multiplicity of equilibria.

3 The Model

• **Preferences and Information:** We consider an organization made of one principal (P) and two agents (A_1 and A_2).[11] Agent A_i produces a good in quantity q_i on the principal's behalf. Players have quasi-linear utility functions defined respectively as:

$$V(q,t) = \tilde{S}(q_1,q_2) - \sum_{i=1}^{2} t_i \quad \text{and} \quad U_i(q,t) = t_i - \theta_i q_i.$$

The vector of goods $q = (q_1, q_2)$ produced respectively by the agents belongs to some set $\mathcal{Q} = \mathcal{Q}_1 \times \mathcal{Q}_2 \subset \mathbb{R}^2_+$. The vector of their monetary transfers $t = (t_1, t_2)$ belongs to some set $\mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2 \subset \mathbb{R}^2$.

The efficiency parameter θ_i is A_i's private information. Types are jointly drawn from the common knowledge non-negative and atomless density function $\tilde{f}(\theta_1, \theta_2)$ whose support is Θ^2 where $\Theta = [\underline{\theta}, \bar{\theta}]$. Sometimes we denote $\theta = (\theta_1, \theta_2)$ the vector of types. For future reference, we will also denote the marginal density and the corresponding cumulative distribution, respectively, as:[12]

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

The principal's surplus function $\tilde{S}(\cdot)$ is increasing in each of its arguments q_i and concave in $q = (q_1, q_2)$. For simplicity, we shall also assume that $\tilde{S}(\cdot)$ is symmetric.

This formulation encompasses three cases of interest to whom we shall devote more attention in the sequel:

- *Unrelated projects:* $\tilde{S}(\cdot)$ is separable in both q_1 and q_2 and thus can be written as $\tilde{S}(q_1, q_2) = S(q_1) + S(q_2)$ for some function $S(\cdot)$ that is assumed to be increasing and concave with the Inada condition $S'(0) = +\infty$ and $S(0) = 0$.

- *Perfect Substitutability:* $\tilde{S}(\cdot)$ is in fact a function of the total production $q_1 + q_2$ only: $\tilde{S}(q_1, q_2) = S(q_1 + q_2)$ for some increasing and concave function $S(\cdot)$ still satisfying 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.

[11] Our framework can be extended in a straightforward manner to settings with more than two agents at the cost of some notational burden.

[12] In the case of independent types, $\tilde{f}(\theta_1, \theta_2) = f(\theta_1) f(\theta_2)$.

With unrelated projects, the only externality between agents is an informational one and goes through the possible correlation of their cost parameters. This correlation may help the principal to better design incentives for truthful behavior. Perfect substitutability arises instead in the context of a procurement auction for an homogenous good. Perfect complementarity occurs for instance in the case of complementary inputs.[13]

• **Mechanisms:** In standard mechanism design, messages are *public*, i.e., the reports made by A_i on his type is observed by A_{-i} before the given allocation requested by the mechanism gets implemented. We shall now relax this assumption and focus on the case of *private communication*. Each agent A_i privately communicates with the principal some message m_i. Then, the principal releases a report $\hat{m}_i$ to A_{-i} before implementing the requested transfers and quantity for A_{-i}. Each agent A_i observes just the private message m_i that he sends to the principal and the report $\hat{m}_{-i}$ made by the principal on the message m_{-i} the latter has received from agent A_{-i}.

A mechanism is a pair $(g(\cdot), \mathcal{M})$. The outcome function $g(\cdot) = (g_1(\cdot), g_2(\cdot))$ is itself a pair of outcome functions. $g_i(\cdot)$ maps the communication space $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ into the set of feasible allocations $\mathcal{Q}_i \times \mathcal{T}_i$. The outcome function $g_i(\cdot)$ associates to any message vector $m = (m_i, \hat{m}_{-i})$ from the joint communication space $\mathcal{M} = \mathcal{M}_i \times \mathcal{M}_{-i}$ an output $q_i(m_i, \hat{m}_{-i})$ and a transfer $t_i(m_i, \hat{m}_{-i})$ for agent A_i.

To avoid inferences by A_i on the true report made by A_{-i} to the principal just by checking whether the transfer and output given to A_{-i} are consistent with his own private report to the principal and on the message he received, we assume that the output q_{-i} and transfer t_{-i} are not observed by A_i.

With private communication, the principal once informed on an agent's report might manipulate this report to extract more from the other agent if he finds it attractive. Of course, in the background the Court of Law that can observe the private messages m sent by the agents but enforce the released messages $\hat{m}$ is corruptible and colludes with the principal. In this sense, our modelling captures a case for weak institutions where the opacity of transactions leaves scope to all sorts of gaming.

• **Timing:** 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, all agents simultaneously accept or refuse this mechanism. If agent A_i refuses, he gets no transfer ($t_i = 0$) and produces nothing ($q_i = 0$) so that he obtains a payoff normalized to zero. Fourth, agents privately and simultaneously send messages $m = (m_1, m_2)$ to the principal. Fifth, and this is the novelty of our modelling, the principal privately reports the messages $\hat{m}_1$ to A_2 and $\hat{m}_2$ to A_1. Finally, the corresponding outputs

[13] By a quick change of set-up, perfect substitutability is also the relevant case to treat standard auctions while perfect complementarity is the relevant case to treat public good problems.

and transfers for each agent are implemented according to the reported messages.

For most of the paper, our equilibrium concept is perfect Bayesian equilibrium (thereafter PBE).[14]

• **Benchmark:** Without already entering into the details of the analysis, let us consider the case of a mechanism relying on public messages. If types are correlated, a by-now standard result in the literature, is that the first-best outcome can be either achieved (with discrete types) or arbitrarily approached (with a continuum of types). In sharp contrast with what economic intuition commends, there is no trade-off between efficiency and rent extraction in such correlated environments. In the case of unrelated projects, for instance, the (symmetric) first-best output requested from each agent trades off the marginal benefit of production against its marginal cost, namely:

$$S'(q^{FB}(\theta_i, \theta_{-i})) = \theta_i, \quad i = 1, 2. \tag{1}$$

When types are instead independently distributed, it is also well-known that the first-best outcome can no longer be implemented. Because asymmetric information gives information rents to the agents and those rents are viewed as costly by the principal, there is now a real trade-off between efficiency and rent extraction. The marginal benefit of production must be equal to the *virtual* marginal cost. With unrelated projects, the (symmetric) second-best output is therefore given by the so-called *Baron-Myerson* outcome[15] for each agent:

$$S'(q^{BM}(\theta_i)) = \theta_i + \frac{F(\theta_i)}{f(\theta_i)}. \tag{2}$$

Provided that the *Monotone Hazard Rate Property* holds, namely $\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right) > 0 \quad \forall \theta \in \Theta$, $q^{BM}(\theta_i)$ is indeed the solution.[16]

It should be straightforward to observe that this Baron-Myerson outcome is obtained when the principal contracts separately with each agent on the basis of his own report only. This would also be the solution if the principal was a priori restricted to use bilateral contracts with each agent even in settings with correlated types. If types are correlated, the discrepancy between (1) and (2) measures then the loss when going from a multilateral contracting environment to a bilateral contracting one.

[14]Except in Section 6.2 where we study also dominant strategy implementation as the implementation concept for the agents' behavior and generalize some of our results.

[15]Baron and Myerson (1982).

[16]Otherwise, bunching may arise at the optimal contract. See Laffont and Martimort (2002, Chapter 3) for instance.

4 A Simple Example

To fix ideas and already give some preliminary insights on the general analysis that will be performed later on, let us consider a very simple example where the principal's ability to manipulate information significantly undermines optimal contracting. A buyer (the principal) wants to procure one unit of a good from a single seller (the agent). The gross surplus that accrues to the principal when consuming this unit is S. The seller's cost may take two values $\theta \in \{\underline{\theta}, \bar{\theta}\}$ (where $\Delta\theta = \bar{\theta} - \underline{\theta} > 0$) with respective probabilities ν and $1 - \nu$. The following conditions hold:

$$\bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta > S > \bar{\theta}. \tag{3}$$

The right-hand side inequality simply means that trade is efficient with both types of seller under complete information. The left-hand side inequality instead captures the fact that trade is no longer efficient with the high cost seller when there is asymmetric information. The buyer makes then an optimal take-it-or-leave-it offer to the seller at a price $\underline{\theta}$. Only an efficient seller accepts this offer and trades.

Let us now suppose that the buyer learns a signal $\sigma \in \{\underline{\theta}, \bar{\theta}\}$ on the agent's type ex post, i.e., once the agent has already reported his cost parameter. This signal is informative on the agent's type, and more specifically,

$$\text{proba}\{\sigma = \underline{\theta}|\underline{\theta}\} = \text{proba}\{\sigma = \bar{\theta}|\bar{\theta}\} = \rho > \frac{1}{2} > \text{proba}\{\sigma = \underline{\theta}|\bar{\theta}\} = \text{proba}\{\sigma = \bar{\theta}|\underline{\theta}\} = 1 - \rho.$$

Let assume that the signal σ is publicly verifiable. The price paid by the buyer for one unit of the good should in full generality be a function of the seller's report on his cost as well as the realized value of the signal. Let denote by $t(\theta, \sigma)$ this price.

Looking for transfers that would implement the first-best production decision, incentive compatibility for both types of seller requires now respectively:

$$\begin{aligned}
\rho t(\underline{\theta}, \underline{\theta}) + (1-\rho)t(\underline{\theta}, \bar{\theta}) &\geq \rho t(\bar{\theta}, \underline{\theta}) + (1-\rho)t(\bar{\theta}, \bar{\theta}) \\
(1-\rho)t(\bar{\theta}, \underline{\theta}) + \rho t(\bar{\theta}, \bar{\theta}) &\geq (1-\rho)t(\underline{\theta}, \underline{\theta}) + \rho t(\underline{\theta}, \bar{\theta}).
\end{aligned}$$

Normalizing at zero the seller's outside opportunities, the respective participation constraints of both types (assuming that both types produce) can be written as:

$$\begin{aligned}
\rho t(\underline{\theta}, \underline{\theta}) + (1-\rho)t(\underline{\theta}, \bar{\theta}) - \underline{\theta} &\geq 0 \\
(1-\rho)t(\bar{\theta}, \underline{\theta}) + \rho t(\bar{\theta}, \bar{\theta}) - \bar{\theta} &\geq 0.
\end{aligned}$$

It is well known from the work of Riordan and Sappington (1988) that the buyer can extract all surplus from the seller and implement the first-best outcome by properly designing the price schedule: It suffices (among many other possibilities) to set price lotteries which bind all incentive and participation constraints:

$$t(\underline{\theta},\underline{\theta}) = \frac{\rho}{2\rho-1}\underline{\theta} > \underline{\theta} > 0, \qquad t(\underline{\theta},\bar{\theta}) = -\frac{1-\rho}{2\rho-1}\underline{\theta} < 0,$$
$$t(\underline{\theta},\bar{\theta}) = -\frac{1-\rho}{2\rho-1}\bar{\theta} < 0, \qquad t(\bar{\theta},\bar{\theta}) = \frac{\rho}{2\rho-1}\bar{\theta} > \bar{\theta} > 0.$$

This mechanism punishes the seller whenever his report conflicts with the public signal. Otherwise the seller is rewarded and paid more than his marginal cost.

Consider now the case where the principal *privately* observes σ. The price scheme above can no longer be used by the principal since it is manipulable. Once the seller has already reported his type, the buyer may want to claim that he receives conflicting evidence on the agent's report to pocket the corresponding punishment instead of giving the reward. To avoid those manipulations by the principal, the price must be independent of the realized signal:

$$t(\theta,\underline{\theta}) = t(\theta,\bar{\theta}) \qquad \forall\theta \in \{\underline{\theta},\bar{\theta}\}.$$

With this *non-manipulability constraint*, we are back to the traditional screening model *without* ex post information. Given Assumption 3, it is suboptimal for the buyer to procure the good in all states of nature.

This simple example illustrates the consequences of having the principal manipulate information which, if otherwise public, would be used for screening purposes. In the sequel and contrary to this simple example with only one agent, information is no longer exogenously produced but results from contracting with another agent. Second, the non-manipulability is derived rather than assumed. Moreover, and again in sharp contrast with the above example where output was fixed (one unit of the good had to be produced irrespectively of the observed/reported signal σ), the non-manipulability of a mechanism by the principal may require distorting both outputs and transfers. Although the kind of lotteries used above lose much of their content, they might still have some value if output is accordingly distorted.

5 Revelation and Taxation Principles with Private Communication

5.1 The Revelation Principle

Let us come back to our general model. To start the analysis, we look for a full characterization of the set of allocations that can be achieved as PBEs of the overall contracting game where the principal first offers a mechanism $(g(\cdot), \mathcal{M})$ using a priori any arbitrary communication space $\mathcal{M}$ but with private communication and, second, may then manipulate the agents' reports when releasing each of those reports to the other party.

For any agents' strategy vector $m^*(\cdot)$, $\sup m^*(\cdot)$ denotes the support of the strategies, i.e., the set of messages m that are sent with strictly positive probability given $m^*(\cdot)$.

For a fixed mechanism $(g(\cdot), \mathcal{M})$, let us define the continuation PBEs that such mechanism induce as follows:

Definition 1 *: A continuation PBE for any arbitrary mechanism $(g(\cdot), \mathcal{M})$ is a triplet $\{m^*(\cdot), \hat{m}^*(\cdot), d\mu(\theta|m)\}$ such that:*

- *The agents' strategy vector $m^*(\theta) = (m_1^*(\theta_1), m_2^*(\theta_2))$ from $\Theta \times \Theta$ into $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ forms a Bayesian equilibrium given the principal's manipulation strategy $\hat{m}^*(\cdot)$*

$$m_i^*(\theta_i) \in \arg\max_{m_i \in \mathcal{M}_i} E_{\theta_{-i}} \left(t_i(m_i, \hat{m}_{-i}^*(m_i, m_{-i}^*(\theta_{-i})) - \theta_i q_i(m_i, \hat{m}_{-i}^*(m_i, m_{-i}^*(\theta_{-i}))))|\theta_i \right); \tag{4}$$

- *The principal's manipulation $\hat{m}^*(\cdot)$ from $\mathcal{M}$ onto satisfies*

$$\hat{m}^*(m) \in \arg\max_{(\hat{m}_1, \hat{m}_2) \in \mathcal{M}} \tilde{S}(q_1(m_1, \hat{m}_2), q_2(\hat{m}_1, m_2)) - \sum_{i=1}^{2} t_i(m_i, \hat{m}_{-i}), \quad \forall m = (m_1, m_2) \in \mathcal{M}; \tag{5}$$

- *The principal's posterior beliefs $d\mu(\theta|m)$ on the agents' types follow Bayes's rule whenever possible (i.e., when $m \in \sup m^*(\cdot)$) and are arbitrary otherwise.*

Given a mechanism $(g(\cdot), \mathcal{M})$, a continuation PBE $\{m^*(\cdot), \hat{m}^*(\cdot), d\mu(\theta|m)\}$ induces an allocation $a = g \circ \hat{m}^* \circ m^*$ which maps Θ^2 on $\mathcal{Q} \times \mathcal{T}$.

Definition 2 *: A mechanism $(g(\cdot), \mathcal{M})$ is non-manipulable if and only if $\hat{m}^*(m) = m$, for all $m \in \sup m^*(\cdot)$ at a continuation PBE.*[17]

Definition 3 *: A direct mechanism $(\bar{g}(\cdot), \Theta^2)$ is truthful if and only if $m^*(\theta) = \theta$, for all $\theta \in \Theta$ at a continuation PBE.*

We are now ready to state:

[17]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 PBEs.

Proposition 1 *: **The Revelation Principle with Private Communication.** Any allocation $a(\cdot)$ achieved at a PBE of any arbitrary mechanism $(g(\cdot), \mathcal{M})$ with private communication can also be implemented as a truthful and non-manipulable perfect Bayesian equilibrium of a direct mechanism $(\bar{g}(\cdot), \Theta^2)$.*

The Bayesian incentive compatibility constraints describing the agents' behavior are written as usual:

$$\underset{\theta_{-i}}{E}\left(t_i(\theta_i, \theta_{-i}) - \theta_i q_i(\theta_i, \theta_{-i})|\theta_i\right) \geq \underset{\theta_{-i}}{E}\left(t_i(\hat{\theta}_i, \theta_{-i}) - \theta_i q_i(\hat{\theta}_i, \theta_{-i})|\theta_i\right) \quad \forall(\theta_i, \hat{\theta}_i) \in \Theta^2. \quad (6)$$

The following *non-manipulability* constraints stipulate that the principal will not misrepresent to one agent what he has learned from another agent's report:

$$\tilde{S}(q_1(\theta_1, \theta_2), q_1(\theta_1, \theta_2)) - \sum_{i=1}^{2} t_i(\theta_i, \theta_{-i}) \quad \geq \quad \tilde{S}(q_1(\theta_1, \hat{\theta}_2), q_2(\hat{\theta}_1, \theta_2)) - \sum_{i=1}^{2} t_i(\theta_i, \hat{\theta}_{-i}),$$
$$\forall(\theta_1, \theta_2, \hat{\theta}_1, \hat{\theta}_2) \in \Theta^4. \quad (7)$$

In the sequel, we will analyze in much details several cases of interest (unrelated projects or various production externalities) and will trace out the impact of the non-manipulability constraint (7) on optimal mechanisms in those different contexts.

5.2 The Taxation Principle

Beforehand, we propose an alternative formulation of the problem which clarifies the impact of private communication and uncovers a link between our analysis and the common agency literature. Consider indeed the following three-stage mechanism:

- At stage 1, the principal offers menus of nonlinear prices $\{T_i(q_i, \hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ which stipulate a payment for each agent as a function of how much he produces and which type he reports.
- At stage 2, agents report simultaneously and non-cooperatively their types and thus pick schedules among the offered menus. The menu is truthful and each agent chooses truthfully the schedule corresponding to his own type.
- At stage 3, the principal chooses how much output to request from each agent.

The important point to notice is that the principal optimally chooses the agents' outputs conditionally on what he has learned from their truthful reports. This aspect of the game is clearly reminiscent of the common agency literature. As in this literature, the player at the nexus of all contracts optimally reacts to the others' choices. However, and in sharp contrast, there is still a bit of commitment in the game we are analyzing here since the principal initially chooses the menu of nonlinear prices available to the informed agents. In common agency games, not only the players moving first would be the informed agents but there would be a priori no restriction in the deviations that they could envision.

Proposition 2 *: **The Taxation Principle.***

- *Any allocation $a(\theta)$ achieved at a continuation PBE of a non-manipulable direct Bayesian mechanism $(\bar{g}(\cdot), \Theta)$ with private communication can alternatively be implemented as a continuation PBE of a modified common agency game which requires each agent A_i to choose truthfully a nonlinear price from menus $\{T_i(q_i, \hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ and then the principal to choose outputs.*

- *Conversely, any allocation $a(\theta)$ achieved at a continuation PBE of a modified common agency game which has agents choosing truthfully nonlinear prices from menus $\{T_i(q_i, \hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ and then the principal choosing outputs can alternatively be implemented as a truthful continuation PBE of a non-manipulable direct Bayesian mechanism $(\bar{g}(\cdot), \Theta)$ with private communication.*

Proposition 2 shows the exact nature of the non-manipulability constraint in practice: The principal can only use personalized nonlinear prices to reward the agents. Of course, those schedules are designed in an incentive compatible way. More general mechanisms, especially complex lotteries making the monetary payment of one agent depend on the whole vector of the agents' announcements, are manipulable and thus cannot be used in any credible way by the principal. The Taxation Principle above also shows that non-manipulability does not necessarily imply bilateral contracting. Retaining the possibility to pick q after that agents have chosen the nonlinear prices still allows the principal to exploit informational and production externalities if any.

6 Unrelated Projects

6.1 Bayesian Implementation

To familiarize ourselves with the non-manipulability constraint and its consequences on mechanism design, let us start with the simplest case where agents work on projects without any production externality. The principal's gross surplus function is separable:

$$\tilde{S}(q_1, q_2) = \sum_{i=1}^{2} S(q_i).$$

Written in terms of direct mechanisms, the non-manipulability constraint (7) yields that there exists an arbitrary function $h_i(\theta_i)$ such that:

$$S(q_i(\theta_i, \theta_{-i})) - t_i(\theta_i, \theta_{-i}) = h_i(\theta_i) \tag{8}$$

Equation (8) shows that each agent is made residual claimant for the part of the principal's objective function which is directly related to his own output. The nonlinear price which achieves this objective is a so-called *sell-out contract*:

$$T_i(q, \theta_i) = S(q) - h_i(\theta_i). \tag{9}$$

Everything happens thus as if agent A_i had to pay upfront an amount $h_i(\theta_i)$ to have the right to produce on the principal's behalf. Then, the agent enjoys all returns $S(q)$ on the project he is running for the principal; $h_i(\theta_i)$ being the principal's payoff in his relationship with A_i which does not depend on the amount produced. Of course, fixed-fees are fixed so that participation by all types is ensured.

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

$$U_i(\theta_i) = \underset{\theta_{-i}}{E} \left(S(q_i(\theta_i, \theta_{-i})) - \theta_i q_i(\theta_i, \theta_{-i}) | \theta_i\right) - h_i(\theta_i). \tag{10}$$

Individual rationality implies:

$$U_i(\theta_i) \geq 0 \qquad \forall i, \quad \forall \theta_i \in \Theta. \tag{11}$$

Bayesian incentive compatibility can be written as:

$$U_i(\theta_i) = \arg\max_{\hat{\theta}_i \in \Theta_i} \underset{\theta_{-i}}{E} \left(S(q_i(\hat{\theta}_i, \theta_{-i})) - \theta_i q_i(\hat{\theta}_i, \theta_{-i}) | \theta_i\right) - h_i(\hat{\theta}_i) \quad \forall i, \ \forall \theta_i \in \Theta. \tag{12}$$

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 (12). This similarity makes the analysis of the set of non-manipulable incentive compatible allocations look very close to that with independent types. This will trigger some strong resemblance in the properties of the optimal mechanisms.

Assuming differentiability of $q_i(\cdot)$,[18] simple revealed preferences arguments show that $h_i(\cdot)$ is itself differentiable. The local first-order condition for Bayesian incentive compatibility becomes thus:[19]

$$\dot{h}_i(\theta_i) = \underset{\theta_{-i}}{E}\left((S'(q_i(\theta_i,\theta_{-i})) - \theta_i)\frac{\partial q_i(\theta_i,\theta_{-i})}{\partial\theta_i}|\theta_i\right) \quad \forall i,\ \forall\theta_i \in \Theta; \qquad (13)$$

Consider thus any output schedule $q_i(\cdot)$ which is monotonically decreasing in θ_i and which lies below the first-best. Then (13) shows that necessarily, the $h_i(\cdot)$ function that implements this output schedule is necessarily also decreasing in θ_i. In other words, less efficient types are requested to pay lower up-front payments. The Bayesian incentive constraint (13) captures then the trade-off faced by an agent with type θ_i. By exaggerating his type, this agent will have to pay a lower up-front payment. However, he will also produce less and enjoy a lower expected surplus from residual claimancy. Incentive compatibility is achieved when those two effects just compensate each other.

To highlight the trade-off between efficiency and rent extraction, it is useful to rewrite incentive compatibility in terms of the agents' information rent. (13) becomes:

$$\dot{U}_i(\theta_i) = -\underset{\theta_{-i}}{E}(q_i(\theta_i,\theta_{-i})|\theta_i) + \underset{\theta_{-i}}{E}\left((S(q_i(\theta_i,\theta_{-i})) - \theta_i q_i(\theta_i,\theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right). \qquad (14)$$

To better understand the right-hand side of (14), consider an agent with type θ_i willing to mimic a less efficient type $\theta_i + d\theta_i$. By doing so, this agent produces the same amount than this less efficient type at a lower marginal cost. This gives a first source of information rent to type θ_i which is worth:

$$\underset{\theta_{-i}}{E}(q_i(\theta_i,\theta_{-i})|\theta_i)\, d\theta_i.$$

Note that this source of rent is there whether there is correlation or not.

By mimicking this less efficient type, type θ_i affects also how the principal interprets the information contained in the other agent's report to adjust θ_i's own production. The corresponding marginal rent is the second term on the right-hand side of (14):

$$-\underset{\theta_{-i}}{E}\left((S(q_i(\theta_i,\theta_{-i})) - \theta_i q_i(\theta_i,\theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right) d\theta_i.$$

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

$$-\underset{\theta_{-i}}{E}\left(\frac{\partial q_i(\theta_i,\theta_{-i})}{\partial\theta_i}|\theta_i\right) + \underset{\theta_{-i}}{E}\left((S'(q_i(\theta_i,\theta_{-i})) - \theta_i)\frac{\partial q_i(\theta_i,\theta_{-i})}{\partial\theta_i}\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right) \geq 0$$
$$\forall i = 1, 2,\ \forall\theta_i \in \Theta. \qquad (15)$$

[18] Because conditional expectations depend on A_i's type, one cannot also derive from revealed preferences arguments that $q_i(\cdot)$ is itself monotonically decreasing in θ_i.

[19] We postpone the analysis of the global incentive compatibility constraints to the Appendix.

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

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

subject to constraints (11) to (15).

To get sharp predictions on the solution, we need an important assumption on the type distribution which generalizes to our correlated environment the well-known assumption of monotonicity of the virtual cost:

Assumption 1 *Monotonicity of 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 always non-negative, strictly increasing in θ_i and decreasing in θ_{-i}.

This assumption generalizes to setting that involve some correlation the standard assumption on virtual types. It ensures that optimal outputs are decreasing functions of own types, a condition which is neither sufficient nor necessary for implementability as it can be seen from (15) but which remains a useful ingredient to ensure that this latter condition holds.

Assumption 1 holds as long as the following three other assumptions are satisfied:

Assumption 2 *Weak correlation:*[20]

$$\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i) \textit{ is close enough to zero, for all } (\theta_i, \theta_{-i}) \in \Theta^2.$$

Assumption 3 *Monotone hazard rate property (MHRP):*

$$\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right) > 0 \textit{ for all } \theta \in \Theta..$$

[20]In Appendix B, we analyze also the case of a strong correlation for a model with discrete types.

Assumption 4 *Monotone likelihood ratio property (MLRP):*

$$\frac{\partial}{\partial \theta_{-i}} \left(\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \right) \geq 0 \textit{ for all } (\theta_i, \theta_{-i}) \in \Theta^2.$$

Assumptions 3 and 4 are standard in Incentive Theory. They will help us to build intuition on some of the results below.

Proposition 3 : ***Unrelated Projects.*** *Assume that Assumptions 1 and 2 both hold. 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*

  $$S'(q^{SB}(\theta_i, \theta_{-i})) = \varphi(\theta_i, \theta_{-i}), \tag{16}$$

  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

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

- *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, there is a strong similarity between incentive constraints for a non-manipulable Bayesian mechanism and for independent types. This similarity suggests that the trade-off between efficiency and rent extraction that occurs under independent types carries over in our context. This intuition is confirmed by equation (16) which highlights the output distortion capturing this trade-off even when types may be correlated.

When types are independently distributed, the right-hand side of (16) is the same as that of (2). The principal finds useless the report of one agent to better design the other's incentives. He must give up some information rent to induce information revelation anyway. Outputs are accordingly distorted downward to reduce those rents and the standard Baron-Myerson distortions follow. The important point to notice is that the optimal multilateral contract with unrelated projects and independent types can be implemented with a pair of bilateral contracts which are *de facto* non-manipulable by the principal. The non-manipulability constraint has no bite in this case.

When types are instead correlated, the agents' rent can be (almost) fully extracted in this context with a continuum of types[21] and the first-best output can be implemented. Of course, this result relies on the use of complex lotteries linking an agent's payment to what the other reports. Those schemes are manipulable and thus no longer used with private communication.

A similar logic to that of Section 4 applies here with an added twist. Indeed, in our earlier example, non-manipulability constraints imposed only a restriction on transfers since output was fixed at one unit. When output may also vary, non-manipulability constraints impose only that the principal's payoff remains constant over all possible transfer-output pairs that he offers to an agent. This still allows the principal to link also agent A_i's payment to what he learns from agent A_{-i}'s report as long as A_i's output varies accordingly. Doing so, the principal may still be able to incorporate some of the benefits of correlated information in the design of each agent's contract. The multilateral contract signed with both agents does better than a pair of bilateral contracts. If we switch to the interpretation in terms of nonlinear prices, retaining control on the quantity produced by each agent allows the principal to still somewhat exploit the informational externalities.

To understand the nature of the output distortions and the role of the correlation, it is useful to compare the solution found in (16) with the standard Baron-Myerson formula (2) which corresponds also to the optimal mechanism had the principal contracted separately with each agent. As already noticed, this pair of bilateral contracts is of course non-manipulable since each agent's output and payment depend only on his own type. Whether communication is public or private does not matter. Let us see how those bilateral contracts affect the agents' information rent. Using (14), we observe that the second term on the right-hand side is null for a bilateral contract implementing $q^{BM}(\theta_i)$ since

$$\underset{\theta_{-i}}{E}\left(\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right) = 0. \tag{17}$$

By departing from the Baron-Myerson outcome, one affects this second term and reduce the agent's information rent. Think now of the principal as using A_{-i}'s report to improve his knowledge of agent A_i's type. Suppose that the principal starts from the 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 "maximum likelihood estimator," the principal should infer how likely it is that A_i lies on his type by simply observing A_{-i}'s report.

From Assumption 4 and condition (17), 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 θ_i if he learns from A_{-i}, $\theta_{-i} = \theta^*_{-i}(\theta_i)$. Nothing a priori unknown has been learned from A_{-i}'s report in that case. The only concern of the principal remains reducing the first-term

[21]McAfee and Reny (1992).

on the right-hand side of (14). For this type θ_i, the optimal output is still equal to the Baron-Myerson solution.

Think now of an observation $\theta_{-i} > \theta^*_{-i}(\theta_i)$. Because Assumption 4 holds, it is much likely that the principal infers from such observation that A_i is less efficient than what it pretends to be. Such signal let the principal think that the agent 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 in the Baron-Myerson solution. Instead, a signal $\theta_{-i} < \theta^*_{-i}(\theta_i)$ is more likely to confirm the agent's report if he exaggerates his type. Curbing these incentives requires increasing further the distortion beyond the Baron-Myerson solution.

Corollary 1 *: Under the assumptions of Proposition 3, the following output ranking holds*

$$q^{SB}(\theta_i, \theta_{-i}) \geq q^{BM}(\theta_i) \quad \Leftrightarrow \quad \theta_{-i} \geq \theta^*_{-i}(\theta_i) \quad \forall \theta_i \in \Theta,$$

where $\theta^*_{-i}(\theta_i)$ *is increasing.*

Remark 1: With correlated types it is no longer true that the local second-order condition given by equation (15) is always sufficient to guarantee global incentive compatibility even if the agents' utility function satisfies the Spence-Mirrlees condition. However, Assumption 2 ensures that the mechanism identified in Proposition 3 is globally incentive compatible so that our approach remains valid.[22] ■

6.2 Dominant Strategy Implementation

We show here that if we strengthen the implementation concept and require that agents play dominant strategies, informational externalities can no longer be exploited and the principal cannot do better than offering bilateral contracts. Therefore, the Baron-Myerson outcome becomes optimal even with correlated types in such an environment.

The notions of private communication and non-manipulability are actually independent of the implementation concept that is used to describe the agents' behavior. Our framework can be straightforwardly adapted to dominant strategy implementation. For any arbitrary mechanism $(g(\cdot), \mathcal{M})$, a dominant strategy continuation equilibrium is then defined as follows:

Definition 4 *: A continuation dominant strategy equilibrium is a triplet* $\{m^*(\cdot), \hat{m}^*(\cdot), d\mu(\theta|m)\}$ *such that:*

[22]See the Appendix for details.

- $m^*(\theta) = (m_1^*(\theta_1), m_2^*(\theta_2))$ *from* $\Theta \times \Theta$ *into* $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ *forms a dominant strategy equilibrium given the principal's manipulation strategy* $\hat{m}^*(\cdot)$

$$m_i^*(\theta_i) \in \arg\max_{m_i \in \mathcal{M}_i} t_i(m_i, \hat{m}_{-i}^*(m_i, m_{-i})) - \theta_i q_i(m_i, \hat{m}_{-i}^*(m_i, m_{-i})), \forall m_{-i} \in \mathcal{M}_{-i}; \tag{18}$$

- *The principal's manipulation* $\hat{m}^*(\cdot)$ *from* $\mathcal{M}$ *onto satisfies*

$$\hat{m}^*(m) \in \arg\max_{(\hat{m}_1, \hat{m}_2) \in \mathcal{M}} \tilde{S}(q_1(m_1, \hat{m}_2), q_2(\hat{m}_1, m_2)) - \sum_{i=1}^{2} t_i(m_i, \hat{m}_{-i}), \quad \forall m = (m_1, m_2) \in \mathcal{M}; \tag{19}$$

- *The principal's posterior beliefs on the agents' types are derived following Bayes's rule whenever possible (i.e., when* $m \in \sup m^*(\cdot)$*) and are arbitrary otherwise.*

We immediately adapt our previous findings to get:

Proposition 4 *: **The Revelation Principle for Dominant Strategy Implementation with Private Communication.** Any allocation* $a(\cdot)$ *achieved at a dominant strategy equilibrium of any arbitrary mechanism* $(g(\cdot), \mathcal{M})$ *with private communication can alternatively be implemented as a truthful and non-manipulable dominant strategy equilibrium of a direct mechanism* $(\bar{g}(\cdot), \Theta^2)$.

Under dominant strategy implementation, non-manipulability for unrelated projects is still characterized by the condition:

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

The analysis of the dominant strategy incentive and participation constraints is standard. If we denote $u_i(\theta_i, \theta_{-i}) = t_i(\theta_i, \theta_{-i}) - \theta_i q_i(\theta_i, \theta_{-i})$ the ex post rent received by an agent with type θ_i, dominant strategy incentive compatibility amounts to the following implementability conditions:

$$q_i(\theta_i, \theta_{-i}) \text{ weakly decreasing in } \theta_i, \text{ for all } \theta_{-i},$$

and

$$u_i(\theta_i, \theta_{-i}) = u_i(\bar{\theta}, \theta_{-i}) + \int_{\theta_i}^{\bar{\theta}} q_i(u, \theta_{-i}) du. \tag{20}$$

When the set of incentive feasible allocations is independent of the agents' beliefs, one may also strengthen the participation condition and impose ex post participation constraints:

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

Proposition 5 *Under dominant strategy implementation and ex post participation, the optimal non-manipulable mechanism can be achieved with a pair of bilateral contracts implementing the Baron-Myerson outcome for each agent,* $(t_i^{BM}(\theta_i), q_i^{BM}(\theta_i))$ *such that*

$$t_i^{BM}(\theta_i) = \theta_i q_i^{BM}(\theta_i) + \int_{\theta_i}^{\bar{\theta}} q_i^{BM}(u)du.$$

Remark 2: Bilateral contracts are suboptimal if we do not impose non-manipulability even under dominant strategy implementation and ex post participation. In that case, when the required monotonicity conditions are satisfied, the optimal quantities are

$$S'(q_i(\theta_i, \theta_{-i})) = \theta_i + \frac{\tilde{F}(\theta_i \mid \theta_{-i})}{\tilde{f}(\theta_i \mid \theta_{-i})} \tag{21}$$

and the optimal mechanism yields a strictly higher payoff than a pair of bilateral contracts. This result highlights the fact that non-manipulability and dominant strategy implementability are clearly two different concepts with quite different implications. One restriction does not imply the other. Moreover, these restrictions justify simple bilateral contracts only when taken in tandem. ■

7 Auction Design

Auction design provides a nice area of application of our theory. The private communication hypothesis seems quite relevant to study auctions organized on the internet. In light of the recent development of such trading mechanisms, it is certainly a major objective to extend auction theory in that direction.[23]

For simplicity, let us first suppose that the principal wants to procure only one unit of a good. Then, q_i can be interpreted as the probability that agent A_i serves himself the principal. Let also denote by S the (constant) valuation of the principal for that unit. Assuming that $S \in [\underline{\theta}, \bar{\theta}]$ to avoid triviality, the first-best is thus:

$$q_i^*(\theta_i, \theta_{-i}) = 1 \quad \Leftrightarrow \quad \theta_i = \min\{S, \theta_{-i}\}.$$

Expressed with direct revelation mechanisms, non-manipulability constraints become:

$$(\theta_1, \theta_2) \in \arg \max_{\{(\hat{\theta}_1, \hat{\theta}_2) \in \Theta^2 \mid \sum_{i=1}^2 q_i(\theta_i, \hat{\theta}_{-i}) \leq 1, q_i(\theta_i, \hat{\theta}_{-i}) \geq 0\}} \left(\sum_{i=1}^2 q_i(\theta_i, \hat{\theta}_{-i}) \right) S - \sum_{i=1}^2 t_i(\theta_i, \hat{\theta}_{-i}).$$

[23]The private communication hypothesis is also a consistent way to give a more active role to the auctioneer and build a general model of "shill bidding."

It is indeed straightforward to see that, for instance, the second-price sealed-bid auction does not belong to that class. Once he has learned all bids, the principal chooses to buy from the agent with the lowest one and may want to pretend that the second-lowest bid is just ϵ above. Anticipating that, agents no longer have a dominant strategy. Instead, the first-price auction is non-manipulable.

Note that the principal may manipulate private reports only up to the extent that the manipulated allocation remains feasible. This explains the constraint $\sum_{i=1}^{2} q_i(\theta_i, \hat{\theta}_{-i}) \leq 1$ in the above maximand. Suppose for a while that we neglect this constraint. By doing so, we thus allow *a priori* more manipulations for the principal and restrict de facto the set of non-manipulable mechanisms with the possible consequence that the truly optimal mechanism may actually not belong to that class. The corresponding more stringent non-manipulability constraints can be written as:

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

Non-manipulable mechanisms are then readily obtained since we are *almost* as in the case of unrelated projects with some separability in the principal's objective function between the surplus obtained from each agent. For this class of non-manipulable mechanisms, there should thus exist a function $h(\cdot)$[24] such that:

$$Sq_i(\theta_i, \theta_{-i}) - t_i(\theta_i, \theta_{-i}) = h(\theta_i) \quad \forall (\theta_i, \theta_{-i}) \in \Theta^2. \tag{22}$$

Of course, the feasibility conditions below remain as constraints of the principal's optimization problem:

$$\sum_{i=1}^{2} q_i(\theta_i, \theta_{-i}) \leq 1 \quad \text{and} \quad q_i(\theta, \theta_{-i}) \geq 0, \quad \forall (\theta_i, \theta_{-i}) \in \Theta^2. \tag{23}$$

Next proposition shows in fact that the above approach is correct and that neglecting feasibility constraints in the characterization of the non-manipulable mechanisms is without loss of generality as long as the probability that an agent wins increases with his opponent's type.

To get a simple design of the optimal mechanism, we impose also:

Assumption 5 *Best-Predictor Property (BPP):*

$$\tilde{f}_{\theta_i}(\theta_i|\theta_i) = 0.$$

[24]Omitting the index i of the functions $h_i(\cdot)$ because of symmetry.

Given the report made by A_{-i}, the most likely type for A_i is this report itself. This property is useful to derive some of the results below, noticeably on the efficiency of the optimal auction.

Proposition 6 *: Assume that the buyer (principal) wants at most one unit of the good, and that Assumptions 1, 2, 3, 4 and 5 hold altogether. Then, the optimal non-manipulable auction mechanism is such that:*

$$q_i^{SB}(\theta_i, \theta_{-i}) = 1 \quad \Leftrightarrow \quad \theta_i \leq \min(\theta_{-i}, \theta^*), \tag{24}$$

where θ^* is implicitly defined as the unique solution to:

$$S = \theta^* + \frac{F(\theta^*)}{f(\theta^*)}. \tag{25}$$

Two points are worth to be stressed. First, whenever the good is produced at all, it is produced by the most efficient agent. Nevertheless, the set $[\underline{\theta}, \theta^*]$ of types who produce with some probability is affected by the non-manipulability constraint as it can be seen from comparing (25) with the full information condition which is obviously less stringent.

Second, the non-manipulable optimal auction is quite simple. From (22) and the optimal allocation derived in (24), there exists some function $h^{SB}(\cdot)$ (explicit in equation (28) below) such that agent A_i's payment is:

$$t^{SB}(\theta_i, \theta_{-i}) = \begin{cases} S - h^{SB}(\theta_i) & \text{if } \theta_i \leq \min(\theta_{-i}, \theta^*), \\ -h^{SB}(\theta_i) & \text{otherwise.} \end{cases} \tag{26}$$

This all-pay auction works as follows. All agents choose an entry fee within a proposed menu $\{h^{SB}(\hat{\theta})\}_{\hat{\theta} \in \Theta}$. The agent who pays the highest entry fee gets the right to produce and receives from the principal a bonus S equal to his own valuation for the good. The other agent does not produce.

This mechanism is clearly non-manipulable. Since the winning agent is made residual claimant for the decision to produce or not, the principal loses all possibilities for manipulating rewards or production for that agent. Moreover, the principal finds also no gain in manipulating the mechanism so that the agent with the lowest bid (i.e. lowest entry fee) instead produces. Indeed, this agent is the least efficient one and produces the good less often, reducing the entry fee he is ready to pay.

The expression for the entry fee $h^{SB}(\theta_i)$ is obtained from using the Bayesian incentive compatibility constraint and a boundary condition that states that an agent with type θ^* gets zero information rent. Denoting by $U^{SB}(\theta_i)$ the (symmetric) information rent at the optimal mechanism, and taking into account that an agent produces only if his marginal cost is both less than the threshold θ^* and the other agent's cost, we have:

$$U^{SB}(\theta_i) = \arg\max_{\hat{\theta}_i \in [\underline{\theta}, \theta^*]} (S - \theta_i)(1 - \tilde{F}(\hat{\theta}_i|\theta_i)) - h^{SB}(\hat{\theta}_i). \tag{27}$$

From which, we get:

$$\dot{h}^{SB}(\theta_i) = -(S - \theta_i)\tilde{f}(\theta_i|\theta_i) < 0 \text{ for } \theta_i \leq \theta^*.$$

Integrating and taking into account that $U^{SB}(\theta^*) = 0$, we obtain:

$$h^{SB}(\theta_i) = \begin{cases} (S - \theta^*)(1 - \tilde{F}(\theta^*|\theta^*)) - \int_{\theta_i}^{\theta^*} (S - \theta)\tilde{f}(\theta|\theta) d\theta & \forall \theta_i \leq \theta^* \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

The optimal non-manipulable auction is an all-pay auction with reserve price $h^{SB}(\theta^{*-}) = (S - \theta^*)(1 - \tilde{F}(\theta^*|\theta^*))$. If an agent is ready to pay at least this amount, he will participate to the auction. Otherwise, he does not even participate.

Note that, as the correlation vanishes, non-manipulability constraints have again no bite. This all-pay auction converges towards the optimal allocation that would be obtained without taking into those constraints. Simple integration by parts show that the entry-fee defined in (28) converges indeed towards

$$h^{SB}(\theta_i) = \begin{cases} (S - \theta_i)(1 - F(\theta_i)) - \int_{\theta_i}^{\theta^*} (1 - F(\theta)) d\theta & \forall \theta_i \leq \theta^* \\ 0 & \text{otherwise} \end{cases}$$

which yields to an agent with type $\theta_i \leq \theta^*$ his well-known rent at the optimal mechanism

$$U(\theta_i) = \int_{\theta_i}^{\theta^*} (1 - F(\theta)) d\theta.$$

At zero correlation, the all-pay and the first-price auctions with the same reserve price achieve the same outcome. Both formats allocate the right to produce to the most efficient agent provided his type is lower than θ^*. With some correlation, both mechanisms are non-manipulable and still allocate the right to produce to the same seller. However, the principal's revenue (resp. the agents' rent) is higher (resp. lower) with an all-pay format.

To understand this ranking, it is useful to come back to the expression of the agents' rent in the all-pay auction given by (27) which, by the Envelope Theorem, gives:

$$\dot{U}^{SB}(\theta_i) = -(1 - \tilde{F}(\theta_i|\theta_i)) - (S - \theta_i) \int_{\underline{\theta}}^{\theta_i} \tilde{f}_{\theta_i}(\theta|\theta_i) d\theta. \tag{29}$$

Consider now a first-price auction with reserve price θ^* where bidders adopt the symmetric truthful strategy $b(\cdot)$. The agents' rent in such a Bayesian game is

$$U^1(\theta_i) = \arg\max_{\hat{\theta}_i \in [\underline{\theta}, \theta^*]} (b(\hat{\theta}_i) - \theta_i)(1 - \tilde{F}(\hat{\theta}_i|\theta_i))$$

whose derivative is

$$\dot{U}^1(\theta_i) = -(1 - \tilde{F}(\theta_i|\theta_i)) - (b(\theta_i) - \theta_i) \int_{\underline{\theta}}^{\theta_i} \tilde{f}_{\theta_i}(\theta|\theta_i) d\theta. \tag{30}$$

In both cases, the slope of the agents' information rent is the sum of two terms, one positive, the other being negative from Assumptions 4 and 5. The first (positive) source of rent comes from the fact that, conditionally on winning, an agent with type θ_i gains some information rent from his ability to exaggerate his type and producing at a lower cost than what he announces. The second (negative) source of rent comes from the fact that, as he exaggerates his type, this agent sends "good news" on the other and, this reduces by as much the probability of being awarded the right to produce. In the first price auction, the corresponding benefit is lower than in the all-pay auction since in the latter what the agent loses is the whole social value of producing whereas this is only the private profit he makes in the first-price auction. Both formats are subject to some form of countervailing incentives which reduce rent[25] but the more pronounced ones are for the all-pay auction.

8 A General Approach

The previous section has highlighted the difficulties faced when dealing with non-manipulability constraints directly especially when those constraints are not separable. Although, we were guided by intuition to conjecture the form of the non-manipulable nonlinear prices, it is time to propose a more general approach that enable us to derive second-best distortions in more general environments.

Using again the Taxation Principle derived in Proposition 2, non-manipulability constraints can generally be written as:

$$(q_1(\theta_1, \theta_2), q_2(\theta_1, \theta_2)) \in \arg \max_{(q_1, q_2) \in \mathcal{Q}} \tilde{S}(q_1, q_2) - \sum_{i=1}^{2} T_i(q_i, \theta_i). \tag{31}$$

This formulation is attractive since the optimality conditions above look very much like an incentive compatibility constraint on the principal's side. Keeping q_{-i} as fixed, the optimality condition satisfied by q_i is the same as that one should write to induce this principal to publicly reveal his private information q_{-i}. This remark being made, one can proceed as usual in mechanism design and characterize direct revelation mechanisms $\{t_i(\hat{q}_{-i}|\theta_i); q_i(\hat{q}_{-i}|\theta_i)\}_{\hat{q}_{-i} \in \mathcal{Q}}$ which induce truthful revelation of the parameter $\hat{q}_{-i}$. Of course, this parameter is not exogenously given as in standard adverse selection problem, but is derived endogenously from the equilibrium behavior. Starting then from such

[25]Lewis and Sappington (1989).

direct revelation mechanism, we can then use standard techniques and reconstruct a non-manipulable nonlinear price $T_i(q_i, \theta_i)$ by simply "eliminating" $\hat{q}_{-i}$ from the expressions obtained for $t_i(\hat{q}_{-i}|\theta_i)$ and $q_i(\hat{q}_{-i}|\theta_i)$.

Lemma 1 *The direct revelation mechanism* $\{t_i(\hat{q}_{-i}|\theta_i); q_i(\hat{q}_{-i}|\theta_i)\}_{\hat{q}_{-i}\in\mathcal{Q}}$ *associated to a non-manipulable nonlinear price* $T_i(q_i, \theta_i)$ *is such that:*

- $q_i(q_{-i}|\theta_i)$ *is monotonically increasing (resp. decreasing) in* q_{-i} *and thus a.e. differentiable when the agents' efforts are complements, i.e.,* $\frac{\partial^2 \tilde{S}}{\partial q_1 \partial q_2} > 0$*, (resp. substitutes, i.e.,* $\frac{\partial^2 \tilde{S}}{\partial q_1 \partial q_2} < 0$*).*
- $t_i(q_{-i}|\theta_i)$ *is a.e. differentiable in* q_{-i} *with*

$$\frac{\partial t_i(q_{-i}|\theta_i)}{\partial q_{-i}} = \frac{\partial \tilde{S}}{\partial q_i}(q_i(q_{-i}|\theta_i), q_{-i})\frac{\partial q_i(q_{-i}|\theta_i)}{\partial q_{-i}}. \tag{32}$$

- *Consider any differentiability point where* $\frac{\partial q_i(q_{-i}|\theta_i)}{\partial q_{-i}} \neq 0$ *and denote* $\tilde{q}_{-i}(q_i, \theta_i)$ *the inverse function of* $q_i(q_{-i}|\theta_i)$*. The non-manipulable nonlinear price* $T_i(q_i, \theta_i)$ *is differentiable at such point and its derivative satisfies:*

$$\frac{\partial T_i(q_i, \theta_i)}{\partial q_i} = \frac{\partial \tilde{S}}{\partial q_i}(q_i, \tilde{q}_{-i}(q_i, \theta_i)). \tag{33}$$

It is now necessary to evaluate $\tilde{q}_{-i}(q_i, \theta_i)$ and the nonlinear prices $\{T_1(q_1, \theta_1), T_2(q_2, \theta_2)\}$ as functions of the vector of *anticipated* equilibrium quantities $q^e(\theta)$ that those nonlinear prices induce. Restricting the analysis to allocations such that $q_i^e(\theta_i, \theta_{-i})$ is strictly monotonic in θ_{-i},[26] we can unambiguously define the function $\phi_{-i}(\theta_i, q_i)$ such that:

$$q_i = q_i^e(\theta_i, \phi_{-i}(\theta_i, q_i)).$$

Then, we have:

$$\tilde{q}_{-i}(q_i, \theta_i) = q_{-i}^e(\theta_i, \phi_{-i}(\theta_i, q_i)).$$

Finally, (33) yields the expression of $T_i(q_i, \theta_i)$ as:

$$T_i(q_i, \theta_i) = \int_0^{q_i} \frac{\partial \tilde{S}}{\partial x}(x, q_{-i}^e(\theta_i, \phi_{-i}(\theta_i, x)))dx - H_i(\theta_i) \tag{34}$$

where $H_i(\theta_i)$ is some arbitrary function.

This formula is rather general and can be used to recover some polar cases:

[26] The reader will have noticed that such restriction is of the same kind as those we imposed when deriving the optimal auction.

- *Unrelated Projects:* This case is straightforward. $\frac{\partial \tilde{S}}{\partial x}(x, q^e_{-i}(\theta_i, \phi_{-i}(\theta_i, x)))$ is of the form $\frac{\partial \tilde{S}}{\partial x}(x)$. Direct integration of (34) yields (9).

- *Perfect Substitutability:* Suppose that $\tilde{S}(q_1, q_2) = S(q_1 + q_2)$ so that the agents' outputs are perfect substitutes as in the case of the multi-unit auction. Then, conjecturing that $q^e_i(\theta_i, \theta_{-i})$ is monotonically increasing in θ_{-i}, we have $q^e_{-i}(\theta_i, \phi_{-i}(\theta_i, x)) = 0$ for all $x > 0$. Nonlinear prices are given by:

$$T_i(q_i, \theta_i) = \int_0^{q_i} \frac{\partial \tilde{S}}{\partial x}(x + q^e_{-i}(\theta_i, \phi_{-i}(\theta_i, x)))dx - H_i(\theta_i) \quad (35)$$

  where $H_i(\theta_i)$ is some arbitrary function.

Using (34), we can also express the agent's incentive compatibility constraint as follows:

$$U_i(\theta_i) = \arg\max_{\hat{\theta}_i \in \Theta} \underset{\theta_{-i}}{E} \left(\int_0^{q_i(\hat{\theta}_i, \theta_{-i})} \frac{\partial \tilde{S}}{\partial x}(x, q^e_{-i}(\hat{\theta}_i, \phi_{-i}(\hat{\theta}_i, x)))dx - \theta_i q_i(\hat{\theta}_i, \theta_{-i}) | \theta_i \right) - H_i(\hat{\theta}_i). \quad (36)$$

From this we can derive the optimal second-best distortions:

Proposition 7 *Assume that Assumptions 1, 2 and 5 hold altogether and that $\tilde{S}(\cdot)$ is strictly concave in (q_1, q_2). Then, the optimal non-manipulable mechanism entails outputs such that:*

$$\frac{\partial \tilde{S}}{\partial q_i}(q^{SB}_i(\theta_i, \theta_{-i}), q^{SB}_{-i}(\theta_i, \theta_{-i})) = \varphi(\theta_i, \theta_{-i}), \quad (37)$$

provided $q^{SB}_i(\theta_i, \theta_{-i})$ is non-increasing in θ_i and non-decreasing (resp. non-increasing) in θ_{-i} if outputs are substitutes (resp. complements) as requested by Lemma 1.

This is again a generalized Baron-Myerson formula. The marginal benefit of one activity is equal to its generalized virtual cost defined in Assumption 1.

For further references, it is useful to discuss two examples in more details.

8.1 The Limiting Case of Perfect Complementarity

Consider the following surplus function for the principal:

$$\tilde{S}(q_1, q_2) = \mu(q_1 + q_2) - \frac{q_1^2}{2} - \frac{q_2^2}{2} - \lambda(q_1 - q_2)^2$$

for some parameter $\mu > 0$ and $\lambda > 0$ so that optimal outputs remain non-negative. Using (37) above, it is straightforward to check that, in the limit of λ very large, i.e., for almost perfect complements, both agents produce the same output given by:

$$q^{SB}(\theta_i, \theta_{-i}) = \mu - \frac{\varphi(\theta_i, \theta_{-i}) + \varphi(\theta_{-i}, \theta_i)}{2}. \quad (38)$$

The marginal benefit of the production is equal to the sum of the agents' generalized virtual costs. We will recover this case in more details in Section 9 below. There, we will give up this limit argument and rely on a more direct analysis still relying on the general intuition that what the nonlinear price between P and A_i does is inducing information revelation on the contract signed with A_{-i}.

8.2 Multi-Unit Auctions

The multi-unit auction framework raises new issues coming from the fact that the principal's objective is no longer strictly concave in (q_1, q_2). The principal's gross surplus from consuming q units of the good can be written as $S(q)$ where $S'(0) = +\infty$, $S' > 0$, $S'' < 0$ and $S(0) = 0$.[27]

Proposition 8 *: Assume that Assumptions 1, 2, 3, 4 and 5 hold altogether. Then, the optimal non-manipulable multi-unit auction mechanism entails:*

- *The most efficient agent always produces all output*

$$q_i^{SB}(\theta_i, \theta_{-i}) = q^{SB}(\theta_i, \theta_{-i}) > 0 \quad \Leftrightarrow \quad \theta_i \leq \theta_{-i} \tag{39}$$

where output is given by the modified "Baron-Myerson" formula

$$S'(q^{SB}(\theta_i, \theta_{-i})) = \varphi(\theta_i, \theta_{-i}) \quad \textit{for } \theta_{-i} \geq \theta_i \tag{40}$$

and $q^{SB}(\theta_i, \theta_i) = q^{BM}(\theta_i)$*;*

- *The optimal nonlinear price is defined as*

$$T^{SB}(q, \theta_i) = S(q) - h^{SB}(\theta_i), \quad \forall q \in \ \textit{range}(q^{SB}(\theta_i, \cdot)),\ \forall \theta_i \in \Theta \tag{41}$$

where

$$h^{SB}(\theta_i) = E_{\theta_{-i}}(S(q^{SB}(\theta_i, \theta_{-i})) - \theta_i q^{SB}(\theta_i, \theta_{-i}))|\theta_i)$$
$$- \int_{\theta_i}^{\bar{\theta}} E_{\theta_{-i}} \left(q^{SB}(x, \theta_{-i}) - (S(q^{SB}(x, \theta_{-i})) - \theta_i q^{SB}(x, \theta_{-i})) \frac{\tilde{f}_{\theta_i}(\theta_{-i}|x)}{\tilde{f}(\theta_{-i}|x)} |x \right) dx.$$

The optimal multi-unit auction shares several features with the single unit auction. Again, the right to produce is allocated to the most efficient agent. The principal offers a menu of (symmetric) nonlinear schedules and let agents pick their most preferred choices. The one

[27] The Inada condition ensures that it is always optimal to induce a positive production even in the second-best environment that we consider so that the issue of finding an upper bound on the set of types who may actually produce no longer arises.

having revealed the lowest cost parameter gets the whole market and produces accordingly. Moreover, these nonlinear schedules are designed so that each agent is actually made residual claimant for the principal's surplus function conditionally on winning the auction.

Conditionally on this event, this agent produces an output which is nevertheless modified to take into account what the principal learns from the other agent's report. However, output distortions are always lesser than in the Baron-Myerson outcome. Indeed, the mere fact that an agent wins the auction conveys only "good news" to the principal; the other agent's cost parameter is always greater.

9 Team Production

Let us now consider the case where agents exert efforts for an organization and assume that those efforts are *Leontieff perfect complements*. We will denote $q = \min(q_1, q_2)$ the organization's output and by $S(q)$ the principal's benefit from producing q units of output. Again, we assume that $S'(0) = +\infty,\ S' > 0,\ S'' < 0$ with $S(0) = 0$.

As a benchmark, consider the case where types are independently distributed. Then, both agents produce the same amount and the marginal benefit of such production $q^{LS}(\theta_1, \theta_2)$ must be traded off against the sum of the agents' virtual costs of effort:

$$S'(q^{LS}(\theta_1, \theta_2)) = \sum_{i=1}^{2} \theta_i + \frac{F(\theta_i)}{f(\theta_i)}. \tag{42}$$

More generally, for a given (symmetric) output schedule $q(\cdot)$ offered to the agents, we rewrite the non-manipulability constraints (7) as:

$$(\theta_1, \theta_2) \in \arg \max_{(\hat{\theta}_1, \hat{\theta}_2) \in \Theta^2} S\left(\min(q_1(\theta_1, \hat{\theta}_2), q_2(\hat{\theta}_1, \theta_2))\right) - \sum_{i=1}^{2} t_i(\theta_i, \hat{\theta}_{-i}). \tag{43}$$

In this context with perfect complementarity, it is quite natural to look for incentive schemes such that both the agents' efforts and payments are non-decreasing in both types. Then, the principal has no incentives to lie *"excessively"* to either agent and

$$q_1(\theta_1, \hat{\theta}_2) = q_2(\hat{\theta}_1, \theta_2).$$

Alternatively, imposing such an equality can be justified when agents observe the output of the organization. Using the *modified common agency* formulation proposed in Proposition 2, we observe that the output $q(\theta_1, \theta_2) = q_1(\theta_1, \theta_2) = q_2(\theta_1, \theta_2)$ should now solve:

$$q(\theta_1, \theta_2) \in \arg \max_{q} S(q) - \sum_{i=1}^{2} T_i(q, \theta_i). \tag{44}$$

Again, the non-separability of the non-manipulability constraint raises some technical issues. To solve the problem, we slightly depart from previous observability assumptions and consider the case where agent A_i does not observe $T_{-i}(q, \theta_{-i})$, the menu offered to agent A_{-i}. Guided by the intuition built in Section 8, the nonlinear prices $T_i(q, \theta_i)$ can still be recovered from a direct revelation mechanism that would now induce the principal to reveal everything not known by A_i. Given that both agents' outputs end up being observed by A_i, this information unknown to A_i amounts only to A_{-i}'s type.

When designing a nonlinear schedule $T_i(q, \theta_i)$ to extract the principal's information on A_{-i}, one must take into account that A_i forms conjectures on the equilibrium output $q^e(\theta_1, \theta_2)$ and the nonlinear price $T^e_{-i}(q, \theta_{-i})$ offered to A_{-i} (which is also non-observable by A_i, here). Let define $\phi_{-i}(\theta_i, q)$ such that:

$$q = q^e(\theta_i, \phi_{-i}(\theta_i, q)).$$

Inducing truthtelling from the principal on A_{-i}'s type requires to use a nonlinear price $T_i(q, \theta_i)$ which solves:

$$T_i(q, \theta_i) = \int_0^q \left(S'(x) - \frac{\partial T^e_{-i}}{\partial q}(x, \phi_{-i}(\theta_i, x)) \right) dx - H(\theta_i) \tag{45}$$

where $H(\theta_i)$ is some arbitrary function.

The class of mechanisms satisfying (45) is non-empty. It is routine to show that separable nonlinear prices $T(q, \theta)$ of the form

$$T(q, \theta) = \frac{1}{2} S(q) - h(\theta), \tag{46}$$

where $h(\cdot)$ is some arbitrary function belong to that class.

To characterize the optimal non-manipulable mechanism,[28] we proceed as in the previous sections:

Proposition 9 *: When Assumptions 1 and 2 hold, there exists an optimal non-manipulable mechanism with perfect complements which entails:*

- *A symmetric output $q^{SB}(\theta_1, \theta_2)$ such that:*

$$S'(q^{SB}(\theta_1, \theta_2)) = \sum_{i=1}^{2} \varphi(\theta_i, \theta_{-i}), \tag{47}$$

as long as $q^{SB}(\theta_1, \theta_2)$ is decreasing in both arguments;

[28] As usual in the literature on secret contract offers, we are assuming passive beliefs out of the equilibrium path, see Segal (1999) for instance

- *The marginal payment to A_i is equal to his generalized virtual cost parameter*

$$\frac{\partial T_i}{\partial q}(q(\theta_i, \theta_{-i}), \theta_i) = \varphi(\theta_i, \theta_{-i}) \tag{48}$$

which is non-decreasing in θ_i and non-increasing in θ_{-i}.

- *Both agents always get a positive information rent except when inefficient*

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

The logic of the argument here is very similar to that made earlier although the output distortions differ somewhat due to the specificities of the team production problem. Non-manipulability constraints require that each agent's payment make him somewhat internalize the principal's objective function. Because of the team problem, each agent can only partially internalize the principal's objective and his marginal payment is only a fraction of the principal's marginal benefit of production. Equation (48) shows that the marginal reward to agent A_i decreases as A_{-i} (resp. A_i) becomes less (resp. more) efficient. As a consequence, splitting in half the production surplus between the agents yields with the nonlinear prices given in (46) cannot be optimal.

In this team production framework, the output distortions necessary to reduce both agents' information rents must be compounded as it can be seen on (47) which generalizes the limiting case found on a particular example in (38). Also using (47), we observe that the optimal output converges towards the solution (42) as correlation diminishes.

10 Conclusion

This paper has investigated the consequences of relaxing the assumption of public communication in an otherwise standard mechanism design environment. Doing so paves the way to a tractable theory which responds to some of the most often heard criticisms addressed to the mechanism design methodology.

Each of the particular settings we have analyzed (auctions, team production, more general production externalities) deserves some further studies either by specializing the information structure, by generalizing preferences or by focusing on organizational problems coming from the analysis of real world institutions in particular contexts (political economy, regulation, etc..).

Of particular importance may be the extension of our framework to the case of auctions with interdependent valuations and common values. Our approach for simplifying mechanisms could be an alternative to the somewhat too demanding ex post implementation requested by the recent vintage of the literature on those topics.

The analysis of public good mechanisms and trading mechanisms in correlated environments deserves further analysis.

The introduction of 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 modelling of the vertical collusion and favoritism that may take place in those environments. Second, this congruence may introduce interesting aspects related to the common values element that arises in such environment.

Also, it would be worth to investigate what is the scope for horizontal collusion between the agents in the environments depicted in this paper. Indeed, since an agent's output and information rents still depend on what other claims, there is scope for collusion in Bayesian environments whereas relying on dominant and non-manipulable mechanisms destroys this possibility.

In practice, the degree of transparency of communication in an organization may be intermediate between what we have assumed here and the more usual postulate of public communication. We conjecture that reputation-like arguments on the principal's side may help in circumventing non-manipulability constraints but the extent by which it is so remains to uncover.

All those are extensions that we plan to analyze in further research.

References

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.

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.

Che, Y.-K. and J. Kim (2006): "Robustly Collusion-Proof Implementation," *Econometrica*, 74, 1063-1108.

Chung, K.S. and J. Ely (2005): "Foundations for Dominant Strategy Mechanisms," *forthcoming Review of Economic Studies.*

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

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

Compte, O. and P. Jehiel (2006): "Veto Constraints in Mechanism Design: Inefficiency with Correlated Types," mimeo PSE CERAS.

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

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

Dasgupta, P. and E. Maskin (2000): "Efficient Auctions," *Quarterly Journal of Economics*, 115, 341-388.

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.

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.

Gromb, D. and D. Martimort (2005): "Collusion and the Organization of Delegated Expertise," mimeo IDEI Toulouse.

Hart, O. and J. Tirole (1990): "Vertical Integration and Market Foreclosure," *Brooking Papers in Economic Activity*, 205-285.

Heifetz, A. and Z. Neeman (2006): "On the Generic (Im-)possibility of Surplus Full Extraction in Mechanism Design," *Econometrica*, 74, 213-233.

Jehiel, P., M. Meyer-Ter-Vehn, B. Moldovanu, and W. Zame (2006): "The Limits of Ex-Post Implementation," *Econometrica*, 74, 585-610.

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

Laffont, J.-J. and D. Martimort (2000): "Mechanism Design with Collusion and Correlation," *Econometrica*, 68, 309-342.

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

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.

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

Martimort, D. and H. Moreira (2005): "Common Agency with Informed Principals," Mimeo IDEI Toulouse.

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

— (2003a): "Contractual Externalities and Common Agency Equilibria," *Advances in Theoretical Economics*, Vol. 3, Issue 1, Article 4. http://www.bepress.com/bejte.

— (2005): "On the Robustness of Truthful Equilibria in Common Agency Games," Mimeo IDEI and University of Chicago.

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.

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

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

— (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.

Perry, M. and P. Reny (2002): "An Ex Post Efficient Auction," *Econometrica*, 70, 1199-1212.

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

— (2003): "Negotiation and Take-It-Or-Leave-It in Common Agency," *Journal of Economic Theory*, 111, 88-109.

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

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.

Segal, I. and M. Whinston (2003): "Robust Predictions for Bilateral Contracting with Externalities," *Econometrica*, 71, 757-792.

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

Appendix A

• **Proof of Proposition 1:** Take any arbitrary mechanism $(g(\cdot), \mathcal{M}) = ((g_1(\cdot), \mathcal{M}_1), (g_2(\cdot), \mathcal{M}_2))$ for an arbitrary communication space $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$. 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 θ_i reports a private message $m_i^*(\theta_i)$ to the principal. The strategy $m^*(\theta) = (m_1^*(\theta_1), m_2^*(\theta_2))$ forms a Bayesian-Nash equilibrium between the agents. We make explicit the corresponding equilibrium conditions below.

- 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 publicly reveals the messages $\hat{m}_1(m)$ and $\hat{m}_2(m)$ which maximizes his expected payoff, i.e.,

$$(\hat{m}_1(m), \hat{m}_2(m))$$
$$\in \arg \max_{(\hat{m}_1,\hat{m}_2)\in\mathcal{M}} \int_{\Theta^2} \left\{ \tilde{S}(q_1(m_1, \hat{m}_2)), q_2(\hat{m}_1, m_2) - \sum_{i=1}^{2} t_i(m_i, \hat{m}_{-i}) \right\} d\mu^*(\theta|m). \tag{A.1}$$

  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:

$$(\hat{m}_1^*(m), \hat{m}_2^*(m)) \in \arg \max_{(\hat{m}_1,\hat{m}_2)\in\mathcal{M}} \tilde{S}(q_1(m_1, \hat{m}_2), q_2(\hat{m}_1, m_2)) - \sum_{i=1}^{2} t_i(m_i, \hat{m}_{-i}). \tag{A.2}$$

Let us turn now to the agents' Bayesian incentive compatibility conditions that must be satisfied by $m^*(\cdot)$. For A_i, we have for instance

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

Our proof of a Revelation Principle will now proceeds in two steps. In the first one, we replace the general mechanism $(g(\cdot), \mathcal{M})$ by another general 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 and truthful mechanism $(\bar{g}(\cdot), \Theta)$.

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

$$\tilde{t}_i(m_i, m_{-i}) = t_i(m_i, \hat{m}_i^*(m_i, m_{-i})) \text{ and } \tilde{q}_i(m_i, m_{-i}) = q_i(m_i, \hat{m}_{-i}^*(m_i, m_{-i})) \text{ for } i = 1, 2. \tag{A.3}$$

Lemma 2 *: $(\tilde{g}(\cdot), \mathcal{M})$ 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 $m = (m_1, m_2) \in \mathcal{M}$. By (A.1), we have:

$$\tilde{S}(q_1(m_1, \hat{m}_2^*(m)), q_2(m_2, \hat{m}_1^*(m))) - t_1(m_1, \hat{m}_2^*(m))$$
$$\geq \tilde{S}(q_1(m_1, \tilde{m}_2), q_2(m_2, \hat{m}_1^*(m))) - t_1(m_1, \tilde{m}_2) \quad \forall \tilde{m}_2 \in \mathcal{M}_2.$$

Using the definition of $\tilde{g}(\cdot)$ given in (A.3), we get:

$$\tilde{S}(\tilde{q}_1(m_1,m_2),\tilde{q}_2(m_2,m_1)) - \tilde{t}_1(m_1,m_2)$$

$$\geq \tilde{S}(q_1(m_1,\hat{m}_2^*(m')),\tilde{q}_2(m_2,m_1)) - t_1(m_1,\hat{m}_2^*(m')) \quad \forall m' = (m_1,m_2') \in \mathcal{M}. \qquad \text{(A.4)}$$

Then (A.4) becomes:

$$\tilde{S}(\tilde{q}_1(m_1,m_2),\tilde{q}_2(m_2,m_1)) - \tilde{t}_1(m_1,m_2) \qquad \text{(A.5)}$$

$$\geq \tilde{S}(\tilde{q}_1(m_1,m_2'),\tilde{q}_2(m_2,m_1)) - \tilde{t}_1(m_1,m_2') \quad \forall m_2' \in \mathcal{M}_2. \qquad \text{(A.6)}$$

Given that $\tilde{g}(\cdot)$ is played, the best manipulation made by the principal is $\hat{m}_2^*(m) = m$ 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 $m^*(\theta) = (m_1^*(\theta_1), m_2^*(\theta_2))$ for the agents. Indeed, $m^*(\cdot)$ satisfies by definition the following Bayesian-Nash constraint:

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

which can be rewritten as:

$$m_i^*(\theta_i) \in \arg\max_{m_i} \underset{\theta_{-i}}{E} \left(\tilde{t}_i(m_i,m_{-i}^*(\theta_{-i})) - \theta_i q_i(m_i,m_{-i}^*(\theta_{-i}))|\theta_i\right). \qquad \text{(A.7)}$$

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

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

$$\bar{t}_i(\theta) = \tilde{t}_i(m^*(\theta)) \text{ and } \bar{q}_i(\theta) = \tilde{q}_i(m^*(\theta)) \quad \text{for } i = 1,2. \qquad \text{(A.8)}$$

Lemma 3 : $\bar{g}(\cdot)$ *is truthful in Bayesian incentive compatibility and not manipulable.*

Proof: First consider the non-manipulability of the mechanism $\bar{g}(\cdot)$. From (A.6), we get:

$$\tilde{S}\left(\bar{q}_1(\theta_1,\theta_2),\bar{q}_2(\theta_1,\theta_2)\right) - \bar{t}_1(\theta_1,\theta_2)$$

$$\geq \tilde{S}\left(\tilde{q}_1(m_1^*(\theta_1),m_2'),\tilde{q}_2(m_1^*(\theta_1),m_2^*(\theta_2))\right) - \tilde{t}_1(m_1^*(\theta_1),m_2') \quad \forall m_2' \in \mathcal{M}_2. \qquad \text{(A.9)}$$

Taking $m_2' = m_2^*(\theta_2')$, (A.9) becomes

$$\tilde{S}\left(\bar{q}_1(\theta_1,\theta_2),\bar{q}_2(\theta_1,\theta_2)\right) - \bar{t}_1(\theta_1,\theta_2) \geq \tilde{S}(\bar{q}_1(\theta_1,\theta_2'),\bar{q}_2(\theta_1,\theta_2)) - \bar{t}_1(\theta_1,\theta_2') \quad \forall m_2' \in \Theta. \quad \text{(A.10)}$$

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

Turning to (A.7), it is immediate to check that the agents' Bayesian incentive constraints can be written as:

$$\theta_i \in \arg\max_{\hat{\theta}_i} \underset{\theta_{-i}}{E} \left(\bar{t}_i(\hat{\theta}_i,\theta_{-i}) - \theta_i \bar{q}_i(\hat{\theta}_i,\theta_{-i})|\theta_i\right). \qquad \text{(A.11)}$$

∎

• **Proof of Proposition 2:** Let consider the non-manipulability constraint (7) and define the nonlinear price $T_i(\hat{q}_i, \theta_i)$ as $T_i(\hat{q}_i, \theta_i) = t_i(\theta_i, \theta_{-i})$ for $\hat{q}_i$ such that $\hat{q}_i = q_i(\theta_i, \theta_{-i})$. $T_i(\cdot, \theta_i)$ is defined over the range of $q_i(\theta_i, \cdot)$ that we denote $range(q_i(\theta_i, \cdot))$. For any $\hat{q}_i \in range(q_i(\theta_i, \cdot)), \hat{q}_{-i} \in range(q_{-i}(\theta_{-i}, \cdot))$, we find

$$\tilde{S}(q_i(\theta_i, \theta_{-i}), q_{-i}(\theta_i, \theta_{-i})) - \sum_{i=1}^{2} T_i(q_i(\theta_i, \theta_{-i}), \theta_i) \geq \tilde{S}(\hat{q}_i, \hat{q}_{-i}) - \sum_{i=1}^{2} T_i(\hat{q}_i, \theta_i), \qquad \text{(A.12)}$$

which amounts to

$$(q_i(\theta), q_{-i}(\theta)) \in \arg \max_{(q_1,q_2)\in\mathcal{D}(\theta)} \tilde{S}(q_1, q_2) - \sum_{i=1}^{2} T_i(q_i, \theta_i), \qquad \text{(A.13)}$$

where $\mathcal{D}(\theta) = \prod_{i=1}^{2} range(q_i(\theta_i, \cdot))$. (A.13) is an optimality condition for the principal.

It is straightforward to check the agents' Bayesian incentive compatibility constraints:

$$\theta_i \in \arg\max_{\hat{\theta}_i} \underset{\theta_{-i}}{E} \left(T_i(q_i(\hat{\theta}_i, \theta_{-i}), \hat{\theta}_i) - \theta_i q_i(\hat{\theta}_i, \theta_{-i}) | \theta_i \right). \qquad \text{(A.14)}$$

The modified common agency game $\{T_i(q_i, \hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta}$ is thus Bayesian incentive compatible.

Conversely, consider any equilibrium quantities $(q_i(\theta), q_{-i}(\theta))$ of the modified common agency game and the nonlinear prices $T_i(q_i, \theta_i)$ that sustain this equilibrium. These nonlinear prices satisfy equations (A.13) and (A.14). Define a direct mechanism with transfers $t_i(\theta_i, \theta_{-i}) = T_i(q_i(\theta_i, \theta_{-i}), \theta_i)$ and outputs $q_i(\theta_i, \theta_{-i})$. Equation (A.13) implies

$$\tilde{S}(q_1(\theta_1, \theta_2), q_1(\theta_1, \theta_2)) - t_1(\theta_1, \theta_2) - t_2(\theta_1, \theta_2)$$

$$\geq \tilde{S}(q_1(\theta_1, \hat{\theta}_2), q_2(\hat{\theta}_1, \theta_2)) - t_1(\theta_1, \hat{\theta}_2) - t_2(\hat{\theta}_1, \theta_2), \qquad \forall (\theta_1, \theta_2, \hat{\theta}_1, \hat{\theta}_2) \in \Theta^4. \qquad \text{(A.15)}$$

Hence, this direct mechanism is non-manipulable.

Equation (A.14) implies

$$\underset{\theta_{-i}}{E} \left(t_i(\theta_i, \theta_{-i}) - \theta_i q_i(\theta_i, \theta_{-i}) | \theta_i \right) \geq \underset{\theta_{-i}}{E} \left(t_i(\hat{\theta}_i, \theta_{-i}) - \theta_i q_i(\hat{\theta}_i, \theta_{-i}) | \theta_i \right) \quad \forall (\theta_i, \hat{\theta}_i) \in \Theta^2, \qquad \text{(A.16)}$$

which ensures Bayesian incentive compatibility. ∎

• **Proof of Proposition 3:** First, let us suppose that (11) is binding only at $\theta_i = \bar{\theta}$. Integrating (14) , we get

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

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

$$\underset{\theta_i}{E}(U_i(\theta_i)) = U_i(\bar{\theta}) + \underset{(\theta_{-i},\theta_i)}{E}\left(\frac{F(\theta_i)}{f(\theta_i)}\left(q_i(\theta_i,\theta_{-i}) - (S(q_i(\theta_i,\theta_{-i})) - \theta_i q_i(\theta_i,\theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\right)\right). \tag{A.17}$$

Of course minimizing the agents' information rent requires to set $U_i(\bar{\theta}) = 0$ when the right-hand side in (14) is negative; something that will be checked later. Inserting (A.17) into the principal's objective function and optimizing pointwise yields (16).

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

Monotonicity of $U_i(\theta_i)$: From Assumption 2 ($\tilde{f}_\theta$ is small enough), the second term on the right-hand side of (14) is small relative to the first one and $U_i(\cdot)$ is strictly decreasing.

Second-order conditions: Let us come back to condition (15). For $q^{SB}(\theta_i,\theta_{-i})$ this condition becomes

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

which obviously holds under the assumptions of Proposition 3.

Global incentive compatibility: The global incentive compatibility condition writes as:

$$U_i(\theta_i) \geq U_i(\hat{\theta}_i) + \underset{\theta_{-i}}{E}\left(S(q_i(\hat{\theta}_i,\theta_{-i})) - \theta_i q_i(\hat{\theta}_i,\theta_{-i})|\theta_i\right) - \underset{\theta_{-i}}{E}\left(S(q_i(\hat{\theta}_i,\theta_{-i})) - \hat{\theta}_i q_i(\hat{\theta}_i,\theta_{-i})|\hat{\theta}_i\right).$$

Using the first-order condition, the above constraint rewrites as:

$$\begin{array}{l}\int_{\theta_i}^{\hat{\theta}_i} \underset{\theta_{-i}}{E}\left(q_i(x,\theta_{-i}) - (S(q_i(x,\theta_{-i})) - xq_i(x,\theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|x)}{\tilde{f}(\theta_{-i}|x)}|x\right)dx \geq \\ \underset{\theta_{-i}}{E}\left(S(q_i(\hat{\theta}_i,\theta_{-i})) - \theta_i q_i(\hat{\theta}_i,\theta_{-i})|\theta_i\right) - \underset{\theta_{-i}}{E}\left(S(q_i(\hat{\theta}_i,\theta_{-i})) - \hat{\theta}_i q_i(\hat{\theta}_i,\theta_{-i})|\hat{\theta}_i\right),\end{array} \tag{A.18}$$

When $q_i(\cdot)$ is the second-best schedule and for a fixed strictly positive marginal density $f(\cdot|\cdot)$, both sides of the inequality are continuous functions of the degree of correlation, where correlation is measured by the function $\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)$ and where continuity is with respect to the *supnorm*. For independent types, i.e., $\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i) = 0$, the above inequality becomes

$$\int_{\theta_i}^{\hat{\theta}_i} q_i^{BM}(x)dx \geq (\hat{\theta}_i - \theta_i)q_i^{BM}(\hat{\theta}_i), \tag{A.19}$$

which is clearly satisfied (with a strict inequality as soon as $\hat{\theta}_i \neq \theta_i$) and $q_i^{BM}(\theta_i)$ is strictly decreasing in θ_i. Moreover, under these hypothesis, the local second-order condition, which is also a continuous function of the degree of correlation, strictly holds for independent types since:

$$\frac{\partial q_i^{BM}}{\partial \theta_i}(\theta_i) < 0.$$

Therefore, a continuity argument shows that global incentive compatibility is satisfied for $\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)$ sufficiently small. ■

• **Proof of Proposition 4:** The proof is straightforwardly adapted from that of Proposition 1 by replacing the Bayesian incentive compatibility concept by the dominant strategy incentive compatibility concept. We omit the details. ■

• **Proof of Proposition 5:** The bilateral contracts exhibited in the proposition are such that the inefficient agents' participation constraints are binding, namely $u_i(\bar{\theta}, \theta_j) = 0$ for all $\theta_j \in \Theta$. These contracts satisfy also incentive compatibility. Moreover, they implement the optimal bilateral quantity schedules. They thus maximize the principal's expected payoff within the set of bilateral contracts.

We must check that a multilateral mechanism cannot achieve a greater payoff. Non-manipulability and dominant strategy incentive compatibility imply that there exists functions $h_i(\cdot)$ $(i = 1, 2)$ such that

$$h_i(\theta_i) = S(q_i(\theta_i, \theta_{-i})) - \theta_i q_i(\theta_i, \theta_{-i}) - u_i(\bar{\theta}, \theta_{-i}) - \int_{\theta_i}^{\bar{\theta}} q_i(x, \theta_{-i})dx \quad \forall \theta_{-i}, \tag{A.20}$$

and the program of the principal can be written

$$\max_{\{q(\cdot),h(\cdot)\}} \sum_{i=1}^{2} E_{\theta_i}(h_i(\theta_i))$$

subject to (A.20), $q_i(., \theta_{-i})$ decreasing 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.20) 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, such a mechanism can be implemented with a bilateral contract with A_i, i.e., with transfers $t_i(\theta_i) = t_i(\theta_i, \bar{\theta})$ which depend only on the type of this agent. ■

• **Proof of Proposition 6:** The proof first starts by showing that our a priori restriction in the set of non-manipulable mechanisms is without loss of generality. Then the rest follows quite closely that of Proposition 3.

Feasible Non-Manipulable Mechanisms: To justify that equation (22) does not entail any loss of generality in characterizing non-manipulable mechanisms, let us restate non-manipulability constraints in terms of the (symmetric) nonlinear price $T(q_i, \theta_i)$:

$$(q_1(\theta_1,\theta_2), q_2(\theta_1,\theta_2)) \in \arg \max_{\{(q_1,q_2)\in\mathcal{Q} \mid \sum_{i=1}^{2} q_i \le 1, q_i \ge 0\}} S\left(q_1+q_2\right) - T(q_1,\theta_1) - T(q_2,\theta_2). \tag{A.21}$$

Consider then any quantity schedule $(q_1(\theta_1,\theta_2), q_2(\theta_1,\theta_2))$ implemented through such a nonlinear price and such that $q_i(\theta_i,\theta_{-i})$ is weakly increasing in θ_{-i} and decreasing in θ_i. For such a schedule, define $\underline{q}(\theta_i) = \min_{\theta_{-i}}\{q_i(\theta_i,\theta_{-i})\}$ and $h(\theta_i) = S\underline{q}(\theta_i) - T(\underline{q}(\theta_i), \theta_i)$.

Lemma 4 *Any non-manipulable nonlinear price $T(q,\theta_i)$ is such that*

$$T(q,\theta_i) \le Sq - h(\theta_i) \text{ for any } q \in range(q_i(\theta_i,\cdot)). \tag{A.22}$$

Proof: Observe that choosing $\underline{q}(\theta_i) = \min_{\theta_{-i}}\{q_i(\theta_i,\theta_{-i})\}$ and keeping $q_{-i} = q_{-i}(\theta_i,\theta_{-i})$ is always a feasible option for the principal but it gives him less than choosing $q_i = q_i(\theta_i,\theta_{-i})$. Writing the corresponding condition gives (A.22). ■

Lemma 5 *Any non-manipulable nonlinear price $T(q,\theta_i)$ is such that*

$$Sq - T(q,\theta_i) \text{ is weakly increasing in } q. \tag{A.23}$$

Proof: Fix $q_{-i} = q_{-i}(\theta_i,\theta_{-i})$. Then, observe that the domain of optimization for q_i is $[0, 1-q_{-i}(\theta_i,\theta_{-i})]$. Since, $q_{-i}(\theta_i,\cdot)$ is decreasing in the second argument, we have

$$Sq_i(\theta_i,\theta_{-i}) - T(q_i(\theta_i,\theta_{-i}),\theta_i) \le Sq_i(\theta_i,\theta'_{-i}) - T(q_i(\theta_i,\theta'_{-i}),\theta_i). \tag{A.24}$$

for $\theta'_{-i} \ge \theta_{-i}$. Since $q_i(\theta_i,\theta_{-i})$ is weakly increasing in θ_{-i}, we get the result. ■

With a nonlinear price schedule, A_i's payoff can be written as

$$U(\theta_i) = \max_{\hat{\theta}_i \in \Theta} E_{\theta_{-i}}(T(q_i(\hat{\theta}_i, \theta_{-i}), \hat{\theta}_i) - \theta_i q_i(\hat{\theta}_i, \theta_{-i}) \mid \theta_i).$$

From which, we deduce:

$$\dot{U}(\theta_i) = E_{\theta_{-i}}\left((T(q_i(\theta_i, \theta_{-i}), \theta_i) - \theta_i q_i(\theta_i, \theta_{-i}))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \mid \theta_i\right) - E_{\theta_{-i}}(q_i(\theta_i, \theta_{-i}) \mid \theta_i).$$

By integrating, we get

$$U(\theta_i) = U(\bar{\theta}) + \int_{\theta_i}^{\bar{\theta}} E_{\theta_{-i}}\left(\left(1 + x\frac{\tilde{f}_{\theta_i}(\theta_{-i}|x)}{\tilde{f}(\theta_{-i}|x)}\right) q_i(x, \theta_{-i}) \mid x\right) dx$$

$$- \int_{\theta_i}^{\bar{\theta}} E_{\theta_{-i}}\left(T(q_i(x, \theta_{-i}), x)\frac{\tilde{f}_{\theta_i}(\theta_{-i}|x)}{\tilde{f}(\theta_{-i}|x)} \mid x\right) dx.$$

Lemma 6 *Among all mechanisms satisfying conditions (A.22) and (A.23), the rent $U(\theta_i)$ is minimized when:*

$$T(q, \theta_i) = Sq - h(\theta_i) \text{ for any } q \in range \quad q_i(\theta_i, \cdot). \tag{A.25}$$

Proof: From Lemmas 4 and 5, we deduce that

$$Sq_i(\theta_i, \theta_{-i}) - h(\theta_i) - T(q_i(\theta_i, \theta_{-i}), \theta_i) \text{ is positive and weakly increasing in } \theta_{-i}. \tag{A.26}$$

From Assumption 4, equation (17) and equation (A.26), we obtain that

$$E_{\theta_{-i}}\left((Sq_i(\theta_i, \theta_{-i}) - h(\theta_i) - T(q_i(\theta_i, \theta_{-i}), \theta_i))\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \mid \theta_i\right) \geq 0.$$

Therefore, the rent left to the agents is minimized when (A.25) holds. ■

No loss of generality was indeed introduced by neglecting the constraint $\sum_{i=1}^{2} q_i(\theta_i, \hat{\theta}_{-i}) \leq 1$ when we derived equation (22) as long as the optimal quantity schedule $q_i(\theta_i, \theta_{-i})$ is weakly increasing in θ_{-i} which turns out to be the case at the optimal mechanism as we see below.

Characterization of the Optimal Mechanism: Bayesian incentive compatibility implies

$$\dot{U}_i(\theta_i) = -\underset{\theta_{-i}}{E}(q_i(\theta_i, \theta_{-i})|\theta_i) + \underset{\theta_{-i}}{E}\left((S - \theta_i)q_i(\theta_i, \theta_{-i})\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\Big|\theta_i\right). \tag{A.27}$$

When types are weakly correlated, the right-hand side of (A.27) is decreasing and (11) binds at $\theta_i = \bar{\theta}$ only. Integrating by parts yields:

$$\underset{\theta_i}{E}(U_i(\theta_i)) = \underset{(\theta_i,\theta_{-i})}{E}\left(\frac{F(\theta_i)}{f(\theta_i)} q_i(\theta_i,\theta_{-i})\left(1-(S-\theta_i)\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\right)\right).$$

Using this expression of the expected rent left to agent A_i, we can rewrite the principal's objective function as:

$$\max_{\{q_i(\cdot)\}} \underset{(\theta_{-i},\theta_i)}{E}\left(\sum_{i=1}^{2} q_i(\theta_i,\theta_{-i})\left[(S-\theta_i)\left(1+\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\frac{F(\theta_i)}{f(\theta_i)}\right)-\frac{F(\theta_i)}{f(\theta_i)}\right]\right)$$

subject to (23).

We immediately find that $q_i^{SB}(\theta_i,\theta_{-i}) = 1$ if and only if the two following conditions hold:

- $$\theta_i + \frac{\frac{F(\theta_i)}{f(\theta_i)}}{1+\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\frac{F(\theta_i)}{f(\theta_i)}} < \theta_{-i} + \frac{\frac{F(\theta_{-i})}{f(\theta_{-i})}}{1+\frac{\tilde{f}_{\theta_{-i}}(\theta_i|\theta_{-i})}{\tilde{f}(\theta_i|\theta_{-i})}\frac{F(\theta_{-i})}{f(\theta_{-i})}} \qquad \text{(A.28)}$$
- $$S > \theta_i + \frac{\frac{F(\theta_i)}{f(\theta_i)}}{1+\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\frac{F(\theta_i)}{f(\theta_i)}} \text{ for all } \theta_{-i} \text{ such that (A.28) holds.} \qquad \text{(A.29)}$$

When Assumptions 4 and 5 hold, $\theta_{-i} > \theta_i$ implies then that

$$\theta_i + \frac{\frac{F(\theta_i)}{f(\theta_i)}}{1+\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\frac{F(\theta_i)}{f(\theta_i)}} < \theta_i + \frac{F(\theta_i)}{f(\theta_i)} < \theta_{-i} + \frac{F(\theta_{-i})}{f(\theta_{-i})}$$

where the right-hand side inequality follows from Assumption 3. Finally, we get

$$\theta_{-i} + \frac{F(\theta_{-i})}{f(\theta_{-i})} < \theta_{-i} + \frac{\frac{F(\theta_{-i})}{f(\theta_{-i})}}{1+\frac{\tilde{f}_{\theta_{-i}}(\theta_i|\theta_{-i})}{\tilde{f}(\theta_i|\theta_{-i})}\frac{F(\theta_{-i})}{f(\theta_{-i})}}$$

from using again Assumptions 4 and 5. Finally, (A.28) holds. The optimal auction is efficient.

Second, because Assumption 4 holds, (A.29) holds for all $\theta_{-i} > \theta_i$ if and only if it holds for $\theta_{-i} = \theta_i$. Taking into account Assumption 5 this gives the value of θ^* in (25).■

• **Proof of Lemma 1:** The proof is standard and is thus omitted. See for instance Laffont and Martimort (2002, Chapters 3 and 9). ■

• **Proof of Proposition 7:** Using (34) for differentiable outputs, we obtain:

$$\dot{U}_i(\theta_i) = -\underset{\theta_{-i}}{E}(q_i(\theta_i,\theta_{-i})|\theta_i)$$

$$+\underset{\theta_{-i}}{E}\left(\left(\int_0^{q_i(\theta_i,\theta_{-i})}\frac{\partial\tilde{S}}{\partial x}(x,q^e_{-i}(\theta_i,\phi_{-i}(\theta_i,x)))dx-\theta_i q_i(\theta_i,\theta_{-i})\right)\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}|\theta_i\right)\quad\forall i=1,2,\ \forall\theta_i\in\Theta. \tag{A.30}$$

The rent is decreasing when Assumption 2 holds and thus (11) is binding at $\bar{\theta}$. This yields the following expression of A_i's expected rent:

$$\underset{\theta_i}{E}\left(U_i(\theta_i)\right)=\underset{(\theta_i,\theta_{-i})}{E}\left(\frac{F(\theta_i)}{f(\theta_i)}q(\theta_i,\theta_{-i})\right)$$

$$-\underset{(\theta_i,\theta_{-i})}{E}\left(\left(\int_0^{q(\theta_i,\theta_{-i})}\frac{\partial\tilde{S}}{\partial x}(x,q^e_{-i}(\theta_i,\phi_{-i}(\theta_i,x)))dx-\theta_i q(\theta_i,\theta_{-i})\right)\frac{F(\theta_i)}{f(\theta_i)}\frac{\tilde{f}_\theta(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\right).$$

Inserting these expected rents into the principal's objective function yields the following optimization problem:

$$\max_{\{q(\cdot)\}}\underset{(\theta_1,\theta_2)}{E}\left(\tilde{S}(q_1(\theta_1,\theta_2),q_2(\theta_1,\theta_2))-\sum_{i=1}^2\left(\theta_i+\frac{F(\theta_i)}{f(\theta_i)}\right)q_i(\theta_1,\theta_2)\right)$$

$$+\sum_{i=1}^2\left(\int_0^{q_i(\theta_i,\theta_{-i})}\frac{\partial\tilde{S}}{\partial x}(x,q^e_{-i}(\theta_i,\phi_{-i}(\theta_i,x)))dx-\theta_i q_i(\theta_i,\theta_{-i})\right)\frac{F(\theta_i)}{f(\theta_i)}\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\Bigg).$$

Optimizing with respect to output this strictly concave objective and taking into account that, at the solution, expectations are correct yields (37).

By the same continuity argument as previously, the global incentive compatibility conditions for the agents' incentive problem are still satisfied when Assumption 2 holds. Indeed, $q_i(\theta_i,\theta_{-i})$ is strictly non-increasing in θ_i and non-decreasing in θ_{-i} if outputs are substitutes and non-increasing in θ_{-i} if outputs are complements as requested by Lemma 1. ■

• **Proof of Proposition 8:** The first steps follow those of the Proof of Proposition 7 with the specification of the nonlinear price given in (35). The principal's optimization problem becomes:

$$\max_{\{q(\cdot)\}}\underset{(\theta_1,\theta_2)}{E}\left(S(\sum_{i=1}^2 q_i(\theta_i,\theta_{-i}))-\sum_{i=1}^2\left(\theta_i+\frac{F(\theta_i)}{f(\theta_i)}\right)q_i(\theta_i,\theta_{-i})\right.$$

$$+\sum_{i=1}^2\left(\int_0^{q_i(\theta_i,\theta_{-i})}\frac{\partial\tilde{S}}{\partial x}(x+q^e_{-i}(\theta_i,\phi_{-i}(\theta_i,x)))dx-\theta_i q_i(\theta_i,\theta_{-i})\right)\frac{F(\theta_i)}{f(\theta_i)}\frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)}\Bigg).$$

Agent A_i with type θ_i produces all output when $\theta_i<\theta_{-i}$ since under Assumptions 3, 4 and 5, (A.28) holds. Then $q^e_{-i}(\theta_i,\phi_{-i}(\theta_i,q_i^{SB}(\theta_i,\theta_{-i}))=0$ and the optimal output allocation is given by (40). ■

• **Proof of Proposition 9:** The proof follows the same lines as before. Given that the transfer schedule satisfies (45), the agents' information rent can thus be written as:

$$U_i(\theta_i) = \max_{\hat{\theta}_i} \left\{ \underset{\theta_{-i}}{E} \left(\int_0^{q(\hat{\theta}_i,\theta_{-i})} \left(S'(x) - \frac{\partial T^e_{-i}}{\partial q}(x, \phi_{-i}(x, \hat{\theta}_i)) \right) dx - \theta_i q(\hat{\theta}_i, \theta_{-i}) | \theta_i \right) - H(\hat{\theta}_i) \right\}.$$

Using the Envelope Theorem yields:

$$\dot{U}_i(\theta_i) = -\underset{\theta_i}{E}\left(q(\theta_i, \theta_{-i})|\theta_{-i}\right)$$

$$+\underset{\theta_{-i}}{E} \left(\left(\int_0^{q(\theta_i,\theta_{-i})} \left(S'(x) - \frac{\partial T^e_{-i}}{\partial q}(x, \phi_{-i}(x, \theta_i)) \right) dx - \theta_i q(\theta_i, \theta_{-i}) \right) \frac{\tilde{f}_\theta(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \Big| \theta_i \right).$$

Under Assumption 2, the information rent of an agent is decreasing and the participation constraint of agent A_i is binding only at $\bar{\theta}$. This yields the expression of A_i's expected rent:

$$\underset{\theta_i}{E}\left(U_i(\theta_i)\right) = \underset{(\theta_i,\theta_{-i})}{E} \left(\frac{F(\theta_i)}{f(\theta_i)} q(\theta_i, \theta_{-i}) \right)$$

$$-\underset{(\theta_i,\theta_{-i})}{E} \left(\left(\int_0^{q(\theta_i,\theta_{-i})} \left(S'(x) - \frac{\partial T^e_{-i}}{\partial q}(x, \phi_{-i}(x, \theta_i)) \right) dx - \theta_i q(\theta_i, \theta_{-i}) \right) \frac{F(\theta_i)}{f(\theta_i)} \frac{\tilde{f}_\theta(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \right).$$

Inserting these expected rents into the principal's objective function yields the following optimization problem:

$$\max_{\{q(\cdot)\}} \underset{(\theta_1,\theta_2)}{E} \left(S(q(\theta_1, \theta_2)) - \left(\sum_{i=1}^2 \theta_i + \frac{F(\theta_i)}{f(\theta_i)} \right) q(\theta_1, \theta_2) \right.$$

$$\left. + \sum_{i=1}^2 \left(\int_0^{q(\theta_1,\theta_2)} \left(S'(x) - \frac{\partial T^e_{-i}}{\partial q}(x, \phi_{-i}(x, \theta_i)) \right) dx - \theta_i q(\theta_1, \theta_2) \right) \frac{F(\theta_i)}{f(\theta_i)} \frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \right).$$

Optimizing pointwise yields:

$$\left(1 + \sum_{i=1}^2 \frac{F(\theta_i)}{f(\theta_i)} \frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \right) S'(q(\theta_1, \theta_2))$$

$$= \sum_{i=1}^2 \theta_i \left(1 + \frac{F(\theta_i)}{f(\theta_i)} \frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \right) + \frac{F(\theta_i)}{f(\theta_i)} + \frac{F(\theta_i)}{f(\theta_i)} \frac{\tilde{f}_{\theta_i}(\theta_{-i}|\theta_i)}{\tilde{f}(\theta_{-i}|\theta_i)} \frac{\partial T_{-i}}{\partial q}(q(\theta_1, \theta_2), \theta_{-i}) \quad \text{(A.31)}$$

where we have taken into account that expectations about the nonlinear price $T_i(q, \theta_i)$ are correct in equilibrium.

Also, the first-order condition for (44) can be written as:

$$S'(q(\theta_1, \theta_2)) = \sum_{i=1}^2 \frac{\partial T_i}{\partial q}(q(\theta_1, \theta_2), \theta_i). \quad \text{(A.32)}$$

We are looking for a pair $\left(\frac{\partial T_1}{\partial q}(q(\theta_1,\theta_2),\theta_1), \frac{\partial T_2}{\partial q}(q(\theta_1,\theta_2),\theta_1)\right)$ which solves (A.31) and (A.32). The pair of marginal contributions given in (48) does the job.

If $q^{SB}(\cdot)$ is decreasing in θ_i (which is true for a sufficiently small degree of correlation), the second-order condition of the agent's problem holds and global incentive compatibility is ensured. ■

Appendix B

• **The Simple Discrete Example Continued:** To highlight the role of output distortions in the optimal mechanism, we come back to the simple discrete example of Section 4 with the added twist that the signal used to improve contracting with one agent comes in fact from the (truthful in equilibrium) report of the other agent. Without any possibility to distort the level of production, we saw there that private communication obliges the principal to implement bilateral contracts. We now investigate what happens when the principal's surplus depends on output and is separable and of the form $S(q_1) + S(q_2)$.

We shall assume again that the type space is discrete, $\Theta = \{\underline{\theta}, \bar{\theta}\}$ and we denote by $[p(\theta_i|\theta_j)]_{(i,j)}$ the matrix of conditional probabilities. Slightly abusing notation, we will also denote by $[p(\theta_i,\theta_j)]_{(i,j)}$ the matrix of joint probabilities. Symmetry imposes moreover that $p(\bar{\theta},\underline{\theta}) = p(\underline{\theta},\bar{\theta})$.

Since a non-manipulable mechanism makes the principal indifferent among all transfer-outputs pairs conditionally on an agent's report, there necessarily exist two constants $h(\bar{\theta})$ and $h(\underline{\theta})$ such that:

$$h(\underline{\theta}) = S(q(\underline{\theta},\underline{\theta})) - t(\underline{\theta},\underline{\theta}) = S(q(\underline{\theta},\bar{\theta})) - t(\underline{\theta},\bar{\theta}); \quad h(\bar{\theta}) = S(q(\bar{\theta},\underline{\theta})) - t(\bar{\theta},\underline{\theta}) = S(q(\bar{\theta},\bar{\theta})) - t(\bar{\theta},\bar{\theta}). \tag{B.1}$$

Denoting also by $U(\underline{\theta})$ and $U(\bar{\theta})$ the information rent for the $\underline{\theta}$ and the $\bar{\theta}$ type respectively, the agents' participation constraints can be written as:

$$U(\underline{\theta}) = p(\underline{\theta}|\underline{\theta})(S(q(\underline{\theta},\underline{\theta})) - \underline{\theta}q(\underline{\theta},\underline{\theta})) + p(\bar{\theta}|\underline{\theta})(S(q(\underline{\theta},\bar{\theta})) - \underline{\theta}q(\underline{\theta},\bar{\theta})) - h(\underline{\theta}) \geq 0. \tag{B.2}$$

$$U(\bar{\theta}) = p(\underline{\theta}|\bar{\theta})(S(q(\bar{\theta},\underline{\theta})) - \bar{\theta}q(\bar{\theta},\underline{\theta})) + p(\bar{\theta}|\bar{\theta})(S(q(\bar{\theta},\bar{\theta})) - \bar{\theta}q(\bar{\theta},\bar{\theta})) - h(\bar{\theta}) \geq 0. \tag{B.3}$$

With this notation, Bayesian incentive compatibility constraints become:

$$\begin{aligned} U(\underline{\theta}) &\geq U(\bar{\theta}) + \Delta\theta\left(p(\underline{\theta}|\underline{\theta})q(\bar{\theta},\underline{\theta}) + p(\bar{\theta}|\underline{\theta})q(\bar{\theta},\bar{\theta})\right) \\ &+ \left(p(\underline{\theta}|\underline{\theta}) - p(\underline{\theta}|\bar{\theta})\right)\left(S(q(\bar{\theta},\underline{\theta})) - \bar{\theta}q(\bar{\theta},\underline{\theta}) - (S(q(\bar{\theta},\bar{\theta})) - \bar{\theta}q(\bar{\theta},\bar{\theta}))\right), \end{aligned} \tag{B.4}$$

$$U(\bar{\theta}) \geq U(\underline{\theta}) - \Delta\theta\left(p(\underline{\theta}|\bar{\theta})q(\underline{\theta},\underline{\theta}) + p(\bar{\theta}|\bar{\theta})q(\underline{\theta},\bar{\theta})\right)$$

$$+\left(p(\underline{\theta}|\bar{\theta})-p(\underline{\theta}|\underline{\theta})\right)\left(S(q(\underline{\theta},\underline{\theta}))-\underline{\theta}q(\underline{\theta},\underline{\theta})-(S(q(\underline{\theta},\bar{\theta}))-\underline{\theta}q(\underline{\theta},\bar{\theta}))\right). \quad \text{(B.5)}$$

The optimal mechanism solves thus the following problem:

$$\max_{\{U(\cdot),h(\cdot),q(\cdot)\}} 2(p(\underline{\theta})h(\underline{\theta})+p(\bar{\theta})h(\bar{\theta}))$$

subject to constraints (B.1) to (B.5).

For further references, we will denote by ρ the correlation coefficient defined as $\rho = p(\underline{\theta},\underline{\theta})p(\bar{\theta},\bar{\theta})-p^2(\underline{\theta},\bar{\theta})$ that is assumed to be positive ($\rho \geq 0$).

In this context, the Baron-Myerson outputs are defined as:

$$S'(q^{BM}(\underline{\theta},\underline{\theta}))=S'(q^{BM}(\underline{\theta},\bar{\theta}))=\underline{\theta} \text{ and } S'\left(q^{BM}(\bar{\theta},\underline{\theta})\right)=S'\left(q^{BM}(\bar{\theta},\bar{\theta})\right)=\bar{\theta}+\frac{p(\bar{\theta},\underline{\theta})+p(\underline{\theta},\underline{\theta})}{p(\bar{\theta},\underline{\theta})+p(\bar{\theta},\bar{\theta})}\Delta\theta.$$

Proposition 10 *: Assume that the correlation is weak enough, the optimal non-manipulable mechanism entails:*

- *No distortion at the top for the agents' outputs when they are efficient*

$$S'\left(q^{SB}(\underline{\theta},\underline{\theta})\right)=S'\left(q^{SB}(\underline{\theta},\bar{\theta})\right)=\underline{\theta}; \quad \text{(B.6)}$$

- *A downward distortion for the agents' outputs when they are inefficient*

$$S'\left(q^{SB}(\bar{\theta},\bar{\theta})\right)=\bar{\theta}+\frac{p(\bar{\theta},\underline{\theta})}{p(\bar{\theta},\bar{\theta})+\frac{\rho}{p(\bar{\theta},\underline{\theta})+p(\bar{\theta},\bar{\theta})}}\Delta\theta \quad \textit{and} \quad S'\left(q^{SB}(\bar{\theta},\underline{\theta})\right)=\bar{\theta}+\frac{p(\underline{\theta},\underline{\theta})}{p(\bar{\theta},\underline{\theta})-\frac{\rho}{p(\bar{\theta},\underline{\theta})+p(\bar{\theta},\bar{\theta})}}\Delta\theta, \quad \text{(B.7)}$$

with

$$q^{SB}(\bar{\theta},\bar{\theta})\geq q^{BM}(\bar{\theta},\bar{\theta})\geq q^{SB}(\bar{\theta},\underline{\theta});$$

- *Only the efficient agents get a positive rent*

$$U^{SB}(\underline{\theta})>0=U^{SB}(\bar{\theta}). \quad \text{(B.8)}$$

Proof: We conjecture that (B.3) and (B.4) are the two binding constraints and we will check ex post the validity of this claim.

Suppose it is so, then incorporating $U(\bar{\theta})=0$ and

$$U(\underline{\theta})=\Delta\theta(p(\underline{\theta}|\underline{\theta})q(\underline{\theta},\bar{\theta})+p(\bar{\theta}|\underline{\theta}))q(\bar{\theta},\bar{\theta})$$

$$+(p(\underline{\theta}|\underline{\theta})-p(\underline{\theta}|\bar{\theta}))(S(q(\underline{\theta},\bar{\theta}))-\bar{\theta}q(\underline{\theta},\bar{\theta}))+(p(\bar{\theta}|\underline{\theta})-p(\bar{\theta}|\bar{\theta}))(S(q(\bar{\theta},\bar{\theta}))-\bar{\theta}q(\bar{\theta},\bar{\theta}))$$

into the principal's objective function yields an expression of the maximand which is strictly concave in $q(\underline{\theta},\bar{\theta})$ if and only $\rho < p(\underline{\theta},\bar{\theta})(p(\underline{\theta},\underline{\theta})+p(\underline{\theta},\bar{\theta}))$. Optimizing yields then the conditions (B.6) to (B.7).

Because $S'(q^{SB}(\bar{\theta},\bar{\theta})) > S'(q^{SB}(\underline{\theta},\bar{\theta})) > \bar{\theta}$, we have

$$S(q^{SB}(\underline{\theta},\bar{\theta})) - \bar{\theta}q^{SB}(\bar{\theta},\underline{\theta}) > S(q^{SB}(\bar{\theta},\bar{\theta})^{SB}) - \bar{\theta}q^{SB}(\bar{\theta},\bar{\theta})$$

and the right-hand side of (B.4) is strictly positive since $p(\underline{\theta}|\underline{\theta}) - p(\underline{\theta}|\bar{\theta}) = -(p(\bar{\theta}|\underline{\theta}) - p(\bar{\theta}|\bar{\theta})) = \frac{\rho}{(p(\underline{\theta},\underline{\theta})+p(\underline{\theta},\bar{\theta}))(p(\bar{\theta},\underline{\theta})+p(\bar{\theta},\bar{\theta}))}$.

Then, it is straightforward to check that (B.8) holds and that (B.5) is slack for a sufficiently weak correlation. ■

Proposition 10 is the exact counterpart of Proposition 3 in the discrete types case. The underlying intuition is the same and is straightforwardly adapted. However, the discrete model also allows us to get a result for the polar case of almost perfect correlation:

Proposition 11 *: Assume that the correlation between types becomes almost perfect, i.e., $p(\bar{\theta},\underline{\theta})$ converges towards zero. Then, the principal approximates his first-best expected payoff arbitrarily closely enough even if the manipulability constraints are taken into account.*

Proof: Consider the following output schedule $(q^{FB}(\bar{\theta},\bar{\theta}), \hat{q}_2(p(\bar{\theta},\underline{\theta})), q^{FB}(\bar{\theta},\underline{\theta}), q^{FB}(\underline{\theta},\underline{\theta}))$ where $\hat{q}_2(p(\bar{\theta},\underline{\theta}))$ is the smallest output $\hat{q}_2$ such that:

$$\Delta\theta\left(p(\underline{\theta}|\underline{\theta})\hat{q}_2 + p(\bar{\theta}|\underline{\theta})q^{FB}(\bar{\theta},\bar{\theta})\right)$$
$$+\left(p(\underline{\theta}|\underline{\theta}) - p(\underline{\theta}|\bar{\theta})\right)\left(S(\hat{q}_2) - \bar{\theta}\hat{q}_2 - (S(q^{FB}(\bar{\theta},\bar{\theta})) - \bar{\theta}q^{FB}(\bar{\theta},\bar{\theta}))\right) = 0,$$

When $p(\bar{\theta},\underline{\theta})$ converges towards 0, $q(\underline{\theta},\bar{\theta})(p(\bar{\theta},\underline{\theta}))$ converges towards $\hat{q}_2(0) < q^{FB}(\bar{\theta},\bar{\theta})$ such that

$$S(\hat{q}_2(0)) - (\bar{\theta}-\Delta\theta)\hat{q}_2(0) = S(q^{FB}(\bar{\theta},\bar{\theta})) - \bar{\theta}q^{FB}(\bar{\theta},\bar{\theta}),$$

Clearly, the payoff achieved by the optimal mechanism (keeping q fixed as above) found with (B.3) and (B.4) binding converges towards the first-best as $p(\bar{\theta},\underline{\theta})$ goes to zero, moreover, it clearly satisfy (B.5) and (B.2). ■

Propositions 3, 10 and 11 altogether show not only the value of the non-manipulability constraint but also its limits. When the correlation is weak, the general thrust is that although correlated information is useful and improves on simple bilateral contracting, it does not destroy the basic lesson of standard agency models: Information rents have to be given up and the cost of doing so remains some output distortions.

The non-manipulability constraints have much less bite when correlation is very strong. Indeed, inducing a given agent to truthfully report this "almost" common piece of information is easier and, intuitively, the principal does not need to rely so much on transfer lotteries which would be easily manipulable.

However, focusing on this case of a strong correlation is nevertheless a bit too extreme as a robustness check of the literature on surplus full-extraction. What has been mostly seen as a serious attack of mechanism design is the lack of continuity between the case of even a small amount of correlation and the case of independent types that this literature has uncovered. With even a little bit of correlation, the principal is able to extract all the agents' information rent and implement the first-best. With no correlation, information rents have to be conceded and outputs are distorted downwards to reduce the cost of doing so. Introducing the non-manipulability constraints restores the continuity between the two environments. As it can be seen on formulas (16) and (B.7), the optimal outputs converge towards the Baron-Myerson solution as the correlation diminishes. In the limit of zero correlation, the non-manipulability constraints have no bite. ■