

Mechanism Design with Collusion and Correlation*

Jean-Jacques Laffont**

David Martimort***

Revised Version: December 98

ABSTRACT

In a public good environment with positively correlated types, we characterize optimal mechanisms when agents have private information and can enter collusive agreements. First, we prove a weak-collusion-proof principle according to which there is no restriction for the principal in offering weak-collusion-proof mechanisms. Second, with this principle, we characterize the set of allocations which satisfy individual and coalitional incentive constraints. The optimal weakly collusion-proof mechanism calls for distortions away from first-best efficiency obtained without collusion. Allowing collusion restores continuity between the correlated and the uncorrelated environments. When the correlation becomes almost perfect, first-best efficiency is approached. Finally, the optimal collusion-proof mechanism is strongly ratifiable.

Keywords : Mechanism Design, Collusion, Correlation.

* We thank Jeon Doh-Shin, Antoine Faure-Grimaud, Oliver Hart, Eric Maskin, Patrick Rey, Mike Riordan, Tomas Sjostrom, Yossef Spiegel, Dimitri Vayanos, Robert Wilson, seminar participants at Humbolt University-Berlin, Bocconi University, University of Maryland, Paris-Séminaire Roy, University Pompeu Fabra, University of California at Berkeley, University College London, Stanford Business School, University of Alabama, three referees and Andrew Postlewaite for helpful comments. Financial support from INRA's AIP Contracts is gratefully acknowledged by David Martimort.

** Université des Sciences Sociales, GREMAQ and IDEI, Toulouse.

*** Université de Pau et des Pays de l'Adour, GREMAQ and IDEI, Toulouse.

1 Introduction

The provision of public goods under informational constraints is one of the leading textbook examples of public economics. How should a society made of several agents with heterogeneous tastes for a public good design an incentive mechanism to induce truthful revelation of the agents' valuations? Does efficiency conflict with incentives? How can this conflict, if any, be solved? Finally, what is the distribution of informational rents induced by asymmetric information?

One striking feature of most previous works on these issues is that they deal mainly with the case where agents are unable to form coalitions to collectively manipulate the decision rule.¹ By focusing on the role of individual incentive constraints, an important dimension of resource allocation in society has been neglected: the formation of groups. It is particularly troublesome when these coalitions can significantly reduce the efficiency of the optimal mechanism designed in the absence of *coalition incentive constraints*. As Olson (1965, p. 1) has forcefully emphasized "*groups of individuals with common interests usually attempt to further those common interests [...] and are expected to act on behalf of their common interests much as single individuals are often expected to act on behalf of their personal interest.*" Following this argument, the standard theoretical framework must be amended to allow also for the formation of groups promoting their own collective goals instead of that of society as a whole. The present paper offers a framework in which the consequences of collusion under asymmetric information on both allocative efficiency and the distribution of rents in society can be assessed.

We model a conflict between a social welfare maximizer and a group of agents who benefit from the public good but who do not care about its budgetary cost. Contrary to the standard assumption in mechanism design, the planner has not a perfect control of the communication technology so that he cannot prevent the agents from colluding.² Collusion between these agents is modeled in reduced form as in Laffont and Martimort (1997). A third-party proposes to the colluding agents a side-mechanism to collectively manipulate their sending of messages to the government. As suggested above, this third-party does not internalize the social cost of the project but maximizes only the sum of the agents' utilities. Lastly, no technology for a credible disclosure of information is available to the colluding partners. The mere forming of a coalition does not change

¹Groves (1973) and Green and Laffont (1977) showed that efficiency could be achieved under asymmetric information on the agents' valuations with dominant strategy if budget balance is not a concern. Arrow (1979) and Aspremont and Gérard-Varet (1979) showed that budget balance could be achieved under Bayesian implementation. Laffont and Maskin (1982) and Mailath and Postlewaite (1990) showed that adding the possibility for the agents to veto the mechanism introduces a real conflict between efficiency and incentives leading to the underprovision of the public good. Ledyard and Palfrey (1996) show that this conflict also arises in the absence of participation constraints when incentives conflict with redistribution concerns.

²See Palfrey (1992) for a discussion of this assumption and some of its implications in the case of Bayesian implementation.

informational asymmetries between the agents. Coalition formation takes place under *asymmetric information*.

When agents do not collude, their Bayesian-Nash behavior does not put any constraint on the set of interim individually rational and incentive compatible allocations in a *correlated information environment*. Indeed, as shown by Crémer and McLean (1988) in the case of auction mechanisms, the existence of even a small amount of correlation between the agents' valuations for the public good is enough to allow the principal to elicit this "almost common" information. This result is in sharp contrast with the case of uncorrelated information since then efficiency does conflict with incentives. Hence, the optimal levels of public good exhibit discontinuities when the degree of correlation goes to zero.

This costless extraction of the agents' surplus by the principal suggests also that they are likely to form an active coalition in such a correlated information environment. Nevertheless, a *weak collusion-proofness principle* holds in this context: any equilibrium of the overall game of contract offer cum coalition formation achieves an outcome which can be replicated with a *weakly collusion-proof* grand-mechanism, i.e., a grand-mechanism such that the null side-contract is a continuation equilibrium of the game of coalition formation. Since one agent's acceptance of the side-contract depends on the status quo utility level that he gets from playing non-cooperatively the grand-mechanism offered by the principal, the issue of learning from disagreement arises. Weakly collusion-proof mechanisms are such that the null side-mechanism is a continuation equilibrium of the game of coalition formation sustained with *passive beliefs*.³

The weak collusion-proofness principle provides a tractable description of the set of perfect Bayesian equilibria of the overall game of contract offer cum coalition formation. After having described this set, the principal's welfare is optimized subject to participation, individual and *coalition incentive constraints*. Generally, the efficient levels of public good can no longer be costlessly implemented even in a correlated environment. Taking into account coalition incentive constraints, there exists now a trade-off between efficiency and rent extraction. Distortions in the quantities of public good which are produced in the different states of nature reduce the cost of the binding coalition incentive constraints.

Depending on the degree of correlation, collusion-proofness constraints take quite different forms. For weak positive correlation, collusion-proofness constraints are similar to those that would arise under symmetric information within the coalition. When the degree of correlation diminishes, coalition incentive constraints are then less and less binding and the principal prevents more easily collusion. In the limit of *uncorrelated information*, the principal costlessly obtains collusion-proofness and the contractual outcome is the same as if agents had not been colluding. By adding coalition incentive constraints, one moves then *continuously* from the outcome with a strictly positive correlation to the outcome

³Rubinstein (1985) coined this expression for games with asymmetric information in which equilibrium behavior is sustained with prior beliefs out of the equilibrium path.

with no correlation.⁴

For strong correlation, collusion-proofness constraints under asymmetric information are instead quite different from those obtained when agents can credibly disclose their information. When the correlation becomes almost perfect, there is only a small probability that agents have different valuations for the public good. The principal can shut-down production in this state of nature at almost no social cost. This breaks the coalition agreement and almost achieves the first-best level of expected welfare.

Since it is implemented with Bayesian strategies, the optimal weakly collusion-proof contract is sensitive to the exact beliefs that the agents have at the time of playing this mechanism. Because we focus on the case where valuations for the public good may take only two values, the equilibrium correspondence of this optimal mechanism as posterior beliefs change can be fully described. This step of the analysis allows us to discuss the robustness of the optimal mechanism to a preplay communication stage in which agents may veto or ratify the truthful play of this mechanism and thereby signal some information to each other. It is then possible to show that the optimal weakly collusion-proof mechanism is *strongly ratifiable* in the sense of Cramton and Palfrey (1995).

Collusion in public good mechanisms has been first analyzed by Green and Laffont (1979) who prove that the Groves mechanisms are not robust to coalitions when agents share freely their information. Still with dominant strategy mechanisms but with a continuum of types, Laffont and Maskin (1980) show then that only pooling decision rules can be implemented. Crémer (1996) takes into account asymmetric information within coalitions and shows that the Groves mechanisms are not robust to size-two coalitions but not beyond. Similarly, there exist some relatively negative results in the case of Nash implementation when agents can form coalitions (Maskin (1979)). The Nash environment can be seen as an extreme case of perfect correlation between the agents' types. Our focus on Bayesian implementation with two types brings more positive results.⁵ Under asymmetric information within the coalition, the principal can implement a much larger set of allocations in these strongly correlated environments. More generally, our approach provides a complete description of the set of implementable allocations under collusion.

This paper extends Laffont and Martimort (1997) by stressing the role of correlated information between the agents as a determinant of the strength of their coalition and by making no restriction on the set of available mechanisms. In this previous work, we restricted the analysis to the case of anonymous mechanisms and uncorrelated information.⁶ With no correlation, there always exists a costless weakly collusion-proof implementation

⁴Interestingly, this result requires neither risk-aversion of the agents' utility functions nor limited liability constraints on transfers (Robert (1991)).

⁵See also Laffont and Maskin (1979).

⁶Beside the differences in the informational structures and the set of available mechanisms, this previous paper was dealing with a model of regulation. This latter difference is without consequence on the results.

of the second-best non-cooperative outcome if the principal can offer non-anonymous Bayesian mechanisms. However, collusion still matters when types are correlated even without any exogenous restriction on the set of mechanisms. Moreover, working in a correlated environment, we obtain a full characterization of all transfers in the optimal weakly collusion-proof mechanism. This characterization allows to compute all the ex post rents precisely and to characterize the outcome of the grand-mechanism when beliefs differ from passive ones. This is an important step towards checking the strong ratifiability of the mechanism.

Section 2 presents the model. Section 3 discusses the optimal mechanism when agents do not collude. Section 4 derives the weak collusion-proofness principle and characterizes weakly collusion-proof allocations. Section 5 describes the optimal weakly collusion-proof mechanism. Section 6 discusses strong ratifiability. Section 7 concludes.

2 The Model

2.1 Technology, Preferences and Information

We consider the provision of public good in a partial equilibrium model. A positive amount x of public good can be produced at cost $c(\cdot)$ with $c'(\cdot) > 0$ and $c''(\cdot) > 0$. There are two agents in the economy denoted by $A_i, i \in \{1, 2\}$.⁷ Each of them derives utility $U_i = \theta_i x - t_i$ from consuming an amount x of public good and paying tax t_i . Agent A_i agrees to participate in the public good mechanism when his participation constraint is satisfied.

The agents' valuations for the public good, $\theta_i, i \in \{1, 2\}$, are drawn from a common knowledge joint distribution on Θ^2 where $\Theta = \{\underline{\theta}, \bar{\theta}\}$ is the common support of θ_1 and θ_2 ($\Delta\theta = \bar{\theta} - \underline{\theta}$). We refer to the probabilities $p(\theta_i, \theta_j)$ of each state (θ_i, θ_j) , for $(i, j) \in \{1, 2\}^2$, as the common knowledge prior beliefs. To make notation simpler, we also write: $p(\bar{\theta}, \bar{\theta}) = p_{11}$, $p(\underline{\theta}, \bar{\theta}) = p(\bar{\theta}, \underline{\theta}) = p_{12}$, $p(\underline{\theta}, \underline{\theta}) = p_{22}$, where the equality $p(\underline{\theta}, \bar{\theta}) = p(\bar{\theta}, \underline{\theta})$ is derived from the symmetry between the two agents. Finally, to capture the congruence of the agents' interests, θ_1 and θ_2 are *positively correlated* and $\frac{p_{12}}{p_{11}} \leq \frac{p_{22}}{p_{12}}$. We denote by $\rho = p_{11}p_{22} - p_{12}^2$ the degree of positive correlation ($\rho = 0$ for independent types). For simplifying technicalities, we also assume that $p_{12} \leq p_{11}$.⁸ The conditional beliefs of agent A_i on A_j 's type ($j \neq i$) induced by the joint distribution above are the same for both agents and, slightly abusing notations, are denoted by p .

The government, or principal P , only knows the distribution of the agents' valuations

⁷Restricting to two agents avoids to consider the formation of subcoalitions and significantly simplifies notations. However, our methodology could be extended to more than two agents at the cost of an increase in complexity.

⁸This assumption ensures that the optimal collusion-proof mechanism under asymmetric information never entails bunching in the case of small correlation. It simplifies significantly the exposition.

for the public good but is uninformed on the exact realizations of these shocks at the time of choosing the public good mechanism. His objective is to maximize the sum of the agents' utilities knowing that the deficit for the production of the public good must be covered by distortionary taxation raised elsewhere in the economy. Formally, the social welfare function is written as $SW = \sum_{i=1}^2 U_i - (1 + \lambda)(c(x) - \sum_{i=1}^2 t_i)$, where λ is the exogenous cost of public funds.⁹

2.2 Mechanisms

The principal proposes a grand-mechanism G to the agents. G maps any pair of messages (m_1, m_2) belonging to the product message space $M_1 \times M_2 = M$ (where M_i denotes the message space used by agent A_i) into a triplet $\{x, t_1, t_2\}$. x denotes the amount of public good produced ($x \in X = \mathbb{R}^+$) and t_i ($i \in \{1, 2\}$) is the tax paid by agent A_i to the principal. We denote by $G = \{x(\cdot), t_1(\cdot), t_2(\cdot)\}$ this grand-mechanism.¹⁰

To make notation simpler in the case of direct mechanisms ($M = \Theta \times \Theta$), we denote by \bar{x}, \hat{x} and \underline{x} respectively the levels of public good when both agents claim $\bar{\theta}$, when their claims differ $(\bar{\theta}, \underline{\theta})$ and when they both claim $\underline{\theta}$. We also denote by t_{kl} for $k, l \in \{1, 2\}$ the tax paid by an agent whose type is $\theta_i = \underline{\theta} + (2 - k)\Delta\theta$ when the other agent's type is $\theta_j = \underline{\theta} + (2 - l)\Delta\theta$. Because of symmetry between the agents, the corresponding taxes are independent of the agents' identity.¹¹

2.3 Coalition Formation

An uninformed third-party, T , proposes a side-mechanism $S = \{\phi(\cdot), y_i(\cdot)_{i \in \{1, 2\}}\}$ to the agents to induce their collusive behavior.

- $\phi(\cdot)$ is a collective manipulation of the messages sent to the principal.
- $\{y_i(\cdot)\}_{i \in \{1, 2\}}$ is a pair of side-transfers. The third-party is not a source of money and therefore the coalition's budget is balanced: $\sum_{i=1}^2 y_i(\theta_1, \theta_2) = 0$ for all $(\theta_1, \theta_2) \in \Theta^2$.

From the revelation principle, there is no loss of generality in assuming that S is a

⁹The model is formally equivalent to the Laffont and Tirole (1986) partial equilibrium model of regulation. It could be possible to build a general equilibrium model endogenizing the value of λ . This could be done by introducing a third uninformed agent, say A_3 , in the analysis. By imposing that the sum of the contributions made by this agent and the group $A_1 - A_2$ covers exactly the cost of the public good as in Aspremont and Gérard-Varet (1978), we would be able to endogenize the value of the budget constraint's multiplier. Moreover, because of complete information on A_3 's valuation for the public good, this agent could be forced to always pay his valuation for the public good. One would then be interested by the coalition between A_1 and A_2 against A_3 .

¹⁰Note that we do not restrict a priori the set of mechanisms available to the set of direct mechanisms. Other message spaces than the product of the agents' type spaces Θ^2 can be used by the principal.

¹¹The symmetry of the grand-mechanism is without loss of generality under a non-cooperative behavior as we will see below. We also show in the Appendix that it is without loss of generality in the case of collusion for a small correlation. The amount of public good does not need to depend on the identity of who has a high valuation for the public good when claims differ.

direct mechanism.¹² Therefore $\phi(\cdot)$ and $y_k(\cdot)$ ($k \in \{1, 2\}$) map Θ^2 respectively into the set of measures on M and the set of balanced side-transfers.

Lastly, T is benevolent and maximizes the sum $U_1 + U_2$ of the two colluding agents' utilities obtained by playing the composition of the grand- and the side-mechanism.¹³

2.4 Timing of the Game

The timing of the overall game of contract offer cum coalition formation is as follows (see also Figure 1 for the game tree):

1. Agents learn their respective valuations for the public good.
2. P proposes a grand-mechanism G . If an agent vetoes the grand-mechanism, all agents get their reservation utility normalized exogenously at zero.
3. The third-party proposes a side-mechanism S to the agents and a non-cooperative continuation play of G if anyone refuses this side-contract. If both agents accept S , agents report their types to the third-party who recommends reports into the grand-mechanism and who commits to enforce the corresponding side-transfers.
4. Reports are sent into the grand-mechanism. The decision on the size of the public good is made and taxes are paid by the agents. Side-transfers, if any, are implemented.

The third party's offer of a side-mechanism S on top of the grand-mechanism G induces a two stage game $\hat{\Gamma}(G, S)$. In the first *ratification stage*, agents simultaneously accept or refuse the side-mechanism and may thereby signal their types to each other. In the second *communication stage*, agents send messages either directly to the principal if at least one of them has refused the side-mechanism or to the third-party if both have accepted. The third-party recommends then a collective manipulation of the messages to be sent to the principal. We denote by $\hat{E}(G, S)$ the set of perfect Bayesian equilibria of $\hat{\Gamma}(G, S)$.

We are interested in finding the optimal mechanism G , knowing that the continuation game of coalition formation consists first of a side-mechanism S optimally chosen by the third-party and second of a ratification-communication game $\hat{\Gamma}(G, S)$.

¹²For any grand-mechanism offered by the principal, one can restrict the third-party to use direct side-mechanisms at the final stage of the game of contract offer cum coalition formation.

¹³Using this third-party as a side-contract mechanism designer avoids the difficult issue of informational leakages through contract offers. It eliminates also the problem of finding an extensive form for describing the collusive game between the agents. This third-party paradigm can be seen as a black-box for the repeated interaction by which collusion emerges. This is a modeling short-cut to justify also our assumption that the side-contract is in fact enforceable even if there is no court of justice available to do so. This modeling characterizes the highest bound that can be achieved by the coalition.

Note first that we eliminate equilibria based on weakly dominated strategies at stage 3 of the overall game. Indeed, for any grand-mechanism G , there always exists a continuation equilibrium of the game of coalition formation in which each agent refuses any collusive offer he may receive because he expects that the other agent also refuses this offer anyway. Second, following A_i 's rejection, A_j ($j \neq i$) may have updated his beliefs on A_i 's type. These beliefs affect the non-cooperative play of the grand-mechanism G and therefore the status quo payoffs that A_i gets following a rejection of the side-mechanism S . Therefore, there may exist several side-mechanisms offered as continuation equilibria of the game of coalition formation depending on what is learnt following the rejection of these side-mechanisms.

Let $\{\tilde{p}_1, \tilde{p}_2\}$ be a belief system where \tilde{p}_i are agent A_{-i} 's beliefs on agent A_i if A_{-i} contemplates A_i 's refusal to play the side-mechanism S . We denote by $\Gamma(G, \tilde{p}_i, p_{-i})$ the game of asymmetric information induced by the grand-mechanism G at stage 4 following A_i 's refusal of playing S . In particular $\Gamma(G, p, p)$ denotes this game of asymmetric information when it is played with passive prior beliefs. $E(G, \tilde{p}_i, p_{-i})$ denotes the set of Bayesian-Nash equilibria of $\Gamma(G, \tilde{p}_i, p_{-i})$. Finally, let us denote by $U_i(\theta_i, e_i)$ the payoff of a θ_i agent A_i in an equilibrium $e_i \in E(G, \tilde{p}_i, p_{-i})$. Note that this interim payoff is computed as an expectation with respect to prior beliefs. Indeed, because joint deviations have probability zero in a Bayesian-Nash equilibrium, nothing has been learned on A_{-i} following A_i 's refusal. Therefore, the deviant agent A_i still continues to play the grand-mechanism with his prior beliefs on A_{-i} 's type.

We are interested in collusive continuation equilibria in which no learning occurs from the agreement of playing the side-mechanism S . We can thus define similarly $\Gamma(G \circ S, p, p)$ the game of asymmetric information induced by the composition of the grand-mechanism G and the side-mechanism S at stage 4 following acceptance of playing S by both agents. $E(G \circ S, p, p)$ denotes similarly the set of Bayesian-Nash equilibria of $\Gamma(G \circ S, p, p)$. Since the revelation principle applies at the last stage of the game, there is no loss of generality in considering that $E(G \circ S, p, p)$ contains the truthful equilibrium e^* . Let $U_i(\theta_i)$ denote agent A_i 's payoff in this equilibrium when his type is θ_i .

3 No-Collusion and the First-Best Outcome

It is by now a well known result that the optimal mechanism achieves the first-best outcome in this correlated environment when the implementation concept is Bayesian-Nash equilibrium, even with interim individual rationality constraints (Cr mer and McLean (1988)).¹⁴ The key is to use agent A_2 's report, which is truthful in equilibrium, as a signal correlated with agent A_1 's type and to condition A_1 's taxes on this information. The

¹⁴See also McAfee and Reny (1991) in the case of a continuum of types and Riordan and Sappington (1988) for a related model using ex post information in the case of only one agent.

flexibility in the taxes paid in the different states of nature can then be used to “stochastically” and costlessly deter any incentive to lie. If agent A_1 does not report truthfully his type, the mechanism is designed so that he gets a negative expected payoff. Since this ingenious trick can be used for each agent simultaneously, the revelation of both agents’ types obtains at no cost for the principal.

Obviously, the taxes paid by the agents may be quite large when types become almost uncorrelated.¹⁵ Indeed, the conditional probability that agents have the same types diminishes and the punishments (or rewards) in the different states of nature must increase to induce revelation.

We provide thereafter a simple proof of this first best implementation in our public good setting. To do that, we first need to describe the set of implementable allocations in this non-cooperative context. Without collusion, the revelation principle yields a characterization of the set of implementable allocations with individual incentive compatibility constraints only. Consider the Bayesian incentive compatibility constraint of agent A_1 when he has a high valuation $\bar{\theta}$ for the public good.¹⁶ Multiplying it by $p(\bar{\theta}) = p_{11} + p_{12} > 0$, this constraint writes as:

$$p_{11}(-t_{11} + \bar{\theta}\bar{x}) + p_{12}(-t_{12} + \bar{\theta}\hat{x}) \geq p_{11}(-t_{21} + \bar{\theta}\hat{x}) + p_{12}(-t_{22} + \bar{\theta}\underline{x}), \quad (1)$$

Similarly, when A_1 has a low valuation $\underline{\theta}$ for the public good, his Bayesian incentive compatibility constraint writes as (again multiplying by $p(\underline{\theta}) = p_{12} + p_{22} > 0$):

$$p_{12}(-t_{21} + \underline{\theta}\hat{x}) + p_{22}(-t_{22} + \underline{\theta}\underline{x}) \geq p_{12}(-t_{11} + \underline{\theta}\bar{x}) + p_{22}(-t_{12} + \underline{\theta}\hat{x}), \quad (2)$$

Moreover, A_1 must be induced to participate to the mechanism without knowing A_2 ’s type. The following interim participation constraints must also be satisfied: For a $\bar{\theta}$ agent

$$p_{11}(-t_{11} + \bar{\theta}\bar{x}) + p_{12}(-t_{12} + \bar{\theta}\hat{x}) \geq 0, \quad (3)$$

and for a $\underline{\theta}$ agent

$$p_{12}(-t_{21} + \underline{\theta}\hat{x}) + p_{22}(-t_{22} + \underline{\theta}\underline{x}) \geq 0. \quad (4)$$

The principal maximizes expected welfare defined as:

$$\begin{aligned} SW = & p_{11} \left(2\bar{\theta}\bar{x} - 2t_{11} + (1+\lambda)(2t_{11} - c(\bar{x})) \right) + 2p_{12} \left((\bar{\theta} + \underline{\theta})\hat{x} - t_{12} - t_{21} + (1+\lambda)(t_{12} + t_{21} - c(\hat{x})) \right) \\ & + p_{22} \left(2\underline{\theta}\underline{x} - 2t_{22} + (1+\lambda)(2t_{22} - c(\underline{x})) \right) \\ & \text{subject to constraints (1) to (4).}^{17} \end{aligned}$$

Proposition 1 : *Assume that types are strictly positively correlated, $\rho > 0$, then the optimal provision of public good without collusive behavior entails:*

- *The first-best decision rule $(\bar{x}^*, \hat{x}^*, \underline{x}^*)$ where $c'(\bar{x}^*) = 2\bar{\theta}$, $c'(\hat{x}^*) = \underline{\theta} + \bar{\theta}$ and $c'(\underline{x}^*) = 2\underline{\theta}$.*
- *Participation constraints (3) and (4) are binding and both agent’s types get zero rent.*

¹⁵See the Appendix for explicit formulas of these taxes.

¹⁶By symmetry A_2 faces the same incentive and participation constraints.

Proof: All proofs are in an Appendix.

To implement the first-best outcome, the participation constraints (3) and (4) must be binding and we look for solutions such that the incentive constraints (1) and (2) are also binding.¹⁸ Then the existence of some strictly positive correlation imposes that the system of linear binding constraints (1)-to-(4) is in fact invertible. This invertibility ensures that the first-best schedule of public goods can be implemented at zero cost for the principal

The taxes which implement this optimal allocation of public good are highly dependent on the information structure. When correlation becomes weaker, taxes become increasingly punishing when both agents claim $\bar{\theta}$ and when the agent who pays the tax claims $\underline{\theta}$ and the other claims $\bar{\theta}$. On the contrary, agent A_i is increasingly rewarded when both agents claim being $\underline{\theta}$ and when he claims $\bar{\theta}$ and A_j ($j \neq i$) claims $\underline{\theta}$. Indeed, with positive correlation, it becomes easier to induce revelation from a $\bar{\theta}$ agent A_i if, when he lies, he is heavily punished when facing a $\bar{\theta}$ agent A_j ($j \neq i$) who truthfully reveals.

Even a very small amount of correlation can be used to threaten the agents of being heavily punished for lying on their types and to achieve the first best allocation. However, with no correlation, allocative distortions become necessary to reduce a $\bar{\theta}$ agent's costly informational rent. Therefore, in this non-cooperative setting, the optimal mechanism fails to be continuous with respect to the information structure.

For notational convenience, let \hat{p} denote the probability of a high valuation type in the case of no correlation. We have thus $p_{11} = \hat{p}^2$, $p_{22} = (1 - \hat{p})^2$ and $p_{12} = p_{21} = \hat{p}(1 - \hat{p})$.

Proposition 2 : *Assume that types are independently distributed, $\rho = 0$, then the optimal provision of public good without collusive behavior entails:*

- *The second-best decision rule $(\bar{x}_0^*, \hat{x}_0^*, \underline{x}_0^*)$ where $c'(\bar{x}_0^*) = 2\bar{\theta}$, $c'(\hat{x}_0^*) = \underline{\theta} + \bar{\theta} - \frac{\lambda}{1+\lambda} \frac{\hat{p}}{1-\hat{p}} \Delta\theta$ and $c'(\underline{x}_0^*) = 2\underline{\theta} - 2\frac{\lambda}{1+\lambda} \frac{\hat{p}}{1-\hat{p}} \Delta\theta$.*
- *A $\bar{\theta}$ (resp. $\underline{\theta}$) agent gets a strictly positive (resp. zero) rent.*

Without any correlation, the system of binding equations (1) to (4) can no more be inverted. Only expected taxes are defined from the binding incentive (1) and participation constraints (4) and a $\bar{\theta}$ agent's informational rent becomes costly. Allocative distortions of \hat{x} and \underline{x} are needed to reduce this cost.

4 Collusion

Because the agents get zero rent from the mechanism proposed by the principal if they play non-cooperatively, they are willing to coordinate their messages to countervail the principal's power. The optimal grand-mechanism with a non-cooperative behavior creates endogenously the stakes for some collusive behavior.

¹⁸Cr mer and McLean (1988) show in fact that incentive constraints can be slack.

4.1 The Third-Party's Problem

Following Cramton and Palfrey (1995), we say that a side-mechanism S is *unanimously ratified* for $(e_1, e_2, \tilde{p}_1, \tilde{p}_2)$ if, for all $\theta_i \in \Theta$ and all i ,

$$U_i(\theta_i) \geq U_i(\theta_i, e_i)$$

where $e_i \in E(G, \tilde{p}_i, p_{-i})$ is a non-cooperative equilibrium of $\Gamma(G, \tilde{p}_i, p_{-i})$ and where

$$U_i(\theta_i) = \sum_{\Theta_{-i}} p(\theta_{-i} | \theta_i) \left(y_i(\theta_i, \theta_{-i}) - t_i(\phi(\theta_i, \theta_{-i})) + \theta_i x(\phi(\theta_i, \theta_{-i})) \right), \forall \theta_i \in \Theta;$$

is a θ_i agent A_i 's payoff from playing $e^* \in \Gamma(G \circ S, p, p)$.

Definition 1 : *A continuation collusive equilibrium of the game of coalition formation consists of first, a system of beliefs $\{\tilde{p}_1, \tilde{p}_2\}$ and associated equilibria $e_i \in \Gamma(G, \tilde{p}_i, p_{-i})$ ($i \in \{1, 2\}$) and second, a truthtelling direct side-mechanism S^* which is unanimously ratified for the quadruplet $(e_1, e_2, \tilde{p}_1, \tilde{p}_2)$ and which maximizes the third party's objective function.*

The continuation equilibrium of the game of coalition formation is thus a solution to the third-party's following problem (denoted thereafter (T)):

$$\max_{\{\phi(\cdot), y_k(\cdot)_{(k \in \{1, 2\})}\}} \sum_{(i, j) \in \{1, 2\}^2} p_{ij} \left(-t_1(\phi(\theta_i, \theta_j)) - t_2(\phi(\theta_i, \theta_j)) + (\theta_i + \theta_j)x(\phi(\theta_i, \theta_j)) \right)$$

subject to

$$\text{(BIC)} \quad U_i(\theta_i) \geq \sum_{\Theta_{-i}} p(\theta_{-i} | \theta_i) \left(y_i(\hat{\theta}_i, \theta_{-i}) - t_i(\phi(\hat{\theta}_i, \theta_{-i})) + \theta_i x(\phi(\hat{\theta}_i, \theta_{-i})) \right), \forall (\theta_i, \hat{\theta}_i) \in \Theta^2;$$

$$\text{(BIR)} \quad U_i(\theta_i) \geq U_i(\theta_i, e_i) \text{ for some } e_i \in E(G, \tilde{p}_i, p_{-i}) \quad \forall \theta_i \in \Theta;$$

$$\sum_{k=1}^2 y_k(\theta_i, \theta_{-i}) = 0, \quad \forall (\theta_i, \theta_{-i}) \in \Theta^2.$$

Along an equilibrium path on which the side-contract S^* is unanimously ratified, no learning occurs since unanimous ratification is a pooling strategy. Since nothing is revealed at the ratification stage, Bayesian incentive constraints and final expected utilities are computed with passive beliefs.

4.2 Weakly Collusion-Proof Mechanisms

Definition 2 : *G is weakly collusion-proof if and only if it is a truthtelling direct mechanism and the null side-mechanism $S_0^* = \{\phi^* = Id, (y_k^* = 0)_{k \in \{1, 2\}}\}$ is unanimously ratified for (e^*, e^*, p, p) , where e^* is the truthful equilibrium of G played with passive beliefs.*

In other words, G is weakly collusion-proof if and only if the third-party offers the null side-mechanism and there exists a perfect Bayesian equilibrium in $\hat{E}(G, S_0^*)$ such that agents accept to play e^* sustained with passive beliefs out of the equilibrium path.

Proposition 3 : *To characterize the outcome of any perfect Bayesian equilibrium of the game of grand-mechanism offer cum coalition formation such that a collusive continuation equilibrium occurs on the equilibrium path, there is no loss of generality in restricting the principal to offer weakly collusion-proof mechanisms.*

The logic behind this *weak collusion-proofness principle* is similar to that underlying the standard revelation principle: any equilibrium of the overall game of grand-mechanism offer cum side-contracting gives an allocation which can be replicated with a direct grand-mechanism G offered by the principal himself. This grand-mechanism is such that the coalition still forms at the ratification stage of $\hat{\Gamma}(G, S_0^*)$ and each agent A_i finds optimal to report his valuation θ_i truthfully to the principal thereafter.

One could argue that agents' information is not only restricted to their own valuations but also includes the knowledge of the side-mechanism they use to collude. The principal could try to elicit this information by also asking the agents which side-mechanism they are actually playing. But the third-party could react by inducing further manipulations of those reports of the side-mechanism. These reactions and counterreactions lead naturally to a problem of infinite regress. By restricting the principal to use grand-mechanisms only contingent on the agents' valuations, we cut arbitrarily this process in favor of the colluding agents. This amounts to an incomplete contracting assumption which fits our desire to give to collusive behavior its best chance.¹⁹ Within this incompleteness contractual framework, we are nevertheless able to generalize the revelation principle and benefit from the weak collusion-proofness principle to obtain a simple constructive characterization of the set of implementable allocations.

The next proposition characterizes this set in the case of symmetric grand-mechanisms.

Proposition 4 : *A symmetric grand-mechanism G is weakly collusion-proof if and only if there exists $\epsilon \in [0, 1[$ such that:*

$$-2t_{11} + 2\bar{\theta}\bar{x} \geq -t_1(\tilde{\theta}_1, \tilde{\theta}_2) - t_2(\tilde{\theta}_1, \tilde{\theta}_2) + 2\bar{\theta}x(\tilde{\theta}_1, \tilde{\theta}_2) \quad \forall(\tilde{\theta}_1, \tilde{\theta}_2) \in \Theta^2; \quad (5)$$

$$-t_{12} - t_{21} + \left(\bar{\theta} + \underline{\theta} - \frac{p_{11}}{p_{12}}\epsilon\Delta\theta\right)\hat{x} \geq -t_1(\tilde{\theta}_1, \tilde{\theta}_2) - t_2(\tilde{\theta}_1, \tilde{\theta}_2) + \left(\bar{\theta} + \underline{\theta} - \frac{p_{11}}{p_{12}}\epsilon\Delta\theta\right)x(\tilde{\theta}_1, \tilde{\theta}_2) \quad \forall(\tilde{\theta}_1, \tilde{\theta}_2) \in \Theta^2; \quad (6)$$

$$-2t_{22} + 2\left(\underline{\theta} - \frac{\epsilon p_{12}^2}{p_{12}p_{22} + \epsilon(p_{11}p_{22} - p_{12}^2)}\Delta\theta\right)\hat{x}$$

¹⁹In a related multiprincipal context, Epstein and Peters (1997) show that a similar infinite regress may converge. This convergence would allow one to define universal sets of types and universal complete grand-mechanisms.

$$\geq -t_1(\tilde{\theta}_1, \tilde{\theta}_2) - t_2(\tilde{\theta}_1, \tilde{\theta}_2) + 2 \left(\underline{\theta} - \frac{\epsilon p_{12}^2}{p_{12}p_{22} + \epsilon(p_{11}p_{22} - p_{12}^2)} \Delta\theta \right) x(\tilde{\theta}_1, \tilde{\theta}_2) \quad \forall (\tilde{\theta}_1, \tilde{\theta}_2) \in \Theta^2; \quad (7)$$

If $\epsilon > 0$, the Bayesian incentive compatibility constraint (1) of a $\bar{\theta}$ agent is binding.

The weakly collusion-proof mechanisms described above are such that a $\underline{\theta}$ agent's incentive constraint is not binding. Only a $\bar{\theta}$ agent's incentive constraint may be binding. These mechanisms are the only ones which are of interest as Proposition 5 will confirm.

The parameter ϵ is a discount factor less than one which captures the fact that collusion takes place under asymmetric information. True valuations must be replaced by *virtual* valuations²⁰ in the *coalition incentive constraints* (6) and (7). The third-party problem (T) is constrained by the reservation utilities that the agents obtain from playing non-cooperatively the grand-mechanism and the incentive compatibility constraint at the coalition formation stage. Virtual valuations are then lower than true valuations to take into account the costly multipliers of these constraints.

In the sequel, the downward coalition incentive constraints (rewritten after having used the symmetry of the grand-mechanism) are of particular interest:

$$-2t_{11} + 2\bar{\theta}\bar{x} \geq -t_{12} - t_{21} + 2\bar{\theta}\hat{x}, \quad (8)$$

$$-t_{12} - t_{21} + \left(\bar{\theta} + \underline{\theta} - \frac{p_{11}}{p_{12}} \epsilon \Delta\theta \right) \hat{x} \geq -2t_{22} + \left(\bar{\theta} + \underline{\theta} - \frac{p_{11}}{p_{12}} \epsilon \Delta\theta \right) \underline{x}. \quad (9)$$

(8) says that a coalition made with two $\bar{\theta}$ agents prefers telling collectively the truth to the principal rather than lying and telling that one of the agents is $\underline{\theta}$. (9) says that a $(\bar{\theta}, \underline{\theta})$ coalition prefers telling the truth rather than claiming that both agents are $\underline{\theta}$.

The logic of the first-best non-cooperative implementation described in Section 2 is to offer large penalties and large rewards depending on the states of nature to induce revelation. For instance, the formula for t_{11} (resp. t_{22}) in the Appendix shows that this tax may become extremely large and positive (resp. negative) when the agents' types are less and less correlated. This suggests that the coalition incentive constraint (8) (resp. (9)) is likely to be binding at the optimum of the principal's problem. The extent to which the principal is restricted in using Bayesian transfers comes from the existence of these coalition incentive constraints.

Once coalition incentive constraints are characterized, it is useful to derive necessary monotonicity conditions for the implementability of a schedule of outputs.

Corollary 1 : For a weak correlation, $\rho \leq (p_{12} + p_{22}) \frac{p_{12}^2}{p_{11}}$, the schedule of implementable outputs is increasing ($\underline{x} \leq \hat{x} \leq \bar{x}$) for all $\epsilon \in [0, 1[$. For a strong correlation, $\rho > (p_{12} + p_{22}) \frac{p_{12}^2}{p_{11}}$, the schedule of implementable outputs is non-monotonic ($\bar{x} \geq \underline{x} \geq \hat{x}$) if and only if $\psi(\epsilon) = 1 + 2 \frac{\epsilon p_{12}^2}{p_{12}p_{22} + \epsilon(p_{11}p_{22} - p_{12}^2)} - \frac{p_{11}}{p_{12}} \epsilon < 0$, i.e., for ϵ large enough, otherwise it remains increasing.

²⁰See Myerson (1979) for the standard definition of virtual valuations.

The striking feature of this corollary is that, in the case of strong correlation (i.e., when p_{12} is small enough), non-monotonic schedules of outputs can be implemented by the principal if he chooses ϵ large enough. The reason for this non-monotonicity is that virtual valuations coming from the formation of the coalition under asymmetric information are no longer ranked in the same order as true valuations. When agents' types are almost perfectly correlated, the probability that they both get a high-valuation for the public good is large. Henceforth, the incentive constraint of a $\bar{\theta}$ agent willing to mimic a $\underline{\theta}$ one at the coalition formation stage is also very costly to the third-party from an ex ante point of view. Inducing revelation within the coalition requires therefore large distortions of the optimal manipulation of reports $\phi^*(\underline{\theta}, \bar{\theta})$ with respect to what the third-party could implement under symmetric information. Hence, the sum of the virtual valuations of a $(\underline{\theta}, \bar{\theta})$ coalition may become smaller than that of a $(\underline{\theta}, \underline{\theta})$ one.

This non-monotonicity property will create a substantive difference between collusion under symmetric and asymmetric information when the correlation is strong contrary to the case of a weak correlation.

5 The Optimal Weakly Collusion-Proof Mechanism

We now turn to some normative analysis and optimize the principal's welfare subject to Bayesian individual incentive, participation and coalition incentive constraints. In the sequel, we focus only on the $\bar{\theta}$ agent's Bayesian incentive constraint (1), the downward coalition incentive constraints (8) and (9) and the $\underline{\theta}$ agent's participation constraint (5).²¹

Before writing the principal's problem, let us first introduce four new variables $\bar{u} = -t_{11} + \bar{\theta}\bar{x}$, $\hat{u}_1 = -t_{12} + \bar{\theta}\hat{x}$, $\hat{u}_2 = -t_{21} + \underline{\theta}\hat{x}$, $\underline{u} = -t_{22} + \underline{\theta}\underline{x}$. These are the *ex post rents* obtained by both types of agent in each possible state of nature. Rearranging expected social welfare, individual and coalitional incentive and participation constraints as functions of these variables, the principal's problem (P) rewrites as:

$$\max_{\{x(\cdot); u\}} p_{11} \left((1 + \lambda)(2\bar{\theta}\bar{x} - c(\bar{x})) - 2\lambda\bar{u} \right) + 2p_{12} \left((1 + \lambda)((\bar{\theta} + \underline{\theta})\hat{x} - c(\hat{x})) - \lambda(\hat{u}_1 + \hat{u}_2) \right)$$

$$+ p_{22} \left((1 + \lambda)(2\underline{\theta}\underline{x} - c(\underline{x})) - 2\lambda\underline{u} \right),$$

subject to

$$\text{(BIC)} \quad p_{11}\bar{u} + p_{12}\hat{u}_1 \geq p_{11}\hat{u}_2 + p_{12}\underline{u} + \Delta\theta(p_{11}\hat{x} + p_{12}\underline{x}) \quad (10)$$

$$\text{(CIC)}_1 \quad 2\bar{u} \geq \hat{u}_1 + \hat{u}_2 + \Delta\theta\hat{x} \quad (11)$$

$$\text{(CIC)}_2 \quad \hat{u}_1 + \hat{u}_2 \geq 2\underline{u} + \Delta\theta\underline{x} + \frac{p_{11}}{p_{12}}\epsilon\Delta\theta(\hat{x} - \underline{x}) \quad (12)$$

$$\text{(IR)} \quad p_{12}\hat{u}_2 + p_{22}\underline{u} \geq 0. \quad (13)$$

²¹We check ex post that all other constraints are indeed satisfied.

- For a weak correlation, $\rho \leq (p_{12} + p_{22})\frac{p_{12}^2}{p_{11}}$, the monotonicity condition derived from the coalition incentive constraints is

$$\bar{x} \geq \hat{x} \geq \underline{x}.$$

- For a strong correlation, $\rho \geq (p_{12} + p_{22})\frac{p_{12}^2}{p_{11}}$, the monotonicity condition becomes

$$\bar{x} \geq \hat{x} \geq \underline{x} \text{ if and only if } \psi(\epsilon) \geq 0 \text{ and } \bar{x} \geq \underline{x} \geq \hat{x} \text{ otherwise.}$$

In the sequel, we focus separately on these two polar cases of weak and strong correlations.

5.1 Weak Correlation

Solving the principal's problem with standard Lagrangean technics yields:

Proposition 5 : *Assuming that $0 \leq \rho \leq (p_{12} + p_{22})\frac{p_{12}^2}{p_{11}}$, the symmetric optimal weakly collusion-proof mechanism G^* entails:*

- *A strictly decreasing schedule of outputs $\bar{x}^c > \hat{x}^c > \underline{x}^c$ with “no distortion at the top” $\bar{x}^c = \bar{x}^*$ and downward distortions with respect to the no collusion outcome otherwise: $\hat{x}^c < \hat{x}^*$ and $\underline{x}^c < \underline{x}^*$ where*

$$c'(\hat{x}^c) = \bar{\theta} + \underline{\theta} - \frac{\lambda}{1 + \lambda} \Delta\theta \frac{p_{11}}{2p_{12}} \left(1 + \frac{p_{12}}{p_{12} + \rho} \right) \quad (14)$$

$$c'(\underline{x}^c) = 2\underline{\theta} - \frac{\lambda}{1 + \lambda} \Delta\theta \frac{1}{p_{22}} \left(p_{11} + 2p_{12} - \frac{p_{12}p_{11}}{p_{12} + \rho} \right). \quad (15)$$

- *The Bayesian incentive constraint of a $\bar{\theta}$ agent (10) is always binding. The downward coalition incentive constraints (11) and (12) are both strictly binding when $\rho > 0$. The participation constraint of a $\underline{\theta}$ agent (13) is also binding. All remaining constraints are strictly satisfied. A $\bar{\theta}$ agent gets a strictly positive informational rent.*

Binding Constraints: The fact that both coalitions $(\bar{\theta}, \bar{\theta})$ and $(\underline{\theta}, \bar{\theta})$ are prevented from misreporting limits the feasible transfers that could be used by the principal to extract the agents' information. t_{11} cannot be made largely positive as it is in the no-collusion outcome without violating the coalition incentive constraint (11). A $(\bar{\theta}, \bar{\theta})$ coalition would like to avoid bearing these detrimental punishments by mimicking a $(\bar{\theta}, \underline{\theta})$ coalition. (11) must be binding at the optimum. Similarly, a $(\bar{\theta}, \underline{\theta})$ coalition would like to mimic a $(\underline{\theta}, \underline{\theta})$ one to get the corresponding large rewards requested in the no-collusion outcome since t_{22} is then large and negative. (12) must thus also be binding.

Because of these constraints on the set of transfers, a $\bar{\theta}$ agent must be given a strictly positive rent contrary to the case without collusion. Large rewards and punishments can no longer be used by the principal without violating the coalition incentive constraints.

Coalition Incentive Constraints: For a weak correlation, $\epsilon = 0$ at the optimum. Indeed, there is no gain in having ϵ strictly positive since this would only increase the cost

of the coalition incentive constraint (12).²² Interestingly, the binding collusion-proofness constraints take therefore the same form as if agents could credibly share their information within the coalition. Everything happens as if asymmetric information does not really undermine the ability of the group to form. However, the agents' participation constraints being the interim ones, they differ from the case of symmetric information within the coalition.

Output Distortions: Distorting downward \hat{x} and \underline{x} below their first-best values reduces the costs of the individual (10) and coalitional incentive (11) and (12) constraints. The size of the public good must be reduced because of asymmetric information and collusion. Note that this result contrasts with the usual free-rider phenomenon discussed in the literature. As shown for instance in Mailath and Postlewaite (1990), the reduction in the size of the public good is then due to the conflict between individual incentive and individual participation constraints. Here, distortions come from the conflict between coalitional incentive and individual participation constraints. It is because a coalition can form that individual incentives become costly to provide and that allocative distortions are needed in this correlated environment.

Role of the Correlation: This conflict between coalition incentive and participation constraints increases with the amount of positive correlation between the agents' types. When types are more positively correlated, the distortion required on \hat{x}^c to reduce the cost of the coalition incentive constraint (11) is larger since p_{12} is smaller (see (14)). The distortion on \bar{x}^c needed to reduce the cost of the coalition incentive constraint (12) is instead rather small since p_{22} is now relatively large (see (15)).

Ex post Rents: Ex post rents in the optimal weakly collusion-proof mechanism satisfy:

$$\bar{u} < \hat{u}_2 + \Delta\theta\hat{x}^c \tag{16}$$

$$\hat{u}_1 > \underline{u} + \Delta\theta\underline{x}^c \tag{17}$$

$$\hat{u}_2 < 0 \tag{18}$$

$$\bar{u} > \hat{u}_1 \tag{19}$$

$$\underline{u} > 0 \tag{20}$$

$$\hat{u}_2 > \bar{u} - \Delta\theta\bar{x}^c \tag{21}$$

$$\underline{u} > \hat{u}_1 - \Delta\theta\hat{x}^c. \tag{22}$$

(16) and (17) indicate that the truthful strategy of a $\bar{\theta}$ agent is Bayesian. Indeed (16) shows that a $\bar{\theta}$ agent A_1 never wants to tell the truth if he is sure of facing a $\bar{\theta}$ agent A_2 . Instead, (17) shows that he always tells the truth if he is sure that A_2 claims $\underline{\theta}$. Contrary

²²In other words, the Bayesian incentive constraint in the third-party problem (T) is not tight at the optimum. This does not mean that (10) is not costly for the principal since he has a different objective function than that of the third-party.

to what happens in the no-collusion outcome where in fact $\bar{u} < \hat{u}_1$, a $\bar{\theta}$ agent is now rewarded when he faces a $\bar{\theta}$ agent A_2 and punished when he faces a $\underline{\theta}$ agent A_2 . It is only because these rewards cover in expectation the penalties that a $\bar{\theta}$ agent A_1 weakly prefers to tell the truth to the principal.

Instead, (21) and (22) show that the optimal weakly collusion-proof contract is such that the dominant strategy incentive constraints of a $\underline{\theta}$ agent are always strictly satisfied. This dominant strategy requirement for one type somewhat simplifies the optimal mechanism since this type's equilibrium strategy is not sensitive to the exact beliefs that agents have at the time of playing the mechanism.

Nevertheless, the ex post rent of a $\underline{\theta}$ agent A_1 depends also explicitly on A_2 's type. For instance, just as in the no collusion outcome, two $\underline{\theta}$ agents are given strictly positive ex post rents. These agents are still subsidized for the consumption of the public good. On the contrary, a $\underline{\theta}$ agent A_1 facing a $\bar{\theta}$ agent A_2 makes a negative ex post profit. The tax he pays for the public good is very large and his final utility is negative. Intuitively, by setting \hat{u}_2 negative, the principal insures that (11) is not very costly. Satisfying the $\underline{\theta}$ agent's interim participation constraint requires then to set \underline{u} strictly above zero.

5.2 The Polar Case of No Correlation

In the degenerate case of no correlation, a rather striking result obtains:

Proposition 6 : *With no correlation between the agents' types ($\rho = 0$), the optimal weakly collusion-proof grand-mechanism G^* entails the same strictly decreasing schedule of outputs as without collusion, $\bar{x}_0^c = \bar{x}_0^*$, $\hat{x}_0^c = \hat{x}_0^*$ and $\underline{x}_0^c = \underline{x}_0^*$.*

Only the individual incentive (10) and the participation constraints (13) are strictly binding. The multipliers of the coalition incentive constraints (11) and (12) are zero at the optimum of (P) . Non-anonymous transfers implement in a Bayesian and weakly collusion-proof way the optimal second-best contract, i.e., the non-cooperative outcome obtained when there is no correlation between the agents' types. Collusion has no impact when agents do not know more information on each other than what is available to the principal. Intuitively, everything happens as if the principal sells the mechanism to the third-party so that the latter perfectly internalizes the social welfare objective.

Strikingly, the optimal policy moves *continuously* from the case with positive correlation to the case without correlation when coalition incentive constraints are taken into account. Allowing collusion restores continuity of the optimal contract with respect to the information structure. For any degree of positive correlation, there are enough binding constraints to pin down the values of the optimal transfers in all states of nature. When the degree of correlation converges to zero, these transfers converge and the limiting transfers implement the second-best outcome in a collusion-proof way. Not only the expected

values of these limiting transfers but also their precise values in all states of nature are now perfectly determined even in the case of no correlation.²³

5.3 The Polar Case of Almost Perfect Correlation

With an almost perfect correlation, (p_{12} close to zero), both agents have almost always the same type.

As a benchmark, suppose first that the agents can credibly exchange information on their types because, for instance, the third-party is endowed with a technology making this information verifiable within the coalition. Assume also that agents agree to play the grand-mechanism before they learn each other types so that their participation constraints remain unchanged with respect to 5.1. Collusion-proofness constraints must now be written with true valuations instead of virtual ones. This has two implications: First, the binding downward coalition-proofness constraints take the same form as (11) and (12) when $\epsilon = 0$. Second, monotonicity of the schedule of outputs ($\bar{x} \geq \hat{x} \geq \underline{x}$) is now always necessary for implementability.

If this monotonicity constraint would not be binding at the optimum, the optimal levels of public good would still be given by (14) and (15). However, when p_{12} is small enough, it is easy to check that the monotonicity constraint $\hat{x} \geq \underline{x}$ would be violated. Hence, we get:

Proposition 7 : *With almost perfect correlation between the agents' types (p_{12} small enough but positive) and symmetric information within the coalition, the optimal weakly collusion-proof grand-mechanism G^* entails partial pooling $\bar{x}^c = \bar{x}^* > \hat{x}^c = \underline{x}^c = x^{cP}$ with x^{cP} defined by:*

$$c'(\hat{x}^{cP}) = 2 \frac{p_{22}\underline{\theta} + p_{12}(\underline{\theta} + \bar{\theta})}{2p_{12} + p_{22}} - 2 \frac{\lambda}{1 + \lambda} \Delta\theta \left(\frac{p_{11} + p_{12}}{2p_{12} + p_{22}} \right). \quad (23)$$

Taking into account collusion-proofness constraints under symmetric information in an almost perfectly correlated environment undermines quite significantly the achievement of the efficient outcome. The optimal allocation entails now lots of pooling. When the agents' types are strongly correlated, we have seen in Section 2 that rewards and punishments necessary to implement the first-best outcome are relatively small. However, when agents collude under symmetric information, their collusion becomes now also harder to prevent. As a result, the optimal collusion-proof mechanism becomes less responsive to their messages. In this framework, the optimal contract looks like an incomplete contract making only partial use of the information.²⁴

²³Coalition incentive constraints somehow compactify the set of feasible allocations just as risk-aversion or limited liability does (see Robert (1991)).

²⁴A last interpretation is worth stressing. Assume that the agents' types are now perfectly correlated ($p_{12} = 0$). The optimal pooling allocation becomes $c'(x^{cP}) = 2\underline{\theta} - 2 \frac{\lambda}{1+\lambda} \frac{p_{11}}{p_{22}} \Delta\theta$. This is the optimal

Things are quite different under asymmetric information within the coalition. For a strong correlation, non-monotonic schedules of outputs can be implemented by the principal if ϵ is large enough. Because now $\hat{x} < \underline{x}$, the collusion-proofness constraint (12) is relaxed when ϵ is as large as possible. An upper bound of the principal's welfare obtains therefore for $\epsilon = 1$.

Proposition 8 : *With almost perfect correlation and asymmetric information within the coalition, the optimal weakly collusion-proof grand-mechanism G^* entails: $\bar{x}^c = \bar{x}^* > \underline{x}^c > \hat{x}^c = 0$ with \underline{x}^c defined by:*

$$c'(\underline{x}^c) = 2\theta - \frac{\lambda}{1 + \lambda} \frac{2p_{12}}{p_{22}} \Delta\theta.$$

The expected rent of the $\bar{\theta}$ agent is $\frac{p_{11}\bar{u} + p_{12}\hat{u}_1}{p_{11} + p_{12}} = \frac{p_{12}}{p_{11} + p_{12}} \Delta\theta \underline{x}^c$. Moreover, $\hat{u}_2 = \underline{u} = 0$, $\hat{u}_1 = \left(1 - \frac{p_{11}}{p_{12}}\right) \Delta\theta \underline{x}^c$ and $\bar{u} = \Delta\theta \underline{x}^c$.

For a strong correlation, the principal can now offer non-monotonic schedules of outputs in a collusion-proof way. In particular, cancelling the production of the public good when agents' types are different and still keeping a positive production when types are the same becomes a valuable option. This strategy is not very costly from an ex ante allocative point of view since p_{12} is small. However, it relaxes quite significantly (12) when ϵ is positive. In the limiting case where p_{12} is almost zero, the R.H.S. of (12) is so largely negative that this coalition incentive constraint does not matter any more for the principal. \hat{u}_1 can be set at a very large negative value and still this constraint can be easily satisfied. Choosing such a punishment in case of different reports also relaxes quite significantly (11) and the principal's problem is almost as without collusion.

Since collusion does not affect too much social welfare, the principal can set \underline{x}^c to a level close to its first best value \underline{x}^* . The punishment for claiming to have different types are so effective that a $\bar{\theta}$ agent's expected rent is now close to zero. Therefore, with almost perfect correlation, the optimal collusion-proof contract achieves an ex ante social welfare close to its full information value.

6 Robustness

We focus in this section on the case of a small correlation.

6.1 Multiplicity of Equilibria of G^*

When played with passive beliefs, the optimal weakly collusion-proof mechanism G^* has several pure strategy equilibria, i.e., $|E(G^*, p, p)| > 1$. The first one is the truthful sym-

distortion in a one principal-agent model where the principal is facing directly a third-party endowed with a utility function being the sum of the agents' utility functions.

metric equilibrium e^* that we have derived above. However, there exist also two asymmetric non-truthful pure strategy equilibria. e_1^* (resp. e_2^*) is one such equilibrium where A_1 (resp. A_2) always claims to be a $\underline{\theta}$ agent and A_2 (resp. A_1) reveals truthfully his type.

Indeed, from (21) and (22), a $\underline{\theta}$ agent always claims truthfully his type. A $\bar{\theta}$ agent A_2 anticipates that A_1 always claims $\underline{\theta}$. Then, from (17), he reports truthfully. Thus, agent A_2 always reveals his type. Anticipating that A_2 always reveal his type, from the binding constraint (10), a $\bar{\theta}$ agent A_1 is indifferent between lying or not and lies in this equilibrium.

In this asymmetric equilibrium, denoted thereafter e_1^* , A_1 gets the same ex ante and interim payoffs as in e^* , $U_1(\theta_1, e_1^*) = U(\theta_1)$ for all θ_1 . Instead, A_2 's interim payoffs differ from that in e^* . Indeed, we have $U_2(\bar{\theta}, e_1^*) = \hat{u}_1 < U(\bar{\theta}) = \frac{p_{11}\bar{u} + p_{12}\hat{u}_1}{p_{11} + p_{12}}$ (from (19)) and $U_2(\underline{\theta}, e_1^*) = \underline{u} > U(\underline{\theta}) = 0$ (from (20)). However, using the values of ex post rents given in the Appendix, A_2 's ex ante payoff is the same as in e^* .

In fact, in equilibrium e_1^* , a $(\bar{\theta}, \underline{\theta})$ coalition reports $(\underline{\theta}, \underline{\theta})$ when it is indifferent between collectively claiming $(\bar{\theta}, \underline{\theta})$ and $(\underline{\theta}, \underline{\theta})$ ((12) is indeed binding for the ex post rents defined by G^*). This indifference of the coalition is broken in favor of the principal so that, among all payoff equivalent equilibria from the third-party's point of view, the agents play the most preferred by the principal. The fact that $E(G^*, p, p)$ is not a singleton is not a big problem since all these equilibria yield the same aggregate payoff to the agents.²⁵

6.2 Strong Collusion-Proofness

The fact that $E(G^*, p, p)$ is not a singleton shows also that G^* is never *strongly collusion-proof*, i.e., $\hat{E}(G^*, S_0^*)$ is not reduced to a singleton either. Indeed, there exist other equilibria of $\hat{\Gamma}(G^*, S_0^*)$ than unanimous ratification and subsequent play of e^* . To show this result, it is enough to exhibit such an equilibrium of $\hat{\Gamma}(G^*, S_0^*)$. Unanimous veto of S_0^* and subsequent play of the equilibrium e_1^* is such an equilibrium of $\hat{\Gamma}(G^*, S_0^*)$ sustained with passive beliefs.

This observation suggests two things. First, the ratification stage enlarges the set of equilibria of G^* . Second, we should test the robustness of G^* to *credible* equilibria of the game of coalition formation. We now turn to this issue.

6.3 Strong Ratifiability

As we have seen above, the ratification stage of $\hat{\Gamma}(G^*, S_0^*)$ adds in fact a preplay communication stage to G^* . At this stage, agents are allowed to make binary preplay announcements ("Veto" or "Accept") which may signal some information on their types. Cramton

²⁵This contrasts with Ma, Moore and Turnbull (1988) where Pareto-dominant non-cooperative and non-truthful equilibria may be a threat to the principal and must therefore be eliminated by using indirect message games.

and Palfrey (1995) have analyzed similar mechanism design problems and have proposed the notion of *strong ratifiability* of a mechanism against itself to test its robustness to such a cheap-talk stage.²⁶ To understand this notion, let us first define *credible veto beliefs*:

Definition 3 : A belief system $\{\tilde{p}_1, \tilde{p}_2\}$ on Θ is a *credible veto system* of the truthful decision rule e^* if, for each i , there exist a non-cooperative Bayesian equilibrium $e_i \in E(G^*, \tilde{p}_i, p_{-i})$ (with $e_i \neq e^*$) and refusal probabilities $v_i(\theta_i)$, $\forall \theta_i \in \Theta, \forall i \in \{1, 2\}$ which together satisfy:

1. $v_i(\theta_i) > 0$ for some $\theta_i \in \Theta$;
2. $v_i(\theta_i) = 1$ for all $\theta_i \in \Theta$ such that $U(\theta_i) < U_i(\theta_i, e_i)$
3. $v_i(\theta_i) = 0$ for all $\theta_i \in \Theta$ such that: $U(\theta_i) > U_i(\theta_i, e_i)$
4. \tilde{p}_i satisfies Bayes' rule, given the prior distribution p and the refusal probabilities $v_i(\cdot)$:

$$\tilde{p}_i(\theta_i|\theta_j) = \begin{cases} \frac{p(\theta_i|\theta_j)v_i(\theta_i)}{\sum_{\theta_k \in \Theta_i, v(\theta_k) > 0} p(\theta_k|\theta_j)v_k(\theta_k)} & \text{for } \theta_i \text{ such that } v_i(\theta_i) > 0 \\ 0 & \text{for } \theta_i \text{ such that } v_i(\theta_i) = 0. \end{cases}$$

Credibility of beliefs requires that vetoing e^* should be interpreted as coming from the subset of types who are the most likely to benefit from this deviation when a non-deviant agent's beliefs put all weight on this particular subset. This definition captures the rational expectation reasoning that the different types of the deviant agent make when they envision vetoing the decision rule e^* . Those types who gain from vetoing e^* effectively deviate and refuse to play e^* with some probability when the non-deviant agent interprets these deviations in a rational way. The set of such types is called a *credible veto set*. Still following Cramton and Palfrey (1995), let also define:

Definition 4 : G^* is *strongly ratifiable* if and only if there does not exist a credible veto system or, if for all credible veto sets and all credible veto beliefs \tilde{p}_i , there exists a non-cooperative equilibrium $e_i \in E(G^*, \tilde{p}_i, p_{-i})$ such that $U(\theta_i) = U_i(\theta_i, e_i)$ for all i and for all θ_i belonging to the credible veto set.

Strong ratifiability captures the idea that no type of agent is willing to credibly veto the play of the truthful equilibrium e^* . If it is strongly ratifiable, G^* is robust to cheap-talk equilibria such that (possibly strict) subsets of types may deviate in a credible way.

²⁶See Matthews and Postlewaite (1989) and Palfrey and Srivastava (1991) for other analyses allowing more general cheap-talk stages in mechanism design.

Definition 3 is actually a step in the construction of perfect sequential equilibria made by Grossman and Perry (1986). Therefore, G^* is strongly ratifiable if and only if all perfect sequential equilibria of $\hat{\Gamma}(G^*, S_0^*)$ are payoff equivalent to unanimous ratification of e^* .

Proposition 9 : *The optimal weakly collusion-proof grand-mechanism G^* is strongly ratifiable.*

The proof consists in describing the equilibrium correspondence $E(G^*, \tilde{p}_1, p_2)$ when \tilde{p}_1 varies. From now on, we abuse notations and denote by \tilde{p}_1 (resp. p) the probability (resp. the conditional prior probability) that A_2 assigns to A_1 being $\bar{\theta}$. Starting from $E(G^*, p, p) = \{e^*, e_1^*, e_2^*\}$, we get $E(G^*, \tilde{p}_1, p) = \{e_2^*\}$ for optimistic beliefs $\tilde{p}_1 > p$ and $E(G^*, \tilde{p}_1, p) = \{e^*, e_1^*\}$ for pessimistic beliefs $\tilde{p}_1 < p$.

The characterization of this correspondence makes easier to isolate the veto sets which can credibly refuse to play e^* . A $\bar{\theta}$ agent A_1 cannot be a credible veto set since G^* has a unique equilibrium e_2^* when A_2 holds optimistic beliefs on A_1 and this equilibrium gives to a $\bar{\theta}$ agent A_1 strictly less utility than e^* . On the contrary, a $\underline{\theta}$ agent A_1 can be a credible veto set since G^* has e^* and e_1^* for equilibria when A_2 holds pessimistic beliefs on A_1 . However, by refusing to ratify e^* , this credible veto set cannot obtain more utility than following unanimous ratification of e^* since, with pessimistic beliefs, we have $E(G^*, \tilde{p}_1, p) = \{e^*, e_1^*\}$ and both equilibria yield payoff $U_1(\underline{\theta}, e_1^*) = U(\underline{\theta})$ to a $\underline{\theta}$ agent A_1 anyway.

7 Conclusion

When agents collude to influence collective decision on public goods and their valuations for the public good are positively correlated, there exists a trade-off between efficiency and rent extraction. The optimal weakly collusion-proof contract depends on the degree of correlation.

This correlation is also a crucial determinant of the group's ability to collude. In particular, a strong positive correlation allows the principal to use asymmetric information within the coalition to undermine significantly its countervailing power.

Adding coalition incentive constraints restores also continuity between the correlated and the uncorrelated information environments. Collusion does not matter in an uncorrelated environment with risk-neutral agents and the non-cooperative outcome is implementable in a collusion-proof way.

The benefit from focusing on a discrete two type modeling of asymmetric information is twofold. First, it has first given us a tractable characterization of the set of collusion-proofness constraints. Second, it is also a key simplification to test the robustness of the optimal mechanism to some form of preplay communication since it becomes then

possible to fully describe the equilibrium correspondence of the optimal weak collusion-proof mechanism G^* when beliefs change.

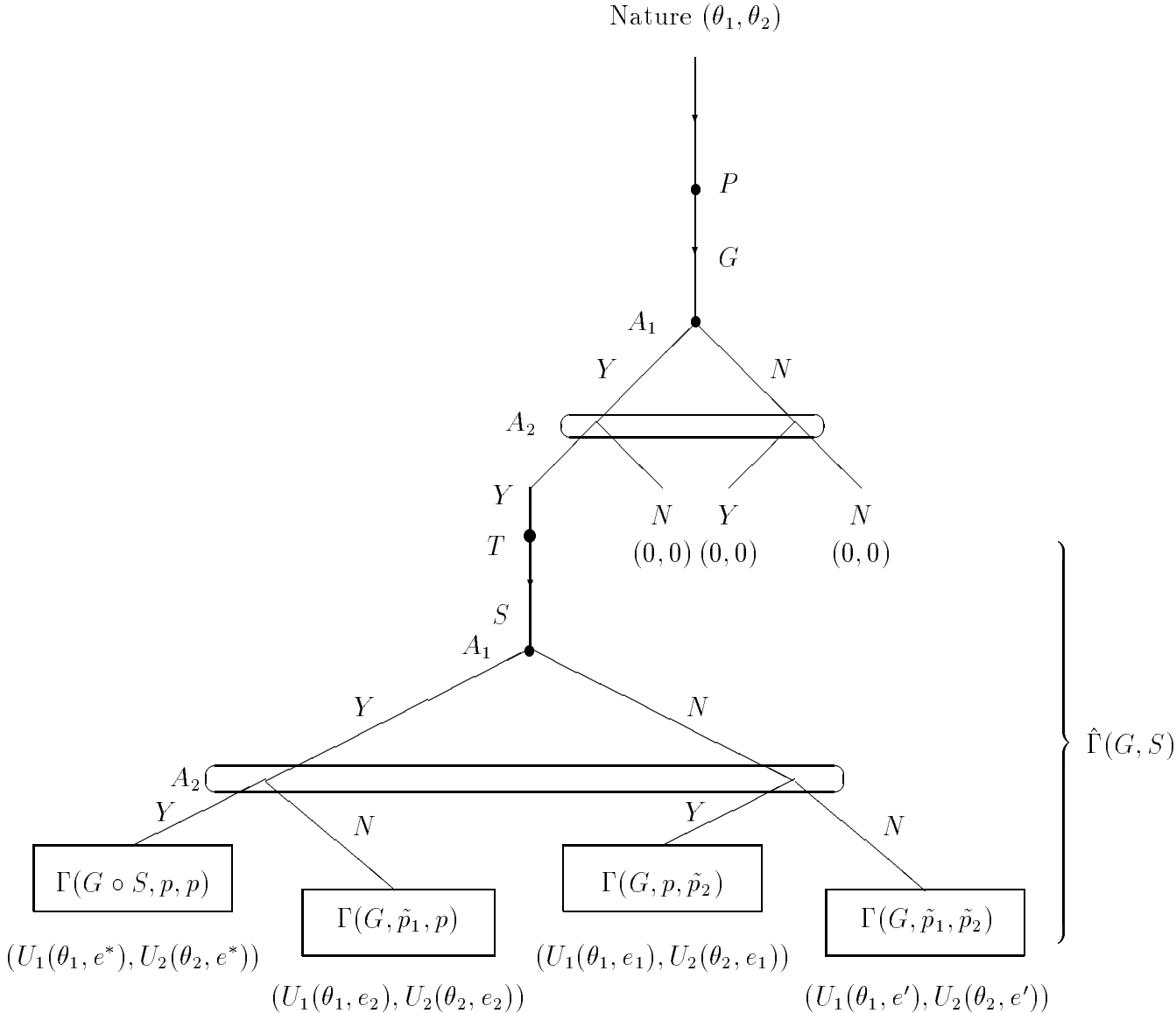
Many lessons of this paper are independent of the specifics of the model, like the principal's objective function or the preferences of agents within the coalition. Therefore, these results would also go through in other environments like auctions, regulation of duopoly, design of incentive schemes within the firm and arbitration mechanisms. More generally, our results suggest that collusion stakes always exist in those correlated environments and that the efficiency of yardstick mechanisms depend significantly on the amount of correlation.

REFERENCES

- Arrow, K.**, 1979, "The Property Rights Doctrine and Demand Revelation under Incomplete Information", in *Economics and Human Welfare*, Academic Press.
- Aspremont, C. d' and L.A. Gérard-Varet**, 1979, "Incentives and Incomplete Information," *Journal of Public Economics*, 11: 25-45.
- Baron, D. and R. Myerson**, 1982, "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50: 911-930.
- Cramton, P. and T. Palfrey**, 1995, "Ratifiable Mechanisms and Learning from Disagreement," *Games and Economic Behavior*, 10: 255-283.
- Crémer, J.**, 1996, "Manipulation of Groves mechanisms," *Games and Economic Behavior*, 13: 39-73.
- Crémer, J. and R. McLean**, 1988, "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56: 1247-1258.
- Epstein, L. and M. Peters**, 1997, "A Revelation Principle for Competing Mechanisms," mimeo, University of Toronto.
- Green, J. and J.J. Laffont**, 1977, "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, 45: 427-438.
- Green, J. and J.J. Laffont**, 1979, "On Coalition Incentive Compatibility," *Review of Economic Studies*, 46: 243-254.
- Grossman, S. and M. Perry**, 1986, "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39: 97-119.
- Groves, T.**, 1973, "Incentives in Teams," *Econometrica*, 41: 587-601.
- Laffont, J.J. and D. Martimort**, 1997, "Collusion under Asymmetric Information," *Econometrica*, 65: 875-911.
- Laffont, J.J. and E. Maskin**, 1979, "A Differentiable Approach to Expected Utility Maximizing Mechanisms," in *Aggregation and Revelation of Preferences*, ed. J.J. Laffont, North-Holland, Amsterdam.
- Laffont, J.J. and E. Maskin**, 1980, "A Differential Approach to Dominant Strategy Mechanisms," *Econometrica*, 48: 1507-1520.
- Laffont, J.J. and E. Maskin**, 1982, "The Theory of Incentives: An Overview," in *Advances in Economic Theory*, ed. W. Hildenbrand, Cambridge University Press.

- Laffont, J.J. and J. Tirole**, 1986, "Using Cost to Regulate Firms," *Journal of Political Economy*, 95: 921-937.
- Ledyard, J. and T. Palfrey**, 1996, "Interim Efficiency in a Public Goods Problem," mimeo, Caltech.
- Ma, A., J. Moore and S. Turnbull**, 1988, "Stopping Agents from Cheating," *Journal of Economic Theory*, 46: 355-372.
- Mailath, G. and A. Postlewaite**, 1990, "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies*, 57: 351-368.
- Maskin, E.**, 1979, "Implementation and Strong-Nash Equilibrium," Chapter 23 in *Aggregation and Revelation of Preferences*, ed. J.J. Laffont, North-Holland.
- Matthews, S. and A. Postlewaite**, 1989, "Pre-Play Communication in Two-Person Sealed-Bid Double Auctions," *Journal of Economic Theory*, 48: 238-263.
- McAfee, P. and P. Reny**, 1991, "Correlated Information and Mechanism Design," *Econometrica*, 60: 395-421.
- Myerson, R.**, 1979, "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 47: 61-73.
- Palfrey, T.**, 1992, "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design" in *Advances in Economic Theory, vol. 1*, ed. J.J. Laffont, 182-282. Cambridge University Press.
- Palfrey, T. and S. Srivastava**, 1989, "Efficient Trading Mechanisms with Preplay Communication," *Journal of Economic Theory*, 55: 17-40.
- Riordan, M. and D. Sappington**, 1988, "Optimal Contracts with Ex Post Information," *Journal of Economic Theory*, 45: 189-199.
- Robert, J.**, 1991, "Continuity in Auction Design," *Journal of Economic Theory*, 55: 169-179.
- Rubinstein, A.**, 1985, "Choices of Conjectures in a Bargaining Game with Incomplete Information," in *Game Theoretic Models of Bargaining*, ed. A. Roth, 99-114, Cambridge University Press.

Figure 1: Game Tree



Appendix

Proof of Proposition 1: When (3) and (4) are binding the principal's expected welfare rewrites as $SW = (1 + \lambda) \left(p_{11}(2\bar{\theta}\bar{x} - c(\bar{x})) + 2p_{12} \left((\bar{\theta} + \underline{\theta})\hat{x} - c(\hat{x}) \right) + p_{22}(2\underline{\theta}\underline{x} - c(\underline{x})) \right)$ if there exist transfers such that the incentive compatibility constraints (1) and (2) hold. Optimizing this expression yields the first-best decision rule. Since we look for transfers such that (1) and (2) are also binding, $(t_{11}, t_{12}, t_{21}, t_{22})$ must solve:

$$p_{11}t_{11} + p_{12}t_{12} = \bar{\theta}(p_{11}\bar{x} + p_{12}\hat{x}), \quad (24)$$

$$p_{11}t_{21} + p_{12}t_{22} = \bar{\theta}(p_{11}\hat{x} + p_{12}\underline{x}), \quad (25)$$

$$p_{12}t_{21} + p_{22}t_{22} = \underline{\theta}(p_{12}\hat{x} + p_{22}\underline{x}), \quad (26)$$

$$p_{12}t_{11} + p_{22}t_{12} = \underline{\theta}(p_{12}\bar{x} + p_{22}\hat{x}). \quad (27)$$

Those equalities are satisfied for some transfers when the matrix

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{pmatrix}$$

is invertible. This holds since its determinant is $\rho = p_{11}p_{22} - p_{12}^2 > 0$. Solving for the taxes: $t_{11} = \frac{1}{\rho} \left(\bar{\theta}(p_{11}\bar{x} + p_{12}\hat{x})p_{22} - \underline{\theta}(p_{12}\bar{x} + p_{22}\hat{x})p_{12} \right)$, $t_{12} = \frac{1}{\rho} \left(\underline{\theta}(p_{12}\bar{x} + p_{22}\hat{x})p_{11} - \bar{\theta}(p_{11}\bar{x} + p_{12}\hat{x})p_{12} \right)$, $t_{21} = \frac{1}{\rho} \left(\bar{\theta}(p_{11}\hat{x} + p_{12}\underline{x})p_{22} - \underline{\theta}(p_{12}\hat{x} + p_{22}\underline{x})p_{12} \right)$, $t_{22} = \frac{1}{\rho} \left(\underline{\theta}(p_{12}\hat{x} + p_{22}\underline{x})p_{11} - \bar{\theta}(p_{11}\hat{x} + p_{12}\underline{x})p_{12} \right)$. These taxes become very large as ρ goes to zero. Indeed, we have $-t_{11} + \bar{\theta}\bar{x} = -\frac{1}{\rho}\Delta\theta p_{12}(p_{12}\bar{x} + p_{22}\hat{x})$, and $-t_{22} + \underline{\theta}\underline{x} = \frac{1}{\rho}\Delta\theta p_{12}(p_{11}\hat{x} + p_{12}\underline{x})$. Hence, t_{11} goes towards plus infinity and t_{22} goes towards minus infinity when ρ goes to zero. Similarly, t_{21} goes towards plus infinity and t_{12} goes towards minus infinity when ρ goes to zero.

Proof of Proposition 2: The proof is standard and thus omitted.

Proof of Proposition 3: Let us consider a perfect Bayesian equilibrium of the overall game of grand-mechanism offer cum coalition formation such that a side-contract is unanimously ratified. This P.B.E is in fact a triplet $\{G^*; S^*; (\tilde{p}_1, \tilde{p}_2)\}$ where:

- G^* from $M = M_1 \times M_2$ into $D = X \times T^2$ maps the messages (m_1, m_2) sent by the agents into an allocation (public good, taxes). G^* maximizes the principal's welfare taking into account the continuation equilibrium of the game of coalition formation.
- S^* is a side-mechanism which, since the revelation principle applies at the last stage of the game, can be taken as being a direct mechanism mapping $\Theta \times \Theta$ into the set of measures on message spaces. Let e^* be its truthful equilibrium. S^* maximizes the sum of the agents' expected utilities subject to individual Bayesian incentive constraints, budget balance and individual rationality constraints $U_i(\theta_i) \geq U_i(\theta_i, e_i)$ for some $e_i \in E(G, \tilde{p}_i, p_{-i})$ and some $\{\tilde{p}_1, \tilde{p}_2\}$.
- $\{\tilde{p}_1, \tilde{p}_2\}$ is the system of out-of equilibrium posterior beliefs used respectively by agent A_2 and A_1 to assess respectively A_1 and A_2 's type following their respective veto of the

side-mechanism S^* . These are the beliefs used in the non-cooperative play of G to compute the status quo payoffs $U_i(\theta_i, e_i)$.

- Consider now the new grand-mechanism $\tilde{G} = G^* \circ S^*$. We prove that there exists a P.B.E. of the overall game of contract offer cum coalition formation in which the principal offers \tilde{G} which is a direct mechanism from $\Theta \times \Theta$ into $D = X \times T^2$, the third-party offers the null side-mechanism $S_0^* = \{\phi = Id, y_k = 0\}$ and this choice is sustained by passive beliefs, $\tilde{p}_1 = \tilde{p}_2 = p$. Because S^* solves (T) with reservation utilities $U_i(\theta_i, e_i)$, the null side-mechanism solves (T) with reservation utilities $U_i(\theta_i)$. Indeed, suppose it is not the case, then there would exist a side-mechanism \tilde{S} such that the third-party can achieve a strictly greater payoff for the coalition than with S^* . Since by definition $U_i(\theta_i) \geq U_i(\theta_i, e_i)$ the third-party's payoff from offering $S^* \circ \tilde{S}$ in the first place would be strictly greater than that achieved with S^* . This would contradict that S^* is optimal when G^* is offered.
- Hence, offering the grand-mechanism \tilde{G} insures to the principal that there is a P.B.E. of the continuation game sustained with passive beliefs in which the null-side-mechanism is unanimously ratified.

Proof of Proposition 4 and Corollary 1: We first solve for the third-party's optimal side-mechanism sustained by a reversion to a non-cooperative equilibrium of the mechanism $G(\cdot)$ played with passive beliefs. Then, we identify conditions such that $S_0^* = \{\phi = Id, y_k = 0\}$, i.e., the null side-mechanism, is the solution to (T) . We conclude by deriving the monotonicity conditions that must be satisfied by such a weakly collusion-proof contract.

- Since, we are not interested in grand-mechanisms such that the $\underline{\theta}$ agent incentive constraint is binding (this constraint will be satisfied ex post), we write the third-party's problem as (the lowerscript denotes the index of the agent concerned with the transfer and let $\phi(\theta_i, \theta_j) = \phi_{ij}$ for simplicity):

$$\max_{\{\phi(\cdot), y(\cdot)\}} \sum_{(i,j) \in \{1,2\}^2} p_{ij} \left(-t_1(\phi_{ij}) - t_2(\phi_{ij}) + (\theta_i + \theta_j)x(\phi_{ij}) \right)$$

subject to

- Budget-balance:

$$\sum_{k=1}^2 y_k(\theta_i, \theta_j) = 0 \quad \forall (\theta_i, \theta_j) \in \Theta^2, \quad (28)$$

- Incentive constraints for respectively the $\bar{\theta}$ agents A_1 and A_2 :

$$\begin{aligned} & p_{11} \left(-t_1(\phi_{11}) - y_1(\bar{\theta}, \bar{\theta}) + \bar{\theta}x(\phi_{11}) \right) + p_{12} \left(-t_1(\phi_{12}) - y_1(\bar{\theta}, \underline{\theta}) + \bar{\theta}x(\phi_{12}) \right) \\ & \geq p_{11} \left(-t_1(\phi_{21}) - y_1(\underline{\theta}, \bar{\theta}) + \bar{\theta}x(\phi_{21}) \right) + p_{12} \left(-t_1(\phi_{22}) - y_1(\underline{\theta}, \underline{\theta}) + \bar{\theta}x(\phi_{22}) \right), \end{aligned} \quad (29)$$

$$\begin{aligned} & p_{11} \left(-t_2(\phi_{11}) - y_2(\bar{\theta}, \bar{\theta}) + \bar{\theta}x(\phi_{11}) \right) + p_{21} \left(-t_2(\phi_{21}) - y_2(\underline{\theta}, \bar{\theta}) + \bar{\theta}x(\phi_{21}) \right) \\ & \geq p_{11} \left(-t_2(\phi_{12}) - y_2(\bar{\theta}, \underline{\theta}) + \bar{\theta}x(\phi_{12}) \right) + p_{21} \left(-t_2(\phi_{22}) - y_2(\underline{\theta}, \underline{\theta}) + \bar{\theta}x(\phi_{22}) \right), \end{aligned} \quad (30)$$

- Participation constraints for respectively the $\bar{\theta}$ agents A_1 and A_2 :

$$p_{11} \left(-t_1(\phi_{11}) - y_1(\bar{\theta}, \bar{\theta}) + \bar{\theta}x(\phi_{11}) \right) + p_{12} \left(-t_1(\phi_{12}) - y_1(\bar{\theta}, \underline{\theta}) + \bar{\theta}x(\phi_{12}) \right) \geq (p_{11} + p_{12})U_1(\bar{\theta}, e_1), \quad (31)$$

$$p_{11} \left(-t_2(\phi_{11}) - y_2(\bar{\theta}, \bar{\theta}) + \bar{\theta}x(\phi_{11}) \right) + p_{21} \left(-t_2(\phi_{21}) - y_2(\underline{\theta}, \bar{\theta}) + \bar{\theta}x(\phi_{21}) \right) \geq (p_{11} + p_{21})U_2(\bar{\theta}, e_2), \quad (32)$$

for some equilibria $e_i \in \Gamma(G, p, p)$ ($i \in \{1, 2\}$).

- Participation constraints for respectively the $\underline{\theta}$ agents A_1 and A_2 :

$$p_{21} \left(-t_1(\phi_{21}) - y_1(\underline{\theta}, \bar{\theta}) + \underline{\theta}x(\phi_{21}) \right) + p_{22} \left(-t_1(\phi_{22}) - y_1(\underline{\theta}, \underline{\theta}) + \underline{\theta}x(\phi_{22}) \right) \geq (p_{21} + p_{22})U_1(\underline{\theta}, e_1), \quad (33)$$

$$p_{12} \left(-t_2(\phi_{12}) - y_2(\bar{\theta}, \underline{\theta}) + \underline{\theta}x(\phi_{12}) \right) + p_{22} \left(-t_2(\phi_{22}) - y_2(\underline{\theta}, \underline{\theta}) + \underline{\theta}x(\phi_{22}) \right) \geq (p_{12} + p_{22})U_2(\underline{\theta}, e_2). \quad (34)$$

Let introduce the following multipliers $\tau(\theta_i, \theta_j)$ for (28), δ_i for (29) and (30). $\bar{\nu}_i$ for (31) and (32) $\underline{\nu}_i$ for (33) and (34). We write the Lagrangean L of the maximization problem above as:

$$L = E(U_1 + U_2) + \sum_{i=1}^2 \delta_i (\text{BIC})_i(\bar{\theta}) + \sum_{i=1}^2 \bar{\nu}_i (\text{BIR})_i(\bar{\theta}) + \sum_{i=1}^2 \underline{\nu}_i (\text{BIR})_i(\underline{\theta}) + \sum_{(\theta_i, \theta_j)} \tau(\theta_i, \theta_j) (\text{BB}) (\theta_i, \theta_j).$$

Optimizing with respect to $y_1(\bar{\theta}, \bar{\theta})$ and $y_2(\bar{\theta}, \bar{\theta})$ yields respectively:

$$\tau(\bar{\theta}, \bar{\theta}) - p_{11}(\delta_1 + \bar{\nu}_1) = 0, \quad (35)$$

$$\tau(\bar{\theta}, \bar{\theta}) - p_{11}(\delta_2 + \bar{\nu}_2) = 0. \quad (36)$$

Optimizing with respect to $y_1(\bar{\theta}, \underline{\theta})$ and $y_2(\bar{\theta}, \underline{\theta})$ yields respectively:

$$\tau(\bar{\theta}, \underline{\theta}) - p_{12}(\delta_1 + \bar{\nu}_1) = 0, \quad (37)$$

$$\tau(\bar{\theta}, \underline{\theta}) + p_{11}\delta_2 - p_{12}\underline{\nu}_2 = 0. \quad (38)$$

Optimizing with respect to $y_1(\underline{\theta}, \bar{\theta})$ and $y_2(\underline{\theta}, \bar{\theta})$ yields respectively:

$$\tau(\underline{\theta}, \bar{\theta}) + p_{11}\delta_1 - p_{21}\underline{\nu}_1 = 0, \quad (39)$$

$$\tau(\underline{\theta}, \bar{\theta}) + p_{21}(\delta_2 + \bar{\nu}_2) = 0. \quad (40)$$

Finally, optimizing with respect to $y_1(\underline{\theta}, \underline{\theta})$ and $y_2(\underline{\theta}, \underline{\theta})$ yields respectively:

$$\tau(\underline{\theta}, \underline{\theta}) + p_{12}\delta_1 - p_{22}\underline{\nu}_1 = 0, \quad (41)$$

$$\tau(\underline{\theta}, \underline{\theta}) + p_{21}\delta_2 - p_{22}\underline{\nu}_2 = 0. \quad (42)$$

- Optimizing with respect to ϕ_{11} yields:

$$\phi_{11}^* \in \arg \max_{\phi_{11}} p_{11} \left(-t_1(\tilde{\phi}_{11}) - t_2(\tilde{\phi}_{11}) + 2\bar{\theta}x(\tilde{\phi}_{11}) \right)$$

$$+(\bar{\nu}_1 + \delta_1)p_{11} \left(-t_1(\tilde{\phi}_{11}) + \bar{\theta}x(\tilde{\phi}_{11}) \right) + (\bar{\nu}_2 + \delta_2)p_{11} \left(-t_2(\tilde{\phi}_{11}) + \bar{\theta}x(\tilde{\phi}_{11}) \right).$$

Taking into account (35) and (36), $\delta_1 + \bar{\nu}_1 = \delta_2 + \bar{\nu}_2$, and simplifying yields:

$$\phi_{11}^* \in \arg \max_{\tilde{\phi}_{11}} \left(-t_1(\tilde{\phi}_{11}) - t_2(\tilde{\phi}_{11}) + 2\bar{\theta}x(\tilde{\phi}_{11}) \right). \quad (43)$$

• Optimizing with respect to ϕ_{12} yields:

$$\begin{aligned} \phi_{12}^* \in \arg \max_{\tilde{\phi}_{12}} & p_{12} \left(-t_1(\tilde{\phi}_{12}) - t_2(\tilde{\phi}_{12}) + (\bar{\theta} + \underline{\theta})x(\tilde{\phi}_{12}) \right) \\ & + p_{12}(\delta_1 + \bar{\nu}_1) \left(-t_1(\tilde{\phi}_{12}) + \bar{\theta}x(\tilde{\phi}_{12}) \right) - \delta_2 p_{11} \left(-t_2(\tilde{\phi}_{12}) + \bar{\theta}x(\tilde{\phi}_{12}) \right) \\ & + \underline{\nu}_2 p_{12} \left(-t_2(\tilde{\phi}_{12}) + \underline{\theta}x(\tilde{\phi}_{12}) \right). \end{aligned} \quad (44)$$

Using (37) and (38), $\delta_1 + \bar{\nu}_1 = -\delta_2 \frac{p_{11}}{p_{12}} + \underline{\nu}_2$. Inserting into (44) yields:

$$\phi_{12}^* \in \arg \max_{\tilde{\phi}_{12}} \left(-t_1(\tilde{\phi}_{12}) - t_2(\tilde{\phi}_{12}) + \left(\bar{\theta} + \underline{\theta} - \frac{p_{11}\epsilon_1}{p_{12}} \Delta\theta \right) x(\tilde{\phi}_{12}) \right) \quad (45)$$

with $\epsilon_1 = \frac{\delta_2}{1+\delta_1+\bar{\nu}_1}$. Similarly

$$\phi_{21}^* \in \arg \max_{\tilde{\phi}_{21}} \left(-t_1(\tilde{\phi}_{21}) - t_2(\tilde{\phi}_{21}) + \left(\bar{\theta} + \underline{\theta} - \frac{p_{11}\epsilon_2}{p_{21}} \Delta\theta \right) x(\tilde{\phi}_{21}) \right) \quad (46)$$

with $\epsilon_2 = \frac{\delta_1}{1+\delta_2+\bar{\nu}_2}$.

• Optimizing with respect to ϕ_{22} yields:

$$\begin{aligned} \phi_{22}^* \in \arg \max_{\tilde{\phi}_{22}} & p_{22} \left(-t_1(\tilde{\phi}_{22}) - t_2(\tilde{\phi}_{22}) + 2\underline{\theta}x(\tilde{\phi}_{22}) \right) + p_{22}\underline{\nu}_1 \left(-t_1(\tilde{\phi}_{22}) + \underline{\theta}x(\tilde{\phi}_{22}) \right) \\ & + p_{22}\underline{\nu}_2 \left(-t_2(\tilde{\phi}_{22}) + \underline{\theta}x(\tilde{\phi}_{22}) \right) - p_{12}\delta_1 \left(-t_1(\tilde{\phi}_{22}) + \bar{\theta}x(\tilde{\phi}_{22}) \right) - p_{21}\delta_2 \left(-t_2(\tilde{\phi}_{22}) + \bar{\theta}x(\tilde{\phi}_{22}) \right). \end{aligned} \quad (47)$$

Note again that $\delta_1 + \bar{\nu}_1 = -\delta_2 \frac{p_{11}}{p_{12}} + \underline{\nu}_2$ and $\delta_2 + \bar{\nu}_2 = -\delta_1 \frac{p_{11}}{p_{12}} + \underline{\nu}_1$. Using (41) and (42), one also gets $p_{22}(1 + \underline{\nu}_1) - p_{12}\delta_1 = p_{22}(1 + \underline{\nu}_2) - p_{12}\delta_2 = p_{22} \left(1 + \delta_1 + \bar{\nu}_1 \right) + \frac{\delta_2 \rho}{p_{12}} = p_{22} \left(1 + \delta_2 + \bar{\nu}_2 \right) + \frac{\delta_1 \rho}{p_{12}} = B$. Simplifying into (47) yields then:

$$\phi_{22}^* \in \arg \max_{\tilde{\phi}_{22}} \left(-t_1(\tilde{\phi}_{22}) - t_2(\tilde{\phi}_{22}) + 2\underline{\theta}x(\tilde{\phi}_{22}) - \frac{(\delta_1 + \delta_2)p_{12}}{B} \Delta\theta x(\tilde{\phi}_{22}) \right).$$

Put differently,

$$\phi_{22}^* \in \arg \max_{\tilde{\phi}_{22}} \left(-t_1(\tilde{\phi}_{22}) - t_2(\tilde{\phi}_{22}) + \left(2\underline{\theta} - \frac{\epsilon_1 p_{12}}{p_{22} + \epsilon_1 \frac{\rho}{p_{12}}} \Delta\theta - \frac{\epsilon_2 p_{12}}{p_{22} + \epsilon_2 \frac{\rho}{p_{12}}} \Delta\theta \right) x(\tilde{\phi}_{22}) \right). \quad (48)$$

In the sequel, we consider symmetric grand-mechanisms such that $\epsilon_1 = \epsilon_2 = \epsilon$ and $t_1(\theta_1, \theta_2) = t_2(\theta_2, \theta_1)$.²⁷ We then have:

$$\phi_{22}^* \in \arg \max_{\tilde{\phi}_{22}} \left(-t_1(\tilde{\phi}_{22}) - t_2(\tilde{\phi}_{22}) + 2 \left(\underline{\theta} - \frac{\epsilon p_{12}}{p_{22} + \epsilon \frac{\rho}{p_{12}}} \Delta\theta \right) x(\tilde{\phi}_{22}) \right). \quad (49)$$

²⁷Asymmetric mechanisms are discussed in the Proof of Proposition 5.

- In a weakly collusion-proof mechanism $\phi_{ij}^* = (\theta_i, \theta_j)$. Inserting into (43), (45), (46) and (49) yields constraints (5), (6) and (7).
- Note that $\epsilon = \frac{\delta}{1+\delta+\bar{p}} \in [0, 1[$. Moreover, $\delta > 0$ when the Bayesian incentive constraints (29) and (30) are binding in (T) .
- Note also that (31), (32), (33) and (34) are binding for a weakly collusion-proof mechanism. Hence, for such a mechanism, the slackness conditions obtained from the Lagrangean's optimization do not give any information on ϵ . Therefore, the principal has some flexibility in choosing this variable.
- We now check for the monotonicity of the schedule of outputs when the mechanism is weakly collusion-proof. From (43) and (44) taken for $\phi_{ij}^* = (\theta_i, \theta_j)$, we have:

$$\begin{aligned}
-2t_{11} + 2\bar{\theta}\bar{x} &\geq -t_{12} - t_{21} + 2\bar{\theta}\hat{x} \\
-t_{12} - t_{21} + \left(\bar{\theta} + \underline{\theta} - \frac{p_{11}}{p_{12}}\epsilon\Delta\theta\right)\hat{x} &\geq -2t_{11} + \left(\bar{\theta} + \underline{\theta} - \frac{p_{11}}{p_{12}}\epsilon\Delta\theta\right)\bar{x}.
\end{aligned}$$

Summing these two inequalities yields $\Delta\theta\left(1 + \frac{p_{11}}{p_{12}}\epsilon\right)(\bar{x} - \hat{x}) \geq 0$, which is satisfied for $\bar{x} \geq \hat{x}$ (since $\epsilon \geq 0$). Proceeding in a similar way and using (44) and (49) for the coalitions $(\bar{\theta}, \underline{\theta})$ and $(\underline{\theta}, \underline{\theta})$, another revealed preference argument tells us that $\left(1 + \frac{2p_{12}^2\epsilon}{p_{12}p_{22} + \epsilon(p_{11}p_{22} - p_{12}^2)} - \frac{p_{11}\epsilon}{p_{12}}\right)(\hat{x} - \underline{x}) \geq 0$. Denoting by $\psi(\epsilon)$ the first term on the left-hand-side of the latter inequality. $\psi(\epsilon)$ is concave in ϵ (since $\psi''(\epsilon) = \frac{-4p_{12}^3p_{22}\rho}{(p_{12}p_{22} + \epsilon\rho)^3} < 0$). Hence, it is either minimum at 0 or 1. $\psi(0) = 1 > 0$ and $\psi(1) > 0$ if and only if $\rho < (p_{22} + p_{12})\frac{p_{12}^2}{p_{11}}$. For a weak correlation, namely $\rho < (p_{22} + p_{12})\frac{p_{12}^2}{p_{11}}$, $\psi(\epsilon)$ is always positive and the monotonicity condition becomes $\hat{x} \geq \underline{x}$. For a strong correlation, namely $\rho > (p_{22} + p_{12})\frac{p_{12}^2}{p_{11}}$, $\psi(\epsilon)$ is negative for ϵ close enough to one and the monotonicity condition becomes then $\underline{x} \geq \hat{x}$.

Proof of Proposition 5:

- We denote by $\alpha, \beta, \gamma, \nu$, the multipliers respectively of (10), (11), (12) and (13). Optimizing with respect to \bar{u} yields:

$$2\lambda p_{11} = \alpha p_{11} + 2\beta. \quad (50)$$

Optimizing with respect to \hat{u}_1 yields:

$$2\lambda p_{12} = \alpha p_{12} - \beta + \gamma. \quad (51)$$

Optimizing with respect to \hat{u}_2 yields:

$$2\lambda p_{12} = -\alpha p_{11} - \beta + \gamma + \nu p_{12}. \quad (52)$$

Optimizing with respect to \underline{u} yields:

$$2\lambda p_{22} = -\alpha p_{12} - 2\gamma + \nu p_{22}. \quad (53)$$

Summing (50) to (53) yields:

$$\nu = \frac{2\lambda}{p_{12} + p_{22}} > 0. \quad (54)$$

Inserting this latter expression into (52) yields:

$$2\lambda p_{12} \left(1 - \frac{1}{p_{12} + p_{22}} \right) = -\alpha p_{11} - \beta + \gamma. \quad (55)$$

Subtracting (51) from (55) yields:

$$\alpha = \frac{2\lambda p_{12}}{(p_{12} + p_{22})(p_{11} + p_{12})} > 0. \quad (56)$$

Inserting this expression into (50) yields then:

$$\beta = \lambda p_{11} \left(1 - \frac{p_{12}}{(p_{12} + p_{22})(p_{12} + p_{11})} \right) > 0. \quad (57)$$

α and β are strictly positive when $\rho = (p_{12} + p_{22})(p_{12} + p_{11}) - p_{12} > 0$. Finally, inserting (57) into (51) gives:

$$\gamma = \lambda(p_{11} + 2p_{12}) \left(1 - \frac{p_{12}}{(p_{12} + p_{22})(p_{12} + p_{11})} \right) > 0. \quad (58)$$

- Since $\gamma > 0$ and since the monotonicity condition implies that $\hat{x} - \underline{x} \geq 0$ for a weak correlation, $\epsilon = 0$ minimizes the cost of constraint (12).
- Optimizing with respect to \bar{x} , \hat{x} and \underline{x} yields $c'(\bar{x}) = 2\bar{\theta}$, $c'(\hat{x}) = \bar{\theta} + \underline{\theta} - \frac{1}{1+\lambda} \frac{1}{2p_{12}} (\alpha p_{11} + \beta) \Delta\theta$, and $c'(\underline{x}) = 2\underline{\theta} - \frac{1}{1+\lambda} \frac{1}{p_{22}} (\alpha p_{12} + \gamma) \Delta\theta$. Using (56), (57) and (58) yields (14) and (15).

Monotonicity of outputs obtains when $1 - \frac{\lambda}{1+\lambda} \frac{p_{11}}{2p_{12}} \left(1 + \frac{p_{12}}{p_{12}+\rho} \right) > -\frac{\lambda}{1+\lambda} \frac{1}{p_{22}} \left(p_{11} + 2p_{12} - \frac{p_{12}p_{11}}{p_{12}+\rho} \right)$. Because $\rho \geq 0$, this property holds when $1 > \frac{\lambda}{1+\lambda} \left(\frac{p_{11}}{p_{12}} - 2\frac{p_{12}}{p_{22}} \right)$. Because $\lambda \geq 0$, the latter inequality also holds when $1 > \frac{p_{11}}{p_{12}} - 2\frac{p_{12}}{p_{22}}$ or when $\rho < p_{12}(p_{12} + p_{22})$. However, for a weak correlation, we have by definition $\rho \leq \frac{p_{12}^2}{p_{11}}(p_{12} + p_{22})$. But this latter left-hand-side is lower than $p_{12}(p_{12} + p_{22})$ when $p_{12} \leq p_{11}$. Hence, monotonicity of outputs is ensured.

- Ex post rents are obtained from solving the system (10) to (13) (with $\epsilon = 0$). After tedious computations, we find:

$$\hat{u}_2 = -\frac{p_{11}p_{22}}{2(p_{12} + \rho)} \Delta\theta(\hat{x}^c - \underline{x}^c) < 0. \quad (59)$$

$$\underline{u} = \frac{p_{11}p_{12}}{2(p_{12} + \rho)} \Delta\theta(\hat{x}^c - \underline{x}^c) > 0. \quad (60)$$

$$\hat{u}_1 - \underline{u} - \Delta\theta \underline{x}^c = \underline{u} - \hat{u}_2 = \frac{p_{11}}{2(p_{11} + p_{12})} \Delta\theta(\hat{x}^c - \underline{x}^c) > 0. \quad (61)$$

$$\bar{u} - \hat{u}_2 - \Delta\theta \hat{x}^c = -\frac{p_{12}}{2(p_{11} + p_{12})} \Delta\theta(\hat{x}^c - \underline{x}^c) < 0. \quad (62)$$

$$\hat{u}_2 - \bar{u} + \Delta\theta\bar{x}^c = \Delta\theta(\bar{x}^c - \hat{x}^c) + \frac{p_{12}}{2(p_{11} + p_{12})}\Delta\theta(\hat{x}^c - \underline{x}^c) > 0. \quad (63)$$

$$\underline{u} - \hat{u}_1 + \Delta\theta\hat{x}^c = \frac{p_{11} + 2p_{12}}{2(p_{11} + p_{12})}\Delta\theta(\hat{x}^c - \underline{x}^c) > 0. \quad (64)$$

• Using (63) and (64), it is immediate to show that a $\underline{\theta}$ agent's incentive constraint is strictly satisfied. A $\bar{\theta}$ agent receives a strictly positive rent in this truthful equilibrium: $U(\bar{\theta}) = \frac{1}{p_{11} + p_{12}}(p_{11}\bar{u} + p_{12}\hat{u}_1) = \Delta\theta\left(\frac{p_{11}\hat{x}^c + p_{12}\underline{x}^c}{p_{11} + p_{12}}\right) - \frac{\rho p_{11}}{p_{12} + \rho}\Delta\theta(\hat{x}^c - \underline{x}^c)$ which is strictly positive since the factor of \underline{x} is strictly positive and the factor of \hat{x} is also strictly positive ($p_{12} + \rho - \rho(p_{11} + p_{12}) > 0$ when $\rho > 0$ and $p_{11} + p_{12} < 1$).

Lastly, monotonicity of the decision rule and the fact that (11) and (12) are binding ensure that the other coalition incentive constraints are satisfied.

Asymmetric Grand-Mechanisms: One may wonder whether the principal could not prevent collusion in a cheaper way by offering a grand-mechanism which treats differently both agents still keeping a symmetric decision rule. In particular, we would have some flexibility in setting $\epsilon_1 \neq \epsilon_2$ and asymmetric transfers in all states of nature. For a weak correlation, one can show that the same monotonicity conditions as under symmetric information within the coalition can be derived, namely $\bar{x} \geq \hat{x} \geq \underline{x}$. Again, the costs of the coalition incentive constraints involving a $(\underline{\theta}, \bar{\theta})$ coalition are minimized when $\epsilon_1 = \epsilon_2 = 0$. There is then a simple argument showing that, in fact, there is always a symmetric mechanism which does at least as well for the principal. First note that the principal's payoff depends only on the sum of the agents' ex post rents. Suppose indeed that the optimal weak collusion-proof grand-mechanism is then asymmetric. Let us denote by G_1^* this mechanism. Because of symmetry, another mechanism G_2^* obtained by permuting all indices is also a solution. The grand-mechanism $\frac{1}{2}G_1^* + \frac{1}{2}G_2^*$ obtained by averaging ex post rents with an equal weight is symmetric. Moreover, it is also individually incentive compatible for both agents and satisfy the same coalition incentive constraints as G_1^* . Hence, it achieves the same welfare as G_1^* and G_2^* . Finally, there is no loss of generality in looking for optimal symmetric mechanisms in the case of a weak correlation.

Proof of Proposition 7: Monotonicity of outputs obtains when $1 - \frac{\lambda}{1+\lambda}\frac{p_{11}}{2p_{12}}\left(1 + \frac{p_{12}}{p_{12} + \rho}\right) > -\frac{\lambda}{1+\lambda}\frac{1}{p_{22}}\left(p_{11} + 2p_{12} - \frac{p_{12}p_{11}}{p_{12} + \rho}\right)$. This condition does not hold when p_{12} is small enough, i.e., for strong correlation. Pooling arises at the optimum. Solving for the solution of the principal's problem under the constraint that $\hat{x} = \underline{x} = x^P$ yields the first order condition: $(2p_{12} + p_{22})c'(x^P) = 2p_{22}\underline{\theta} + 2p_{12}(\underline{\theta} + \bar{\theta}) - \frac{1}{1+\lambda}(\alpha(p_{12} + p_{11}) + \beta + \gamma)\Delta\theta$. Simplifying, we get (16).

Proof of Proposition 8: The optimal contract is obtained for a non-monotonic schedule of outputs, $\bar{x} > \underline{x} > \hat{x}$. Then the principal finds optimal to choose $\epsilon = 1$ to

minimize the cost of (12).²⁸ The binding constraints of (P) are then:

$$\text{(BIC)} \quad p_{11}\bar{u} + p_{12}\hat{u}_1 \geq p_{11}\hat{u}_2 + p_{12}\underline{u} + \Delta\theta(p_{11}\hat{x} + p_{12}\underline{x}) \quad (65)$$

$$\text{(CIC)}_1 \quad 2\bar{u} \geq 2\underline{u} + 2\Delta\theta\underline{x} \quad (66)$$

$$\text{(CIC)}_2 \quad \hat{u}_1 + \hat{u}_2 \geq 2\underline{u} + \Delta\theta\underline{x} + \frac{p_{11}}{p_{12}}\epsilon\Delta\theta(\hat{x} - \underline{x}) \quad (67)$$

where $\epsilon = 1$

$$\text{(IR)} \quad p_{12}\hat{u}_2 + p_{22}\underline{u} \geq 0. \quad (68)$$

- Let denote by respectively α , β , γ and ν the multipliers of these constraints. Proceeding as in **Proof of Proposition 5**, we find $\alpha = \frac{2\lambda p_{12}}{(p_{11}+p_{12})(p_{22}+p_{12})}$, $\beta = \frac{\lambda p_{11}\rho}{p_{12}+\rho}$, $\gamma = \frac{2\lambda p_{12}\rho}{p_{12}+\rho}$, $\nu = \frac{2\lambda}{p_{12}+p_{22}}$.
- Solving for the optimal outputs yields $\bar{x}^c = \bar{x}^*$ and $c'(\hat{x}^c) = \underline{\theta} + \bar{\theta} - \frac{\lambda}{1+\lambda} \frac{p_{11}}{p_{12}+\rho} \left(1 + \frac{1}{p_{12}}\right) \Delta\theta$ when p_{12} is small enough, the positiveness constraint $\hat{x}^c \geq 0$ is binding and $\hat{x}^c = 0$ is in fact optimal. Finally, we have also: $c'(\underline{x}^c) = 2\underline{\theta} - \frac{\lambda}{1+\lambda} \frac{2p_{12}}{p_{22}} \Delta\theta$. Hence, \underline{x}^c converges towards \underline{x}^* when p_{12} becomes arbitrarily small.
- Solving for the values of the ex post rents, $\underline{u} = \hat{u}_2 = 0$, $\hat{u}_1 = \Delta\theta\underline{x}^c \left(1 - \frac{p_{11}}{p_{12}}\right)$ and $\bar{u} = \Delta\theta\underline{x}^c$. It is routine to check that all neglected Bayesian incentive, participation and collusion-proofness constraints are satisfied.

Proof of Proposition 9: To derive the equilibrium correspondence of G^* when posterior beliefs change, first, note that for any belief system, a $\underline{\theta}$ agent (whether a deviant or a non-deviant one) reports truthfully because of dominant strategy for this type. Hence, non-truthful equilibria obtains when a $\bar{\theta}$ agent A_2 (the non-deviant agent) lies and claims he is $\underline{\theta}$. Since a $\bar{\theta}$ agent A_2 is indifferent between claiming $\underline{\theta}$ and $\bar{\theta}$ when G^* is played with passive beliefs ((10) is binding for G^*) and since (16) holds, a $\bar{\theta}$ agent A_2 prefers to claim $\underline{\theta}$ when G^* is played with optimistic beliefs $\tilde{p}_1 > p$. Then, since (17) and (22) hold, A_1 whatever his type reports truthfully. It is thus immediate that $E(G^*, \tilde{p}_1, p) = \{e_2^*\}$ for optimistic beliefs $\tilde{p}_1 > p$. For pessimistic beliefs, a $\bar{\theta}$ agent A_2 always prefers to report instead his true type $\bar{\theta}$. Non-truthful equilibria where A_2 lies cannot hold. However, the deviant agent A_1 still having passive beliefs on A_2 and being indifferent between lying or not (since (10) is binding in G^*) may also lie without changing A_2 's incentives to tell the truth. Thus, we have $E(G^*, \tilde{p}_1, p) = \{e^*, e_1^*\}$ for pessimistic beliefs.

²⁸ ϵ belongs to the open interval $[0, 1[$ but can be made as close as possible to one by increasing the multiplier δ of the incentive constraint in the third party's problem.