Monitoring of Delegated Contracting.*

Michela Cella[†]
(STICERD and LSE)

May 23, 2000
(VERY PRELIMINARY, COMMENTS WELCOME)

Abstract

In this paper we study the delegation of a production process in a three-tier hierarchy. The principal contracts directly only with the agent that produces the final good leaving him in charge of the contract for the production of the intermediate good. We then allow the principal to costlessly monitor the communication between the lower levels of the hierarchy that goes on at the sub-contracting stage. We study two possible scenarios, one in which the principal observes the menu of sub-contracts offered by the first agent and the other in which he can observe the report from the second agent to the first one. In both cases the monitoring damages the first agent and reduces production inefficinecy. We then study how the agent can change the subcontract offer in an attempt to conceal the information that is monitored by the principal.

Keywords: Adverse Selection, Hierarchies, Delegation, Monitoring.
JEL classification:D20, D82, L22, L51

*I am indebted to Kevin Roberts and Antoine Faure-Grimaud for their dedicated supervision. I also wish to thank Michele Arslan, Anna Creti, Luca Deidda, Niko Matouschek, Paolo Ramezzana, Imran Rasul, Cecilia Testa and all the participants to the EOPP internal seminar for helpful comments and suggestions. All remaining errors are mine.

[†]e-mail: m.cella@lse.ac.uk, mailing address: LSE-STICERD, Houghton Street, WC2 2AE London.

1 Introduction.

> In any organisational structure, the nature of communication can be important to the efficiency of the transactions governed by that form. (O. Williamson [1975]).

Delegation of economic activity and subcontracting are widely observed phenomena. The need of exploiting the gains from specialisation is one of the reasons for their diffusion, which has been favoured by the high improvement of communication systems and by the increased sophistication of the available forms of contracts.

Examples of subcontracting can be found in almost any field of economic activity both in the public and private sector. In procurement for example, the purchaser frequently deals only with a number of potential "prime" suppliers, each of whom is granted considerable autonomy over sourcing of components from its subcontractors. Also new forms of private-public partnerships make large use of sub-contracts like the Private Finance Initiative, PFI, launched in the UK in 1992 and designed to bring private financing to public projects. In the private sector, instead, we observe manufacturing firms that only assemble what they once used to produce, having now preferred to contract out the production of various sub-components of the final good (Alcatel is just one example) to more specialised small firms, that in turn may supply more than one firm with their intermediate good. Producers frequently sell their products through a network of distributors who contract with retail agents. Other examples include R&D companies that prefer to delegate some tasks to other centres of research. Finally also the relationship between the board of an investment fund and its managers involves different forms of delegation.

What all these examples have in common is a hierarchical structure, where each level is linked to the lower one by a contract ruling one or more economic activities. Hierarchical decentralisation involves therefore gains from specialisation but brings also extra-costs due to the loss of control over lower levels of these hierarchies. The head of the hierarchy can then try to reduce these costs by regaining some control over the "subordinates", one way to do so is monitoring these delegated relationships to acquire information.

What we do in this work is in fact taking a closer look at the links between members of hierarchies when the informational structure changes because of some monitoring activity by the top level of the hierarchy. We study the contracting over a production process in a very simple delegated environment. More precisely we look at a three-tier hierarchy where a principal wants a final good which is produced by an agent (A_1) using an intermediate good provided by another agent (A_2). Both agents have private information about their marginal costs of production, this sets us in an adverse selection world.

We assume that the principal contracts over the quantity desired of final good directly with A_1 and lets him free to contract with A_2 about the provision of the intermediate good, i.e. the principal cannot contract directly with the second agent. Two contracts will have to be studied, the Grand Contract between the principal and the first agent and the Sub-Contract between the two agents. Given that hierarchies and delegation are widely observed phenomena, we study some features of the agents' strategic behaviour in this specific environment. We therefore impose a delegated structure to our model, without asking what are the reasons that lead to such an organisational mode or comparing it with a centralised one.

Delegation reduces the burden of communication and information processing on the principal but on the other hand it also introduces additional incentives problems. These come from the fact that, even if the principal contracts directly only with one agent, he would like to condition the menu of contracts also on the type of the second agent. Therefore when offering the contract to $A_1$the principal has to give incentives to this agent to truthfully report not only his own type but also the type of the second one, which he will have learned at the sub-contracting stage. In other words, the cost of delegation comes from the fact that the informational rents paid to the agent with whom the principal deals directly can be quite high; this happens because the principal has to reimburse A_1 for the informational rent he has paid to A_2 (after all, these are costs for A_1) and then give him incentives to truthfully report two pieces of information.

It would then seem that monitoring at no cost the communication between the two agents would be profitable for the principal because he would get to know at no cost a piece of information for which he would have to pay otherwise. The direct effect coming from this activity is a gain for the principal, due to a reduction in the transfer he makes to the first agent, but to evaluate thoroughly the effect of monitoring on the profits of the principal we need to study the reaction of the first agent who is the one loosing the most from this activity.

We consider two monitoring possibilities: in the first case the principal observes the report that the second agents makes to the first one; in the second framework the principal observes the menu of subcontracts offered by the first agent to the second one.

When the principal monitors the report from A_2 to A_1 he becomes informed about the type of the second agent, it is the first agent who is made worse off by the monitoring activity even if he still earns a positive rent in some state of the world. the agent could then try to nullify the effect of this monitoring by eliminating the communication with the other agent. A way to do it is offering a "pooling" contract that does not require any report from A_2. We show that indeed the principal design the contract in a way that such deviation it is never profitable, because the pooling contract proves to be too costly.

If instead the principal monitors the sub-contracts offer, he gets to know the type of the first agent (the offer is in fact revealing, we are facing an "informed principal problem). Now the asymmetry of information between the principal and the first agent has disappeared and this allows the principal to leave A_1 with no rent in any state of the world. As an attempt to neutralise the monitoring activity, A_1 can try to delay the revelation of his own type by offering a menu of two pairs of contracts, each pair designed for one of the possible types of A_2 (but that pools across the types of the first agent, leaving for the moment the second agent uninformed). Then, before production takes place, he will reveal his type and it will be clear which contract, inside the chosen menu, is going to be implemented. The principal then obtains no information and he does not gain from monitoring the first agent's offer of sub-contract.

This work is in the stream of literature on collusion and delegation in hierarchies which started with Tirole [1986][1] that gave a clear cut to the way in which organisations and hierarchies were studied in economic theory. They were no longer considered single blocks but networks of overlapping and nested principal-agent relationships where coalition formation and side-contracting are allowed. Since then many articles have been published on this topic, trying to model the additional incentive problems that delegation and collusion can cause even in very simple hierarchies.

More precisely the set-up of our model is taken from Laffont and Martimort [1998] where they compare decentralised and centralised organisation of a production process when there are limits on communication.

An analysis very similar to ours is carried on in Baron and Besanko [1992] but they do not consider both monitoring activities and do not model the possibility of reaction by the agents. Also very related to this, the work by Melumad, Mookherjee and Reichelstein [1995] compares centralised and decentralised structures with similar timing but where the agents do not supply strictly complementary inputs as in our setting.

The structure of the paper is as follows. Section 2 presents the model, utility functions and contracts. Section 3 derives the optimal delegation proof contract in the benchmark case. Section 4 studies the same organisational structure but allows for the monitoring of the report into the sub-contract. Section 5 allows instead for the monitoring of the contract offered. Section 6 concludes. All proofs are in appendix. monitoring of the report into the sub-contract.

2 The Model.

The principal P wants to buy a quantity q of final good. The first agent A_1 produces a quantity $q_1$of the final good using the intermediate good q_2 which is produced by the second agent A_2. Production uses a Leontieff technology such that $q = q_1 = q_2$, in other words the production process is "componetised".[2] As we said before the organisational structure is decentralised, so the principal contracts directly only with the first agent, leaving him the freedom of contracting with the second one.

Each agent A_i $(i = 1, 2)$ faces a constant marginal cost θ_i of producing good i. These marginal costs are independently drawn from the same common knowledge distribution with discrete support $\Theta_i = \Theta = \left\{\underline{\theta}, \overline{\theta}\right\}$, and $\Delta\theta = \overline{\theta} - \underline{\theta} > 0$. With probability ν the agent is efficient, i.e. $\theta_i = \underline{\theta}$. With probability $(1-\nu)$ the agent is inefficient, i.e. $\theta_i = \overline{\theta}$.

Each agent knows only its own cost and not that of the other agent. The principal is uninformed on both agents' costs.

[1]On collusion in hierarchies see also Tirole [1992] and Laffont and Martimort [1997].

[2]Componetised in the sense that the good is formed by putting together components in fixed proportions. The components are produced by different firms or organisational units. As a example we can think of a producer of electricity and a distributor of electricity.

The principal maximizes his revenue minus the monetary transfers to the first agent:

$$W = S(q) - t(q) \tag{2.1}$$

with $S'(\cdot) > 0$, $S''(\cdot) < 0$.

The first agent's utility is given by the monetary transfer received by the principal minus the total costs:

$$U_1 = t - \theta_1 q - y_2 \tag{2.2}$$

where y_2 is the transfer he makes to the second agent at the subcontracting stage.

The second agent's utility is given by:

$$U_2 = y_2 - \theta_2 q \tag{2.3}$$

Had the framework been one of symmetric information, the first best outcome would have been achieved, namely a decreasing schedule of outputs prescribing for each possible pair of agents the quantity that equates marginal benefits and marginal costs of production, that is:

- $S'(q(\underline{\theta}, \underline{\theta})) = 2\underline{\theta}$
- $S'\left(q\left(\underline{\theta}, \overline{\theta}\right)\right) = S'\left(q\left(\overline{\theta}, \underline{\theta}\right)\right) = \underline{\theta} + \overline{\theta}$
- $S'\left(q\left(\overline{\theta}, \overline{\theta}\right)\right) = 2\overline{\theta}$

2.1 The contracts

As we mentioned in the previous section the organisation of the productive activity is decentralised, the principal contracts with A_1 and then the latter contracts with A_2. Therefore we will have to study two contracts, which will be offered by the parties at different stages.

The principal proposes a grand contract, GC, to the first agent that specifies a quantity to be produced and a transfer, i.e. a pair $\left\{q\left(\widehat{\theta}_1, \widehat{\theta}_2\right), t\left(\widehat{\theta}_1, \widehat{\theta}_2\right)\right\}$, where $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are the reported types[3]. This contract takes therefore a form which is standard in the adverse selection literature, the menu offered by the principal to the first agent is composed by four pairs, a quantity and a transfer, one for each possible state of the world. Which pair will actually be implemented depends on the report of A_1 about both agents' types.

At a later stage, A_1, who is the one allowed to communicate with A_2, offers a side contract, SC, to the second agent that consists of a *manipulation-function*[4] of reports and a transfer, i.e. $\left\{\Phi\left(\theta_1, \widetilde{\theta}_2\right), y_2\right\}$, where $\widetilde{\theta}_2$ is the report from A_2 to A_1.

[3] The Revelation Principle applies in this framework so we focuse only on direct mechanims, both for the grand contract and for the sub-contract.

[4] This is a function that to any true pair of types assigns a pair of messages to be delivered to the principal $\Phi : \Theta^2 \rightarrow M_1 \times M_2$ (we do not allow random messages). Then because of the Revelation Principle the relevant range for $\Phi(\theta_1, \theta_2)$ will be Θ^2.

This side-(or sub-) contract is then an agreement between the two agents on how A_1 shall report the information to the principal and how much A_2 receives for each of the possible reports. Therefore while the schedule for q is decided by the contracting between P and A_1, how much will be produced is determined at the subcontracting stage when the two agents fix the manipulation function given their true types. The manipulation function acts therefore as a commitment device for the first agent: if it was not part of the sub-contract then A_1 could have incentive to renege the agreement reached with A_2 over the reports to be made to the principal.[5]

Throughout the paper we assume that sub-contracting is not contractible, that is the contract between the principal and the first agent cannot specify a particular sub-contract between the two agents.

In order to simplify notation, denote $t\left(\overline{\theta},\overline{\theta}\right) = \overline{t}$; $t\left(\underline{\theta},\overline{\theta}\right) = \widehat{t}_1$; $t\left(\overline{\theta},\underline{\theta}\right) = \widehat{t}_2$; $t\left(\underline{\theta},\underline{\theta}\right) = \underline{t}$ and use a similar notation for $q\left(\cdot\right)$.

2.2 The timing.

The timing of the game is the following:

1. Nature draws θ_i each agent learns his cost.
2. P proposes the grand contract M to A_1.
3. A_1 offers SC to A_2.
4. A_2 accepts or refuses the other agent's offer, if he refuses the game ends and both agents get their reservation utility.
5. A_2 reports to A_1.
6. A_1 accepts or refuses M, if he refuses the game ends.
7. A_1 reports to P according to the manipulation function $\Phi\left(\theta_1,\widetilde{\theta}_2\right)$.
8. Output and monetary transfers are implemented. t to A_1 according to M. y_2 to A_2 according to SC.

With this timing then the first agent accepts the grand contract only at the very end and more important after getting to know the type of the second agent. This means that individual rationality constraints will have to be satisfied ex-post for A_1. Another possibility would have been to consider interim individual rationality constraints for A_1, this means that the first agent accepts the grand contract when he still does not know A_2's type. This variation would amount to inserting stage 6 in our timing between stages 2 and 3. The principal is better off when he has to satisfy these constraints only in expected terms because he can "play" with the slackness of the constraints in different states of the world and since the agent is risk neutral this

[5] In that case the analysis would be much more complicate, various additional incentive constraints should be considered. Unless, of course, the report to be made is included in the sub-contract in another way.

does not bring any extra-cost. Actually the second-best can be achieved in this case (as under centralisation) and delegation has no cost. This is because at the time of accepting the contract the principal and A_1 are in the same situation vis-á-vis A_2, none of them knows his type and this is enough to enable the principal to align the interest of the first agent to his own. [6] In our setting instead, at the moment he has to accept the grand contract, A_1 has an advantage over the principal vis-á-vis A_2. He has two pieces of private information and this will cause the even more distortions in output compared to the second-best.

Therefore by choosing this particular timing we set ourselves in a framework where delegation is truly costly.[7]

3 Benchmark.

In this section we study what can be considered the benchmark case for our analysis, a simple delegation model with no monitoring.

3.1 The side contract.

The overall game is a two stages one so it can be solved backwards. When agent A_1, being of type θ_1, contracts with the bottom agent he actually solves the following problem, $SC(\theta_1)$[8]:

$$\begin{aligned} \max E_{\theta_2}[U_1] \;=\; & \nu\left(t\left(\Phi\left(\theta_1,\underline{\theta}\right)\right)-y_2\left(\theta_1,\underline{\theta}\right)-\theta_1 q\left(\Phi\left(\theta_1,\underline{\theta}\right)\right)\right)+ \\ & (1-\nu)\left(t\left(\Phi\left(\theta_1,\overline{\theta}\right)\right)-y_2\left(\theta_1,\overline{\theta}\right)-\theta_1 q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right)\right) \end{aligned} \tag{3.1}$$

Subject to:

$$y_2\left(\theta_1,\overline{\theta}\right)-\overline{\theta} q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right)=0 \tag{3.2}$$

$$y_2\left(\theta_1,\underline{\theta}\right)-\underline{\theta} q\left(\Phi\left(\theta_1,\underline{\theta}\right)\right)=y_2\left(\theta_1,\overline{\theta}\right)-\underline{\theta} q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right) \tag{3.3}$$

Where (3.2) and (3.3) are the participation constraint of an inefficient second agent and the incentive compatibility constraint of an efficient second agent respectively. By rearranging the constraints and substituting we can derive the individually rational and incentive compatibles transfers:

[6] This is a well established result (see for example laffont and Martimort [1998]). Laffont and Martimort [1997] obtain the same result in a centralised framework with collusion under asymmetric information with nonanonymous transfers. There the two agents enter into the side contract offered by an third party without knowing more on each other than the principal does.

[7] Alternatively one could assume risk aversion for the first agent as in Faure-Grimaud, Laffont and Martimort [2000] and Faure-Grimaud-Martimort [2000].

[8] Note that we are in an informed principal framework, we do not face ant additional problem because the utility functions of both agents are linear (Maskin and Tirole [1990]). Therefore A_1 does not loose from making a revealing offer.

$$y_2\left(\theta_1,\overline{\theta}\right)=\overline{\theta}q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right) \tag{3.4}$$

$$y_2\left(\theta_1,\underline{\theta}\right)=\underline{\theta}q\left(\Phi\left(\theta_1,\underline{\theta}\right)\right)+\Delta\theta q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right) \tag{3.5}$$

This is what the second agent receives from the first one and, by assumption, in our framework these transfers cannot be observed by the principal otherwise he would be able to infer the cost parameter of the first agent (we will allow for this possibility in a subsequent section).

3.2 The Grand Contract.

When offering the grand contract the principal faces some constraints caused by the asymmetry of information between him and the agents; he has to give incentives to the agent he is contracting with, namely A_1, to truthfully report all the valuable information. Since we are in a delegated environment, the first agent has not only to report his own type, but also the second agent's one. This means that the incentive compatibility constraints will be different with respect to those that have to be satisfied in a normal one agent-one principal adverse selection model. In other words, we apply the *Delegation-Proofness Principle*, that says that there is no loss of generality in restricting the analysis to the study of grand mechanisms which are unchanged through the process of delegation, i.e. such that the sub-contract is equal to the "null sub-contract" that is the one where it is optimal for the agents not to manipulate the reports and the manipulation function is the identity function[9], $(\Phi\left(\theta_1,\widetilde{\theta}_2\right)=\left(\theta_1,\widetilde{\theta}_2\right))$.

Lemma 1 *A grand contract, GC, is delegation proof iff the following incentive compatibility constraints are satisfied:*

$$t\left(\underline{\theta},\underline{\theta}\right)-2\underline{\theta}q\left(\underline{\theta},\underline{\theta}\right)\geq t\left(\theta_1,\theta_2\right)-2\underline{\theta}q\left(\theta_1,\theta_2\right) \tag{3.6}$$

$$t\left(\overline{\theta},\underline{\theta}\right)-\left(\overline{\theta}+\underline{\theta}\right)q\left(\overline{\theta},\underline{\theta}\right)\geq t\left(\theta_1,\theta_2\right)-\left(\overline{\theta}+\underline{\theta}\right)q\left(\theta_1,\theta_2\right) \tag{3.7}$$

$$t\left(\underline{\theta},\overline{\theta}\right)-\left(\underline{\theta}+\overline{\theta}+\frac{\nu}{1-\nu}\Delta\theta\right)q\left(\underline{\theta},\overline{\theta}\right)\geq t\left(\theta_1,\theta_2\right)-\left(\underline{\theta}+\overline{\theta}+\frac{\nu}{1-\nu}\Delta\theta\right)q\left(\theta_1,\theta_2\right) \tag{3.8}$$

$$t\left(\overline{\theta},\overline{\theta}\right)-\left(2\overline{\theta}+\frac{\nu}{1-\nu}\Delta\theta\right)q\left(\overline{\theta},\overline{\theta}\right)\geq t\left(\theta_1,\theta_2\right)-\left(2\overline{\theta}+\frac{\nu}{1-\nu}\Delta\theta\right)q\left(\theta_1,\theta_2\right) \tag{3.9}$$

$\forall\left(\theta_1,\theta_2\right)\in\Theta\times\Theta$.

Proof. see the appendix. ■

[9] As is becoming common in the works on delegation we loosely borrow from the collusion literature and the concept of collusion proofness, for a definition see Tirole [1992]. In the collusion framework the null side-contract involves also no transfers between the agent, this of course cannot happen in delegation models where transfers are legitimate.

If these constraints are satisfied then truthtelling is optimal and the report into the grand contract by the first agent will be: $\Phi(\theta_1, \theta_2) = (\theta_1, \theta_2)$.

It is worth noting that a coalition of the kind $\left(\overline{\theta}, \underline{\theta}\right)$ is more efficient[10] than a coalition of the kind $\left(\underline{\theta}, \overline{\theta}\right)$ therefore the former has an incentive to mimic the latter. This difference is due to the fact that A_1, when he is facing an inefficient second agent, has even more incentives to distort his report because of the informational rent he has to pay to an efficient one.[11]

In solving the problem the principal can think of facing a single individual who can be of four different types in a decreasing (or increasing) efficiency order.

The problem the principal is facing is therefore:

$$\begin{aligned} \max E_{\theta_1,\theta_2}[W] &= \nu^2 \left(S\left(\underline{q}\right) - \underline{t}\right) + \nu(1-\nu)\left(S(\widehat{q}_1) - \widehat{t}_1\right) + \\ &\quad + \nu(1-\nu)\left(S(\widehat{q}_2) - \widehat{t}_2\right) + (1-\nu)^2 \left(S(\overline{q}) - \overline{t}\right) \end{aligned} \tag{3.10}$$

Subject to constraints (3.6)-(3.9) and the following individual rationality constraints for A_1:

$$\underline{t} - 2\underline{\theta}\underline{q} + -\Delta\theta\widehat{q}_1 \geq 0 \tag{3.11}$$

$$\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1 \geq 0 \tag{3.12}$$

$$\widehat{t}_2 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_2 - \Delta\theta\overline{q} \geq 0 \tag{3.13}$$

$$\overline{t} - 2\overline{\theta}\overline{q} \geq 0 \tag{3.14}$$

These four individual rationality constraints are *ex-post* constraints because the first agent accepts or refuses the grand-contract after he has received the report by the second agent, so he knows exactly what is the state of the world. He has a double informative advantage with respect to the principal.

The binding constraints are the first three downward IC constraints[12] and the individual rationality constraint of a pair of inefficient agents, namely:

$$\underline{t} - 2\underline{\theta}\underline{q} = \widehat{t}_2 - 2\underline{\theta}\widehat{q}_2 \tag{3.6a}$$

$$\widehat{t}_2 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_2 = \widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1 \tag{3.7a}$$

[10] In other words, the virtual type of a coalition $\left(\overline{\theta}, \underline{\theta}\right)$ is lower then the virtual type of a coalition $\left(\underline{\theta}, \overline{\theta}\right)$; where the virtual type is the relevant type for the principal when he chooses production assignments and it is given by the actual type plus the informational rent.

[11] Remember that the informational rent for an efficient second agent is $\Delta\theta q\left(\Phi\left(\theta_1, \overline{\theta}\right)\right)$ so distorting upward the report will reduce the quantity prescribed for a pair $\left(\theta_1, \overline{\theta}\right)$, but also cause a decrease of the informational rent paid in the other two possibile situations $(\theta_1, \underline{\theta})$.

[12] The single crossing property of utilility functions and the monotonicity of the optimal contract ensure that the other constraints are satisfied.

$$\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_1 = \overline{t} - \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\overline{q} \tag{3.8a}$$

$$\overline{t} - 2\overline{\theta}\overline{q} = 0 \tag{3.14}$$

Proposition 1 *The optimal contact implements a decreasing schedule of outputs* $\underline{q} > \widehat{q}_2 > \widehat{q}_1 > \overline{q}$ *where the prescribed quantities are implicitly defined by:*

- $S'\left(\underline{q}\right) = 2\underline{\theta}$
- $S'\left(\widehat{q}_2\right) = \overline{\theta} + \underline{\theta} + \frac{\nu}{1-\nu}\Delta\theta$
- $S'\left(\widehat{q}_1\right) = \overline{\theta} + \underline{\theta} + \frac{\nu(2-\nu)}{(1-\nu)^2}\Delta\theta$
- $S'\left(\overline{q}\right) = 2\overline{\theta} + \frac{\nu(2-\nu)(1-2\nu)}{(1-\nu)^3}\Delta\theta$

Proof. see the appendix. ■

These quantities are more distorted downwards than the second best ones; the amounts of informational rent is more than double and consequently the principal optimally trades off some efficiency. Comparing these quantities to the second best ones reveals that the further distortions are in the quantities prescribed to pairs where an inefficient second agent is present, this is due to the extra incentive that A_1 must be given to truthfully report the pair of types after he has paid the informational rent to A_2. Hence there is a cost for the principal of not being able to communicate directly with one agent.

4 Monitoring the report.

We now suppose that the principal can costlessly and perfectly monitor the communication between the agents (the report that A_2 makes to A_1), then the *collusive* behaviour of A_1 (misreporting two types) is not feasible anymore.

Note that this impossibility is exogenously imposed, it does not stem from the optimising behaviour of the agents and what we now observe is a mismatch between the organizational and informational structures; the first agent still contracts with the second one but, when reporting, to the principal he cannot manipulate the information about the other because the principal knows it already. In other words we have delegation of production and of contracting but not informational delegation.

The sub-contracting between A_1 and A_2 is not directly affected by this monitoring activity, and since the revelation principle is still valid[13] we can restrict attention to the set of direct mechanisms, as we did in the previous section, without the fear of losing in generality. Hence the incentive compatible transfers between the two agents are still given by:

[13] The intuition of why this is so goes as follows: any inference that A_1 can make on the type of A_2 can be made also by the principal, hence allowing also indirect mechanisms does not change the results.

$$y_2\left(\theta_1,\overline{\theta}\right)=\overline{\theta}q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right) \tag{3.4}$$

$$y_2\left(\theta_1,\underline{\theta}\right)=\underline{\theta}q\left(\Phi\left(\theta_1,\underline{\theta}\right)\right)+\Delta\theta q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right)^{14} \tag{3.5}$$

What changes a lot more is the contracting between the principal and the first agent; when offering the grand contract P has to give incentives to A_1 to reveal only one piece of information, his own type, because he already knows the type of the second agent. We assume verifiability of the information acquired with the monitoring, therefore incentive compatibility needs to hold over two separate pairs of contracts, each pair pools across the types of the second agent.

Lemma 2 *When the principal can monitor the report from A_2 to A_1, a grand contract is incentive compatible if the following constraints are satisfied:*

$$\underline{t}-2\underline{\theta}\underline{q}=\widehat{t}_2-2\underline{\theta}\widehat{q}_2 \tag{4.1}$$

$$\widehat{t}_1-\left(\underline{\theta}+\overline{\theta}+\frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_1=\overline{t}-\left(\underline{\theta}+\overline{\theta}+\frac{\nu}{1-\nu}\Delta\theta\right)\overline{q} \tag{4.2}$$

Proof. see the appendix. ■

Therefore the relevant incentive compatibility constraints are the ones of an efficient first agent paired respectively with an efficient and inefficient second agent, that has to be given some rent in order to reveal that he has a low marginal cost.

These incentive compatibility constraints are not in expected terms because when A_1 reports to the principal they both already know A_2's type. The first agent knows it because the side-contracting stage, and the report it entails, precedes the moment he has to report to the principal; the principal in turn is allowed to listen to the truthful report that A_2 makes to A_1.

The principal must also ensure the participation of the first agent into the grand contract, therefore the following individual rationality constraints have to be satisfied:

$$\underline{t}-2\underline{\theta}\underline{q}+-\Delta\theta\widehat{q}_1\geq 0 \tag{4.3}$$

$$\widehat{t}_1-\left(\underline{\theta}+\overline{\theta}\right)\widehat{q}_1\geq 0 \tag{4.4}$$

$$\widehat{t}_2-\left(\underline{\theta}+\overline{\theta}\right)\widehat{q}_2-\Delta\theta\overline{q}\geq 0 \tag{4.5}$$

$$\overline{t}-2\overline{\theta}\overline{q}\geq 0 \tag{4.6}$$

These are the usual participation constraints for A_1, in the benchmark case only one was binding instead now that the principal monitors and get to know the type

[14] Now the side contract includes a trivial version of the manipulation function which is now $\Phi\left(\theta_1,\theta_2\right)=\left(\widehat{\theta}_1,\theta_2\right)$, in other words only the type of the first agent can be distorted via a non truthful report.

of A_2 an inefficient first agent will be left with his reservation utility irrespectively of the type of second agent he is matched with. Namely:

$$\widehat{t}_2 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_2 - \Delta\theta\overline{q} = 0 \tag{4.5a}$$

$$\overline{t} - 2\overline{\theta}\overline{q} = 0 \tag{4.6a}$$

This is because the principal is extracting only one piece of information, he knows the type of A_2 therefore he is not giving any *extra* rent to A_1 to reveal that the second agent is efficient.

We are now ready to characterise the optimal contract.

Proposition 2 *When the principal can costlessly and perfectly monitor the report of the second agent into the sub-contract the optimal grand contract implements a decreasing schedule of output* $\underline{q} > \widehat{q} > \overline{q}$ ($\widehat{q}_1 = \widehat{q}_2 = \widehat{q}$) *where the prescribed quantities are implicitly defined by:*

- $S'\left(\underline{q}\right) = 2\underline{\theta}$
- $S'\left(\widehat{q}\right) = \left(\underline{\theta} + \overline{\theta}\right) + \frac{\nu}{1-\nu}\Delta\theta$
- $S'\left(\overline{q}\right) = 2\overline{\theta} + 2\frac{\nu}{1-\nu}\Delta\theta$

Proof. *see the appendix.* ■

Had the structure of the organisation be a centralised one these would be the quantities produced; this means that if the principal is allowed to monitor the report made into the sub-contract then the second best can be achieved.[15] He cannot do better than the second best even if he gets to know a piece of information because he receives this information when the second agent is reporting to the first one after he has been given the right rents and incentives to do so. These in turn are costs for A_1 that the principal has to reimburse if he wants to ensure the participation of A_1 in the production process. In other words, in the organisation overall two pieces of information have to be extracted, one by A_1 and one by the principal, exactly like in a centralised setting where both pieces are extracted by the principal.

With the monitoring what disappears is the extra-cost of delegation compared to centralisation, but nothing more:even if we do not have informational delegation anymore we still observe “contractual” delegation and this keeps us in a second best world.

[15] Note in fact that $\widehat{q}_1 = \widehat{q}_2$, simmetry is back in the model because the principal can avoid paying the extra-rent so that the two pairs $\left(\underline{\theta}, \overline{\theta}\right)$ and $\left(\overline{\theta}, \underline{\theta}\right)$ can now be treated equally as in a centralised organisation.

4.1 Can the agent react?

We have seen that monitoring is useful to the principal, when this possibility is available, in fact, he can achieve the second best; he can reduce distortions and lower transfers with respect to the benchmark case. This improvement happens at the expenses of A_1 who gets lower utility in two states of the world, the ones in which he is paired with an efficient second agent because this time the principal does not give him any extra rent to reveal the type of A_2 since the principal already knows this information.

We argue that the to be complete the analysis of this environment should consider the possibility of the agent to react to the monitoring activity carried over by the head of the hierarchy, we will then see whether any profitable deviation exists and if this affects the overall equilibrium of the game.

Since the principal gets an advantage when he listens to the communication between the two agents at the sub-contracting stage a possible reaction is to reduce or eliminate the communication that is monitored[16]. In other words A_1 could offer a contract that does not require a report, therefore that is independent from A_2's type. We call this a pooling sub-contract, because does not separate the types of A_2. More precisely the first agent will offer a set of transfers to the second one as if he was always inefficient, namely:

$$y_2(\theta_1, \theta_2) = \overline{\theta} q(\Phi(\theta_1, \theta_2)) \tag{4.7}$$

By paying always the high marginal cost of production he ensures that both types of A_2 are willing to participate in fact their individual rationality constrains are satisfied:

$$\overline{U}_2 = 0$$

$$\underline{U}_2 = \Delta\theta q(\Phi(\theta_1, \theta_2)) > 0$$

Note that the transfer y_2 and the quantities to be produced are apparently still dependent on both types, this is to be more consistent with what we have done so far and have a more homogeneous notation. Theoretically in this case the message space for the first agent when reporting to the principal is larger than before (it is equal to the one in the benchmark case), when, given the monitoring, A_1 was restricted to the message space $\{m_1, \theta_2\}$ (he had to report the true θ_2). Now the message space is in fact $\{m_1, m_2\}$ but the pooling contract implies that $m_2 = \overline{\theta}$ always. This in turn implies that with direct mechanisms this pooling contract reduces once again the manipulation function to function of one variable: $\Phi(\theta_1, \theta_2) = \left(\widehat{\theta}_1, \overline{\theta}\right)$.

If we restrict ourselves to non-random contracts given that agents can be only of two possible types then this pooling contract is the only possible deviation, no

[16]We assume that the action space of A_1 is given by the set of possible sub-contracts he can offer to A_2.

partial pooling (or semi-separating) is possible, and it is also consistent with the idea of concealing some information to the principal[17].

We solve this game backwards so at the time A_1 has to offer the sub-contract now the decision is more complex, first for a given grand contract he must decide whether to offer a pooling or a separating sub-contract. Then given this choice the principal, who will anticipate the agent's behaviour, will make his offer of the grand contract.

More precisely given a grand contract $GC = \left\{\underline{t}, \underline{q}, \widehat{t}_1, \widehat{q}_1, \widehat{t}_2, \widehat{q}_2, \overline{t}, \overline{q}\right\}$, A_1 will choose the type of subcontract that maximises his expected utility[18].

When offering a pooling sub-contract the utility of A_1 is:

$$U_1(\theta_1) = t\left(\Phi\left(\theta_1, \overline{\theta}\right)\right) - \left(\overline{\theta} + \theta_1\right) q\left(\Phi\left(\theta_1, \overline{\theta}\right)\right) \tag{4.8}$$

the expected utility of a separating offer is instead:

$$\begin{aligned} U_1(\theta_1) &= \nu\left(t\left(\Phi\left(\theta_1, \underline{\theta}\right)\right) - \underline{\theta} q\left(\Phi\left(\theta_1, \underline{\theta}\right)\right) - \Delta\theta q\left(\Phi\left(\theta_1, \overline{\theta}\right)\right) - \theta_1 q\left(\Phi\left(\theta_1, \underline{\theta}\right)\right)\right) + \\ & \quad (1-\nu)\left(t\left(\Phi\left(\theta_1, \overline{\theta}\right)\right) - \overline{\theta} q\left(\Phi\left(\theta_1, \overline{\theta}\right)\right) - \theta_1 q\left(\Phi\left(\theta_1, \overline{\theta}\right)\right)\right) \end{aligned} \tag{4.9}$$

What A_1 will do depends on the grand contract, which in turn will depend on the goals that the principal wishes to achieve through the contract. First of all the principal will want truthful revelation and a separating contract offer because this allows him to obtain the second best which is by definition the best he can reach given the asymmetric information setting.

As we saw before the constraints that need to be satisfied to obtain truthtelling by A_1 are:

$$\underline{t} - 2\underline{\theta}\underline{q} = \widehat{t}_2 - 2\underline{\theta}\widehat{q}_2 \tag{4.1}$$

$$\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_1 = \overline{t} - \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\overline{q} \tag{4.2}$$

Moreover it has to be that the separating contract gives a higher expected utility than the pooling one, that is:

$$\nu U_1\left(\underline{\theta}_1, \underline{\theta}_2\right) + (1-\nu) U_1\left(\underline{\theta}_1, \overline{\theta}_2\right) \geq U_P^*\left(\underline{\theta}_1\right) \tag{4.10}$$

where $U_1\left(\underline{\theta}_1, \underline{\theta}_2\right)$ and $U_1\left(\underline{\theta}_1, \overline{\theta}_2\right)$ are the rents earned by an efficient first agent who is paired with and efficient and inefficient second agent respectively when he offers a separating sub-contract and truthfully reports to the principal. While $U_P^*\left(\underline{\theta}_1\right)$ which we define as:

$$U_P^*\left(\underline{\theta}_1\right) = \max_{\Phi} t\left(\Phi\left(\underline{\theta}_1, \overline{\theta}\right)\right) - \left(\overline{\theta} + \underline{\theta}_1\right) q\left(\Phi\left(\underline{\theta}_1, \overline{\theta}\right)\right) \tag{4.11}$$

[17] Also indirect mechanims are not of any use, any information that they would convey could be monitored by the principal.

[18] With expectations taken over the possibile types of A_2 since the offer is made before knowing the type of the second agent.

is the maximum utility that can be achieved by an efficient first agent that offers a pooling sub-contract.

It is without loss of generality that we limit the analysis to the case of an efficient first agent because an inefficient one will receive his reservation utility regardless of the sub-contract offer.

It is worth noting that $U_P^*(\underline{\theta}_1)$ could be achieved by truthtelling but also by any other report, the following Lemma is of some help in this direction.

Lemma 3 *If a Grand Contract is incentive compatible when the sub-contract offer is separating then it is incentive compatible if the offer is pooling and:*

$$U_P^*(\underline{\theta}_1) = \widehat{t}_1 - \left(\underline{\theta}_1 + \overline{\theta}_2\right)\widehat{q}_1$$

Proof. see the appendix. ■

Having calculated the maximum a first agent can get under any of the two possible sub-contract offers we are now ready to check which will be the preferred move.

Proposition 3 *If the grand contract offered* $GC = \left\{\underline{t}, \underline{q}, \widehat{t}_1, \widehat{q}_1, \widehat{t}_2, \widehat{q}_2, \overline{t}, \overline{q}\right\}$ *is incentive compatible and participation constraints are satisfied then the first agent will always offer a separating sub-contract.*

Proof. see the appendix. ■

Therefore the first agent will not react to the monitoring activity by the principal and the second best can be achieved. This means that the costs of a pooling contract are much higher than the benefits in terms of more flexibility when reporting to the principal. Paying the second agent as if he was always inefficient in order not to get any report that could be monitored is not convenient.

5 Monitoring the sub-contract.

In the previous section we supposed that the principal could monitor the response by the second agent to the contract offer, more precisely he could listen to a report or observe the choice of a particular sub-contract but always without observing the contract itself. We now study what will happen in the hierarchy when the principal can observe the part of communication at the sub-contracting stage that comes from the first agent, he will therefore observe, perfectly and at no cost, the menu of contracts offered by A_1 to A_2 .

As we noted earlier each possible type of A_1 offers a menu of contracts that depends on his own type (in other words the offer is revealing), therefore the principal will be able to know the type of the first agent and he will obtain this information at no cost.

This monitoring does not directly affect the side-contracting stage so, if there is no reaction from the agents, the optimal transfers that induce participation and truthful revelation by A_2 are still given by:

$$y_2\left(\theta_1, \overline{\theta}\right) = \overline{\theta} q\left(\Phi\left(\theta_1, \overline{\theta}\right)\right) \tag{3.4}$$

$$y_2(\theta_1, \underline{\theta}) = \underline{\theta} q(\Phi(\theta_1, \underline{\theta})) + \Delta\theta q\left(\Phi\left(\theta_1, \overline{\theta}\right)\right)^{19} \tag{3.5}$$

As we just said, these are observed by the principal that gets to know θ_1 with certainty, so there is no more asymmetric information between P and A_1 concerning the type of the latter, what is left is asymmetric information (at the report stage) between P and A_1 vis-a-vis A_2 . The principal has to give incentives to A_1 only to report what he has learned from the other agent.

The information that the principal gets through the monitoring is verifiable, therefore the first agent cannot misrepresent his type anymore and incentive compatibility needs to hold over two sets of two contracts each, instead of over the four original contracts (now only one piece of information needs to be reported).

Lemma 4 *When the principal can monitor the sub-contract offer a grand contract is incentive compatible if the following constraints are satisfied:*

$$\underline{t} - 2\underline{\theta}\underline{q} = \widehat{t}_1 - 2\underline{\theta}\widehat{q}_1 \tag{5.1}$$

$$\widehat{t}_2 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_2 = \overline{t} - \left(\underline{\theta} + \overline{\theta}\right)\overline{q}^{20} \tag{5.2}$$

Proof. see the appendix. ■

Therefore the relevant (and binding) incentive compatibility constraints are only those of a coalition made by two efficient agents and an inefficient first agent with an efficient second agent; the individual rationality constraints for the first agent will be binding regardless of his type, in particular when the second agent is inefficient:

$$\overline{t} - \overline{\theta}\overline{q} = 0 \tag{3.14}$$

$$\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1 = 0 \tag{3.12}$$

We can think about this situation as if the principal could see through A_1; we are now able to characterise the contract.

Proposition 4 *When the principal can costlessly and perfectly monitor the side-contract offered to the second agent the optimal contract implements a decreasing schedule of output $\underline{q} > \widehat{q}_2 > \widehat{q}_1 > \overline{q}$, where the prescribed quantities are implicitly defined by:*

- $S'\left(\underline{q}\right) = 2\underline{\theta}$
- $S'(\widehat{q}_2) = \left(\underline{\theta} + \overline{\theta}\right)$

[19] Here again the transfers depend on a very simplified version of the manipulation function, which is now $\Phi(\theta_1, \theta_2) = \left(\theta_1, \widehat{\theta}_2\right)$, this time it's only the secondagent's typoe that can be distorted.

[20] These constraints are not in expected terms because the first agent know the type of the second at the time of reporting to the principal, that is in fact the only valuable piece of information he possesses.

- $S'(\widehat{q}_1) = \left(\underline{\theta} + \overline{\theta}\right) + \frac{\nu}{1-\nu}\Delta\theta$
- $S'(\overline{q}) = 2\overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta$

Proof. see the appendix. ■

It is interesting to note that at this stage the first two quantities are efficient, we observe the "no distortion at the top" condition each time the second agent is efficient, also the schedule of output is monotonic with respect to θ_2. This is because when the principal monitors the contracts offer the asymmetry of information with respect to A_1 disappears, in fact the principal can improve upon the second best, which by definition is the upper bound when you have to pieces of private information.

5.1 The agent's reaction.

The first agent loses a lot from the monitoring activity of the principal, a reaction on his part should go in the direction of concealing in some way that part of information that he transmits via the contract offer.

In order not to reveal any important information to the principal that observes the menu of contracts he can offer to A_2 two menus of two contracts each. He asks for truthful revelation from A_2, but postpones the disclosure of his private information till the very end (at least after he has reported to the principal). The second agent will know later the type of A_1: before production takes place, he will find out what amount of the two prescribed by the chosen side-contract he has to produce. In fact the quantity that the principal demands is dependent on both agents' types.

More precisely the side-contracts take the following form:

$$SC(\theta_1, \underline{\theta}) = \{y_2(\theta_1, \underline{\theta}), \Phi(\theta_1, \underline{\theta}); \quad \theta_1 \in \Theta\}$$

$$SC\left(\theta_1, \overline{\theta}\right) = \left\{y_2\left(\theta_1, \overline{\theta}\right), \Phi\left(\theta_1, \overline{\theta}\right); \quad \theta_1 \in \Theta\right\}$$

These two contracts are designed for an efficient and an inefficient second agent (respectively) but are conditioned on the type of the first agent as well, any type of the second agent will choose the contract designed for himself and wait a later stage to find out exactly what pair of the possible two will be implemented.

Moreover, A_2 is aware of the fact that both transfer and quantity will be later conditioned on the type of A_1 also, but at the acceptance and report stage he does not possess this information (and derives nothing from the contract offered) therefore all his constraints, participation and incentive compatibility, will take an *interim* form.

In particular the relevant ones are:

$$\nu\left(y_2\left(\underline{\theta}, \overline{\theta}\right) - \overline{\theta}q\left(\Phi\left(\underline{\theta}, \overline{\theta}\right)\right)\right) + (1-\nu)\left(y_2\left(\overline{\theta}, \overline{\theta}\right) - \overline{\theta}q\left(\Phi\left(\overline{\theta}, \overline{\theta}\right)\right)\right) = 0 \tag{5.3}$$

$$\begin{aligned} & \nu\left(y_2(\underline{\theta}, \underline{\theta}) - \underline{\theta}q(\Phi(\underline{\theta}, \underline{\theta}))\right) + (1-\nu)\left(y_2\left(\overline{\theta}, \underline{\theta}\right) - \underline{\theta}q\left(\Phi\left(\overline{\theta}, \underline{\theta}\right)\right)\right) = \\ = \; & \nu\left(y_2\left(\underline{\theta}, \overline{\theta}\right) - \underline{\theta}q\left(\Phi\left(\underline{\theta}, \overline{\theta}\right)\right)\right) + (1-\nu)\left(y_2\left(\overline{\theta}, \overline{\theta}\right) - \underline{\theta}q\left(\Phi\left(\overline{\theta}, \overline{\theta}\right)\right)\right) \end{aligned} \tag{5.4}$$

where (4.5) is the individual rationality constraint of an inefficient second agent and (4.6) is the incentive compatibility constraint of an efficient second agent.

The standard techniques employed to solve for the optimal side-contract would require, at this stage, to solve for the transfers first but from the equations above it is evident that our system is underidentified (we have four unknowns and only two equations), so there is some leeway in determining the optimal transfers between the agents. One possibility is to break up the interim constraints and impose the ex-post ones, so that we get an equal number of equations and unknowns.

A more formal justification comes from Mookherjee and Reichelstein [1992] that tells us how in this environment (linear utilities and constant marginal costs) we can equivalently implement this Bayesian allocation in dominant strategies without incurring in any loss for the "principal" (in this case the first agent). But requiring dominant strategy implementation amounts to breaking up the constraints as we have just explained above.

Another justification for proceeding in this way comes from a subtle application of the theorems found in Maskin and Tirole [1990]: they prove that, in a informed-principal relationship with private values and quasilinear utilities, the principal is indifferent between revealing or not his information to the agent. This finding applies to our setting but some more considerations need to be done. If is true that the first agent (the "informed-principal" in our case) does not gain nor loose vis-à-vis A_2 by revealing his type, then the same holds for the latter. Therefore A_1 guarantees to A_2 the satisfaction of the ex-post constraints even if he need not to; this because he is not trying to improve upon the contracting with A_2, he wants to regain some power at the grand-contract stage and he does it by postponing the revelation of his type. Then, if the contract is offered as we proposed above and if the transfers are computed using dominant strategy, the second agent participates and truthfully reveals his information[21] but the principal does not observe anything (and cannot force A_1on his reservation utility).

For all these reasons we can then use the dominant strategy constraints, and note that they are the same ones calculated in the previous section where no monitoring was going on. If transfers in the side contract are the same, then also incentive compatibility and participation constraints for A_1, when contracting with P, are identical to the previous ones. This means that the optimisation problem that the principal solves does not change and the optimal contract is the one derived in Lemma 2. This enables us to state the following.

Proposition 5 *The principal does not gain in monitoring the menu of sub-contracts that the first agent offers to the second one.*

Proof. follows from the arguments above. ■

What we have learned from the above analysis is that, in a delegation model, the principal does not gain if he monitors, even at no cost, the menu of subcontracts

[21] This happens because the contract is identical in the transfers and quantities to the one that would have been offered if the principal was not monitoring. The only difference is that the revelation of the private information of the first agent is slightly delayed.

offered by agents in lower levels of the hierarchy because of the reaction of the agent that looses the most from this activity. In fact, the latter manages to offer a subcontract that allows him to reveal his private information only before production takes place, and all this goes on without affecting the welfare of the agent at the bottom of the hierarchy.

6 Concluding remarks.

We have seen that not every type of monitoring is equally beneficial to the head of a hierarchy because of the possible counteractions by the lower levels. In fact the one that gives a higher utility as a direct effect, the monitoring of contracts, can be nullified by an action of the agent that looses the most. Therefore it seems that the best thing to do, at least in our setting, is to monitor the reports because it is too costly for the agents to avoid it.

This analysis was carried out in the spirit of considering hierarchies as networks of agents that interact through contracts. We therefore wanted to take a closer look at the strategic interactions of members of hierarchical organisations and see whether strategic behaviour could at least weaken some "results" that hold in standard one principal-one agent models.

Despite our attempt to generalise the analysis our setup was quite specific, it is probably worth extending the analysis to a different production function (some degree of substitution allowed) and to a setting where there is correlation between the types of the agents (the monitoring could even have some stronger effect).

References

[1] BARON, D.P. and D. BESANKO (1992) "Information, Control, and Organizational Structure", *Journal of Economics & Management Strategy*, 1:237-275.

[2] FAURE-GRIMAUD, A., J.J. LAFFONT and D. MARTIMORT (2000) "Collusion, Delegation and Supervision with Soft Information", mimeo.

[3] FAURE-GRIMAUD and D. MARTIMORT (2000) "On some Agency Costs of Intermediated Contracting", mimeo.

[4] LAFFONT, J.J. and D. MARTIMORT (1995) "Collusion & Delegation", *Rand Journal of Economics*, 29:280-305.

[5] LAFFONT, J.J. and D. MARTIMORT (1997) "Collusion Under Asymmetric Information", *Econometrica*, 65:875-911.

[6] MASKIN, E. and J. TIROLE (1990) "The Principal-Agent relationship with an Informed Principal: the Case of Private Values", *Econometrica*, 58:379-409.

[7] MELUMAD, D.M., D.MOOKHERJEE and S. REICHELSTEIN (1995) "Hierarchical Decentralization of Incentives Contract", *Rand Journal of Economics*, 26:654-672.

[8] MOOKHERJEE, D. and S. REICHELSTEIN (1992) "Dominant Strategy Implementation of Bayesian Incentive Compatible Rules", *Journal of Economic Theory*, 56:378-399.

[9] TIROLE, J. (1986) "Hierarchies & Bureaucracies: On the Role of collusion in Organizations", *Journal of Law, Economics & Organizations*, 2:181-214.

[10] TIROLE, J. (1992) "Collusion and the Theory of Organization", in J.J. Laffont (ed.), *Advances in Economic Theory*, Cambridge University Press, Cambridge.

Appendix.

Proof of Lemma 1. If we substitute the optimal side transfers (3.4) and (3.5) into the first agent's expected utility function (3.1) we get:

$$\begin{aligned}\max E_{\theta_2}[U_1] &= \nu\left(t(\Phi(\theta_1,\underline{\theta})) - \underline{\theta}q(\Phi(\theta_1,\underline{\theta})) - \Delta\theta q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right) - \theta_1 q(\Phi(\theta_1,\underline{\theta}))\right) + \\ &\quad (1-\nu)\left(t\left(\Phi\left(\theta_1,\overline{\theta}\right)\right) - \overline{\theta}q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right) - \theta_1 q\left(\Phi\left(\theta_1,\overline{\theta}\right)\right)\right)\end{aligned}$$

now we can check for incentive compatibility for any of the possible "coalition" (there are four of them). ■

Then the condition for the optimality of $\Phi(\underline{\theta},\underline{\theta}) = (\underline{\theta},\underline{\theta})$ is the following:

$$\nu\left(t(\underline{\theta},\underline{\theta}) - 2\underline{\theta}q(\underline{\theta},\underline{\theta}) - \Delta\theta q\left(\Phi\left(\underline{\theta},\overline{\theta}\right)\right)\right) + (1-\nu)\left(t\left(\Phi\left(\underline{\theta},\overline{\theta}\right)\right) - \left(\overline{\theta}+\underline{\theta}\right)q\left(\Phi\left(\underline{\theta},\overline{\theta}\right)\right)\right) \geq$$
$$\nu\left(t(\theta_1,\theta_2) - 2\underline{\theta}q(\theta_1,\theta_2) - \Delta\theta q\left(\Phi\left(\underline{\theta},\overline{\theta}\right)\right)\right) + (1-\nu)\left(t\left(\Phi\left(\underline{\theta},\overline{\theta}\right)\right) - \left(\overline{\theta}+\underline{\theta}\right)q\left(\Phi\left(\underline{\theta},\overline{\theta}\right)\right)\right).$$

For $\Phi\left(\underline{\theta},\overline{\theta}\right) = \left(\underline{\theta},\overline{\theta}\right)$ is:

$$\nu\left(t(\Phi(\underline{\theta},\underline{\theta})) - 2\underline{\theta}q(\Phi(\underline{\theta},\underline{\theta})) - \Delta\theta q\left(\underline{\theta},\overline{\theta}\right)\right) + (1-\nu)\left(t\left(\underline{\theta},\overline{\theta}\right) - \left(\underline{\theta}+\overline{\theta}\right)q\left(\underline{\theta},\overline{\theta}\right)\right) \geq$$
$$\nu\left(t(\Phi(\underline{\theta},\underline{\theta})) - 2\underline{\theta}q(\Phi(\underline{\theta},\underline{\theta})) - \Delta\theta q(\theta_1,\theta_2)\right) + (1-\nu)\left(t(\theta_1,\theta_2) - \left(\underline{\theta}+\overline{\theta}\right)q(\theta_1,\theta_2)\right).$$

For $\Phi\left(\overline{\theta},\underline{\theta}\right) = \left(\overline{\theta},\underline{\theta}\right)$ is:

$$\nu\left(t\left(\overline{\theta},\underline{\theta}\right) - \left(\overline{\theta}+\underline{\theta}\right)q\left(\overline{\theta},\underline{\theta}\right) - \Delta\theta q\left(\Phi\left(\overline{\theta},\overline{\theta}\right)\right)\right) + (1-\nu)\left(t\left(\Phi\left(\overline{\theta},\overline{\theta}\right)\right) - 2\overline{\theta}q\left(\Phi\left(\overline{\theta},\overline{\theta}\right)\right)\right) \geq$$
$$\nu\left(t(\theta_1,\theta_2) - \left(\overline{\theta}+\underline{\theta}\right)q(\theta_1,\theta_2) - \Delta\theta q\left(\Phi\left(\overline{\theta},\overline{\theta}\right)\right)\right) + (1-\nu)\left(t\left(\Phi\left(\overline{\theta},\overline{\theta}\right)\right) - 2\overline{\theta}q\left(\Phi\left(\overline{\theta},\overline{\theta}\right)\right)\right).$$

For $\Phi\left(\overline{\theta},\overline{\theta}\right) = \left(\overline{\theta},\overline{\theta}\right)$ is:

$$\nu\left(t\left(\Phi\left(\overline{\theta},\underline{\theta}\right)\right) - \left(\overline{\theta}+\underline{\theta}\right)q\left(\Phi\left(\overline{\theta},\underline{\theta}\right)\right) - \Delta\theta q\left(\overline{\theta},\overline{\theta}\right)\right) + (1-\nu)\left(t\left(\overline{\theta},\overline{\theta}\right) - 2\overline{\theta}q\left(\overline{\theta},\overline{\theta}\right)\right) \geq$$
$$\nu\left(t\left(\Phi\left(\overline{\theta},\underline{\theta}\right)\right) - \left(\overline{\theta}+\underline{\theta}\right)q\left(\Phi\left(\overline{\theta},\underline{\theta}\right)\right) - \Delta\theta q(\theta_1,\theta_2)\right) + (1-\nu)\left(t(\theta_1,\theta_2) - 2\overline{\theta}q(\theta_1,\theta_2)\right).$$

Proof. Simplifying we obtain constraints 3.6-3.9 . This are the conditions for truthtelling, what any coalition gets in the contract leaves it better off than anything else they could have gotten misreporting their types.

Note that we have used first agent's expected utility because the manipulation function is part of the side-contract that is offered before getting to know the second agent's type, then incentive compatibility constraints are ex-post because by the time A_1 reports to the principal he knows the true type of A_2. ■

Proof of Proposition 1. Given the binding constraints 3.6a, 3.7a, 3.8a, 3.14 we can manipulate them and obtain the incentive compatible and individually rational transfers:

- $\underline{t} = 2\underline{\theta}\underline{q} + \Delta\theta\widehat{q}_2 + \frac{\nu}{1-\nu}\Delta\theta\widehat{q}_1 + \frac{1-2\nu}{1-\nu}\Delta\theta\overline{q}$
- $\widehat{t}_2 = \left(\underline{\theta}+\overline{\theta}\right)\widehat{q}_2 + \frac{\nu}{1-\nu}\Delta\theta\widehat{q}_1 + \frac{1-2\nu}{1-\nu}\Delta\theta\overline{q}$
- $\widehat{t}_1 = \left(\underline{\theta}+\overline{\theta}+\frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_1 + \frac{1-2\nu}{1-\nu}\Delta\theta\overline{q}$

- $\overline{t} = 2\overline{\theta}\overline{q}$

We can plot them in the principal's objective function and then maximize with respect to $\underline{q}$, $\widehat{q}_1$, $\widehat{q}_2$ and $\overline{q}$, we then obtain the decreasing schedule of output of Proposition 1. ■

Proof of Lemma 2. as in the case with no monitoring we want the grand contract to be delegation proof, i.e. $\Phi(\theta_1, \theta_2) = (\theta_1, \theta_2)$ but because of the monitoring the agent cannot misreport anymore the type of the second agent and the manipulation function boils down to a trivial version of the previous one $\Phi(\theta_1, \theta_2) = \left(\widehat{\theta}_1, \theta_2\right)$. Given this and the fact that each agent can be only of two types, for each possible coalition it is uniquely defined the coalition they could mimic (for example $\left(\overline{\theta}, \underline{\theta}\right)$ can pretend to be only $(\underline{\theta}, \underline{\theta})$). Therefore applying the same metodology of the proof of Lemma 1, the "coalition" incentive constraints are:

$\underline{t} - 2\underline{\theta}\underline{q} \geq \widehat{t}_2 - 2\underline{\theta}\widehat{q}_2$

$\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_1 \geq \overline{t} - \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\overline{q}$

$\widehat{t}_2 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_2 \geq \underline{t} - \left(\underline{\theta} + \overline{\theta}\right)\underline{q}$

$\overline{t} - \left(2\overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\overline{q} \geq \widehat{t}_1 - \left(2\overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_1$

It is straightforward to show that the only relevant constraints are the downward ones (namely the first two) that are binding at the optimum, the other two will be satisfied if the prescribed schedule of output is monotonic. ■

Proof of Proposition 2. Considering the binding constraints (4.1), (4.2), (4.5a) and (4.6a) allows us to determine the incentive compatible and individually rational transfers, namely:

- $\underline{t} = 2\underline{\theta}\underline{q} + \Delta\theta\widehat{q}_2 + \Delta\theta\overline{q}$
- $\widehat{t}_2 = \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_2 + \Delta\theta\overline{q}$
- $\widehat{t}_1 = \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_1 + \frac{1-2\nu}{1-\nu}\Delta\theta\overline{q}$
- $\overline{t} = 2\overline{\theta}\overline{q}$

We can plot them in the principal's objective function and then maximize with respect to $\underline{q}$, $\widehat{q}_1$, $\widehat{q}_2$ and $\overline{q}$, we then obtain the decreasing schedule of output of Proposition 2. ■

Proof of Lemma 3. Since,when offering a pooling sub-contract, A_1 will always report $\theta_2 = \overline{\theta}$ the relevant incentive constraint is the one of a pair $\left(\underline{\theta}, \overline{\theta}\right)$ that is:

$\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1 \geq \overline{t} - \left(\underline{\theta} + \overline{\theta}\right)\overline{q}$

which is for sure satisfied whenever the incentive compatibility constraint for the same pair under separating sub-contract offer is satisfied. In fact (3.8a) can be rewritten as:

$\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1 \geq \overline{t} - \left(\underline{\theta} + \overline{\theta}\right)\overline{q} + \frac{\nu}{1-\nu}\Delta\theta\left(\widehat{q}_1 - \overline{q}\right).$

Hence when the principal satisfies incentives constraints under separation, the utility of an agent that offers a pooling sub-contract is maximised by truthful report and $U_P^*(\underline{\theta}_1) = \widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1$. ■

Proof of Proposition 3. We need to show that, for a given contract $GC = \left\{\underline{t}, \underline{q}, \widehat{t}_1, \widehat{q}_1, \widehat{t}_2, \widehat{q}_2, \overline{t}, \overline{q}\right\}$ which satisfies incentive compatibility and participation constraints, $\nu U_1(\underline{\theta}_1, \underline{\theta}_2) + (1-\nu) U_1\left(\underline{\theta}_1, \overline{\theta}_2\right) \geq U_P^*(\underline{\theta}_1)$ is always satisfied. Because of the binding incentive constraints under separating:

$U_1(\underline{\theta}_1, \underline{\theta}_2) = \underline{t} - 2\underline{\theta}\underline{q} - \Delta\theta\widehat{q}_1$ and $U_1\left(\underline{\theta}_1, \overline{\theta}_2\right) = \widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1$, if we substitute these expressions into (4.10) it becomes:

$$\nu\left(\underline{t} - 2\underline{\theta}\underline{q} - \Delta\theta\widehat{q}_1\right) + (1-\nu)\left(\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1\right) \geq \widehat{t}_1 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1$$

which is always satisfied when $U_1(\underline{\theta}_1, \underline{\theta}_2) \geq U_1\left(\underline{\theta}_1, \overline{\theta}_2\right)$ which is always true in this setting. ■

Proof of Lemma 4. The proof here parallels that of Lemma 2, with just minor changes. Now A_1 cannot misreport his own type therefore the new version of the manipulation function is: $\Phi(\theta_1, \theta_2) = \left(\theta_1, \widehat{\theta}_2\right)$. The incentive constraints are uniquely determined as well:

$\underline{t} - 2\underline{\theta}\underline{q} \geq \widehat{t}_1 - 2\underline{\theta}\widehat{q}_1$

$\widehat{t}_2 - \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_2 \geq \overline{t} - \left(\underline{\theta} + \overline{\theta}\right)\overline{q}$

$\widehat{t}_1 - \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_1 \geq \underline{t} - \left(\underline{\theta} + \overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\underline{q}$

$\overline{t} - \left(2\overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\overline{q} \geq \widehat{t}_2 - \left(2\overline{\theta} + \frac{\nu}{1-\nu}\Delta\theta\right)\widehat{q}_2$

where the relevant and binding ones are the first two, while the other "upward" constraints will be automatically satisfied if the schedule of output is monotonic. ■

Proof of Proposition 4. Taking the binding constraints (5.1), (5.2), (3.14) and (3.12) (but remembering that participation constraints are binding in any state of the world) we determine the transfers:

- $\underline{t} = 2\underline{\theta}\underline{q} + \Delta\theta\widehat{q}_1$
- $\widehat{t}_2 = \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_2 + \Delta\theta\overline{q}$
- $\widehat{t}_1 = \left(\underline{\theta} + \overline{\theta}\right)\widehat{q}_1$
- $\overline{t} = 2\overline{\theta}\overline{q}$

We can plot them in the principal's objective function and then maximize with respect to $\underline{q}$, $\widehat{q}_2$, $\widehat{q}_1$ and $\overline{q}$, we then obtain the decreasing schedule of output of Proposition 4. ■