Delegated Monitoring versus Arm's Length Contracting[1]

by

Vianney Dequiedt[2] and David Martimort[3]

This version: February 2004

Abstract

This paper analyzes the optimal organization of input suppliers in a procurement context. We discuss when a top firm chooses to induce information sharing and active coordination between its suppliers (consolidation) rather than having them contract at arm's length (delegation). In designing the organization, the firm anticipates that different supplier networks have different costs of inducing information revelation from suppliers. When the cost of information sharing varies, the optimal organization exhibits various discrete degrees of consolidation. There is excessive arm's length contracting compared to the case where the top firm controls this organizational choice itself. The paper offers also a methodology to study in a unified framework various organizational forms analyzed separately in the literature.

Keywords: Organizational Design, Hierarchical Agency, Supplier Networks, Consolidation, Delegation.

[1]We thank Jacques Crémer, Antoine Faure-Grimaud, Dominique Henriet, Jean-Jacques Laffont and seminar participants at ESSET Gersenzee and in Oxford, Toulouse and Venice for helpful comments. We also thank the Editor Bernard Caillaud and two referees for remarks which improved the presentation of the paper.

[2]INRA, GAEL, France.

[3]Université de Toulouse 1, IDEI-GREMAQ and Institut Universitaire de France.

1 Introduction

In the Japanese *Keiretsu*, production relies heavily on sub-contracting. Perrow (1992) reports that Japanese firms deal directly with between 100 to 300 contractors who themselves deal in turn with up to 5000 sub-contractors. Clearly, the large number of partners involved at each layer of contracting makes any direct control of the relationship between contractors by the top firms almost impossible. Nevertheless, top firms devote much attention to building up their network of suppliers. Toyota is an emblematic exemple in that respect. As illustrated by Hines and Rich (1998), Toyota was particularly successful in outsourcing a major part of its competitive advantage to its network of direct and indirect suppliers. This performance is partially explained by a specific organization of suppliers named *kyoryoku kai* or supplier associations. These associations have their roots in the late 1930s with the grouping of twenty of the major Toyota suppliers. Over time these associations have developed to involve almost all Toyota suppliers in the first, second and third tiers of supply. While encouraging such associations, Toyota is not necessarily involved in their day to day functioning. Instead, leadership is delegated to important direct suppliers. The *kyoryoku kai* promotes integration activities among members such as top management group meetings, quality awards and audits, and tries to achieve thereby a better coordination through information exchanges. Such flows of information were not present at that scale in other forms of supplier-buyer relationships and should be viewed as the key factor explaining Toyota achievement. As argued by Baiman and Rajan (2002), the amount of information exchanged among subcontractors is what really distinguishes supplier networks from more traditional arm's length relationships.

Beyond the Toyota example, the design of efficient supplier networks has been a major issue across the whole automotive industry. For instance, it is generally acknowledged that one definitive advantage of Ford over General Motors was obtained by reorganizing the supply chain. With this reorganization, Ford kept only relationships with contractors in charge of assembling parts built by other sub-contractors, and left to those contractors the choice of the degree of control and coordination they would like with their sub-contractors. In recent years, this model has been spread over other industries: the computer and electronics industries or the e-commerce sector have now become intensive users of supplier networks.[1] Despite its practical importance, determining under which conditions such networks emerge remains a theoretical question by large still unsettled. Answering this question is of considerable importance in view of the lively debate among both management scholars and practitioners over the organization of the whole supply chain.

This paper tackles this issue. We present a theory of the optimal organization of

[1] See Baiman and Rajan (2002).

supplier networks based on the private incentives of the suppliers bringing complementary resources to the organization to either consolidate their private information on their technological capabilities or to remain independent units contracting at arm's length. In organizing the network of its suppliers, a firm faces a trade-off between providing incentives for costly information sharing among those suppliers and benefiting from the efficiency gains that such close-knit relationships generate. To exhaust those gains, contracts with suppliers must induce the right amount of information sharing even though sub-contracting is delegated to intermediate contractors. This paper analyzes how the choice of the optimal degree of information sharing interacts with information flows to determine the performances of various contracting modes.

To model supplier networks, we envision the top firm as facing suppliers who are privately informed about their own production costs. This top firm is unable to recommend directly the degree of consolidation between these suppliers. Instead, it must induce the optimal degree of information sharing from its own viewpoint. Credible information sharing requires that one supplier, the contractor, monitors the sub-contractor. Instead of investing in an efficient and close relationship with a partner, the contractor could possibly save on these monitoring costs and contract at arm's length with an independent sub-contractor. Of course, this organizational choice (de facto delegated to the contractor) may differ from what would be chosen by the top firm itself. Inducing information sharing has thus the flavor of a moral hazard problem. Taking into account this dimension of the contractor's strategy is in line with some empirical studies. Yun (1999), in a study of the Korean automotive industry emphasized that the negotiated contracts between the contractor and the subcontractors are based on estimates of the subcontractors' costs which are obtained by the contractor through audit or accounting data analysis.

Since the organizational choice affects the degree of informational asymmetries that the top firm faces in front of its suppliers, it also affects information flows in the organization and the trade-off between allocative efficiency and rent extraction that arises in adverse selection environments. In designing the contours of the organization, the top firm faces a complex incentive problem which thus mixes moral hazard and adverse selection.

This paper generates two sets of results. On the one hand, we offer a unified framework which derives various organizational forms which have been studied in isolation within the received literature as being solutions to the same design problem for different values of the cost of information sharing. On the other hand, we offer a measure of the control loss incurred by the principal in delegating organizational choice to contractors and highlight a systematic bias towards too much arm's length contracting and not enough consolidation.

To understand the first set of results, it is rather intuitive to see that, as the fixed-cost of monitoring increases, efficiency gains may not be enough to always justify information sharing (consolidation) and the principal may prefer that suppliers remain at arm's length (delegation). Under delegation, information on sub-contractors is indirectly learned by the contractor through an incentive compatible contract. Because of this maintained asymmetry of information between agents, delegation entails more allocative inefficiency than consolidation but it also saves on the fixed-cost of monitoring.

Because of monitoring fixed-cost and non-convexities, a marginal shift in exogenous technological or preferences parameters may create significant discontinuities in the optimal incentive schemes and outputs chosen within the organization. One cannot expect a whole continuum of organizational forms to emerge as structural parameters of the model are slightly perturbed. Instead only a discrete number of organizations distinguished by their production and their degree of consolidation emerge in equilibrium.

As the fixed-cost of monitoring increases, the degree of consolidation of the optimal organization diminishes. Mixed organizations emerge when consolidation is chosen only by efficient contractors whereas arm's length contracting is preferred by inefficient ones. The top firm's profit nevertheless varies continuously when one moves from a fully consolidated organization, to a partially consolidated one, to finally reach values of the fixed-cost where arm's length contracting is always preferred. Our model predicts thus that organizations where arm's length contracting prevails yield also lower profits than consolidated organizations.

The existence of the partially consolidated organization illustrates the fact that the preferred organizational form can be influenced by the efficiency of the contractor. In our model, different choices of monitoring intensities among a set of otherwise identical firms can be explained by a difference in the corresponding contractors' efficiency. The channel through which productive efficiency of the contractor has an impact on its decision to monitor is called the *efficiency effect* and can be explained as follows. To solve optimally the efficiency-rent extraction trade-off of the overall organization, the top firm has to link the total production and the efficiency of the contractor. But due to the complementarity of the inputs, increasing total production means increasing production of the sub-contractor which in turn means a higher rent for the sub-contractor in case it is not monitored. As a consequence, a more efficient contractor has more incentives to monitor.

To understand the control loss incurred by the principal when he delegates the organizational choice to the contractor, it is important to see that, the principal could increase his screening ability when monitoring the sub-contractor by himself. When monitoring is delegated, some of its gains are dissipated under the form of extra information rents to the contractor coming from a better coordination in manipulating information to the principal. As a result, consolidation is more attractive when the principal does monitor by himself. When the decision to monitor is delegated to the contractor, there may be too much arm's length contracting.

Taking care of the endogenity of information structures, we give a fresh look to the standard trade-off between allocative efficiency and rent extraction in multi-agent organizations. Earlier works on multi-agent organizations include Baron and Besanko (1992), Gilbert and Riordan (1995), Melumad, Mookherjee and Reichelstein (1995), Mookherjee and Reichelstein (2001) among others. Those authors have shown that delegation entails no loss of generality from the principal's viewpoint. Indeed, the principal achieves the same expected payoff as under centralized contracting, i.e., if he could directly contract with both agents and forbid any communication between them.[2] However, as shown in Baron and Besanko (1992) and Gilbert and Riordan (1995), consolidation performs better than delegation. Inducing information sharing between the agents benefits the principal. To understand why consolidation does not necessarily emerge, one must endogenize the information structure and model explicitly the incentives to create such a consolidated organization. To the best of our knowledge, only Baron and Besanko (1999) have tackled this issue so far. They study the incentives to consolidate when agents can opt out to contract independently with the principal. In their model, the principal takes into account this outside option and offers to the agents enough information rent to let consolidation form so that he benefits from its greater allocative efficiency. Binding outside options lead nevertheless to new output distortions and diminish the benefits of consolidation. A first difference with their paper is that we model explicitly the agent's incentives to monitor his peer. A second difference comes from the fact that we restore the Stackelberg leadership of the principal who wants to offer the best conditions for the emergence of an efficient network of suppliers.

With respect to all the existing literature on delegation and consolidation, our paper is the first one to encompass within the same model these two organizational forms and to derive them as solutions to the same optimization problem for different values of the monitoring cost.

There is a distinct literature on monitoring in three-layer hierarchies which, following Tirole (1986), assumes that the monitor does not produce and learns exogenously a parameter which is privately known by the productive agent. These papers study settings where supervisory information is either hard as in Laffont and Tirole (1991) and Kofman and Lawarrée (1993) or soft as in Faure-Grimaud, Laffont and Martimort (2003) and analyze collusion between the monitor and the agent. We consider soft information but, by contrast, the information learned by the contractor is endogenously determined by

[2] Laffont and Martimort (1997 and 1998) have also shown that, in an environment with independent productive shocks, lateral communication and collusion between the agents would be costless for the principal which could fight this collusion by a clever design of the agents' incentive schemes.

the incentive structure.[3] Moreover, we are particularly interested in situations in which monitoring is delegated to a productive agent. This gives us a natural model to study the supply chain and identify interesting interactions between the supervisory and productive roles of the contractor.

Section 2 presents the model and provides useful benchmarks: costless consolidation and delegation. In Section 3, we derive the agency cost faced by the principal when he delegates the choice on whether monitoring or contracting arm's length to a contractor. We then discuss the principal's optimal organizational choice. Section 4 provides some extensions of our basic framework and discusses the benefits of a fully centralized organization. Proofs are relegated to an Appendix.

2 Model and Benchmarks

• **Agents and Organizations:** We consider an organization involving one top firm or principal P and two suppliers or agents A_1 and A_2. Agents produce essential inputs q_i $(i = 1, 2)$ for the organization. These inputs are perfect complements so that $q = q_1 = q_2$ denotes the final output of the organization. Each agent has private information on his constant marginal cost θ_i. Both costs are independently drawn from the same common knowledge distribution on $\Theta = \{\underline{\theta}, \bar{\theta}\}$ (with $\Delta\theta = \bar{\theta} - \underline{\theta} > 0$) with respective probabilities ν and $1-\nu$. The utility function of A_i is $U_i = z_i - \theta_i q$ where z_i is the net monetary transfer received by A_i. Principal P contracts directly with the contractor A_1 and delegates to him the task of contracting with the sub-contractor A_2. Since A_1 receives a transfer t from the principal P and gives back a transfer y to A_2, we have $z_1 = t - y$ and $z_2 = y$. Finally, each agent has a reservation payoff normalized to zero.

The principal's profit is $V = S(q) - t$ where $S(\cdot)$ is increasing and concave ($S' > 0$, $S'' < 0$) and satisfies the Inada conditions ($S'(0) = +\infty$, $S'(+\infty) = 0$) with $S(0) = 0$. Under complete information, the first-best outputs are such that marginal benefit equals marginal cost: $S'(q^*(\theta_1, \theta_2)) = \theta_1 + \theta_2$.

On top of his productive role, the contractor A_1 can also monitor the sub-contractor A_2. A_1 learns perfectly A_2's efficiency parameter if he incurs a fixed-cost c. We assume that c does not depend on A_1's efficiency.[4] A_1's information on A_2 is soft, i.e., it is not possible to transmit it credibly to the principal. The benefit of using A_1 as a monitor comes from his comparative advantage in monitoring A_2.

The network of suppliers has close links when those monitoring costs are incurred. As

[3] Strausz (1997) considers endogenous but hard information.

[4] This assumption suppresses one simple possible impact of A_1's efficiency on its decision to monitor and allows us to identify clearly the *efficiency effect* stressed below.

in the literature on network formations,[5] this gives the contractor access to the profits available to the downstream supplier and drives its incentives to form such a link.

Without loss of generality, P's grand-contract offered to A_1 is a truthful direct mechanism $GC = \{t(\hat{\Phi}); q(\hat{\Phi})\}$ where $t(\cdot)$ is A_1's transfer received by P, $q(\cdot)$ is the output produced and $\hat{\Phi}$ is the report made by A_1 on all information that he has learned by sub-contracting with A_2, i.e., on the whole vector of cost parameters (θ_1, θ_2). For convenience, we denote outputs by $\underline{q} = q(\underline{\theta}, \underline{\theta}), \hat{q}_1 = q(\underline{\theta}, \bar{\theta}), \hat{q}_2 = q(\bar{\theta}, \underline{\theta})$ and $\bar{q} = q(\bar{\theta}, \bar{\theta})$. To express the incentive and participation constraints in a simple way, it is useful to define the complete information payoff of the (A_1, A_2) coalition in the various states of nature as $\underline{u} = t(\underline{\theta}, \underline{\theta}) - 2\underline{\theta}\underline{q}$, $\hat{u}_1 = t(\underline{\theta}, \bar{\theta}) - (\underline{\theta} + \bar{\theta})\hat{q}_1$, $\hat{u}_2 = t(\bar{\theta}, \underline{\theta}) - (\underline{\theta} + \bar{\theta})\hat{q}_2, \bar{u} = t(\bar{\theta}, \bar{\theta}) - 2\bar{\theta}\bar{q}$. Instead of reasoning over transfers and outputs, we will repeatedly use rent-output pairs to illustrate the trade-offs between allocative efficiency and rent extraction in the different organizational forms.

• **Timing:** It is described on Figure 1 below.

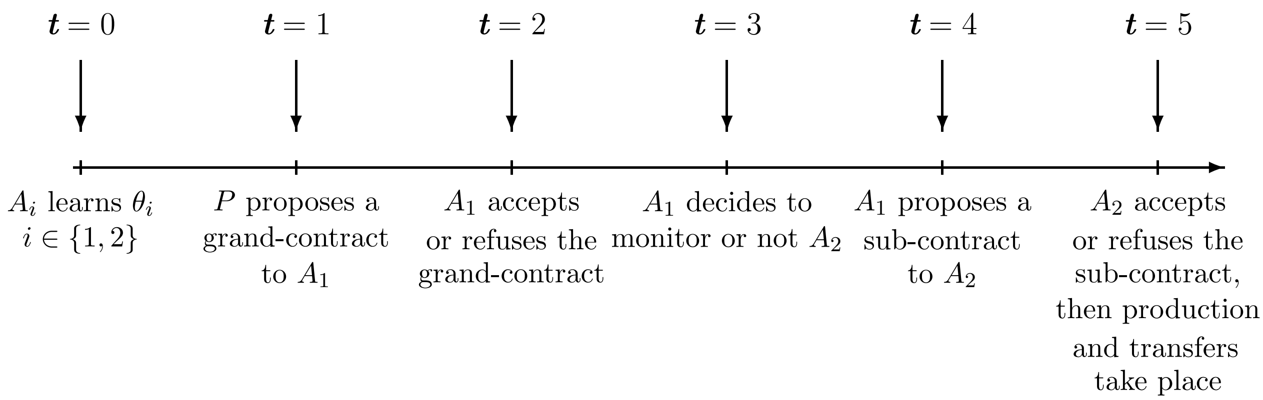

Figure 1: Timing.

This sequence of events corresponds to the case where the contractor accepts the project before meeting with input suppliers. This timing is particularly relevant for a long-lasting project which commits suppliers for a long period of time. The suppliers arrange then their organization as a best response to the procurement contract proposed by the top firm.

• **Costless Consolidation and Delegation:** Two polar situations have been stressed by the existing literature: costless consolidation and delegation. [6] We can conduct their

[5]See Jackson and Wolinski (1996), Bala and Goyal (2000) and Kranton and Minehart (2001) among others. Contrary to this literature which neglects the issue of information, we stress the information role of building close relationships.

[6]Most of the existing literature relied on a framework with a continuum of types which turns out to be quite difficult to use when one wants to analyze the incentives to monitor or not. We recast below the results of the literature in a discrete type environment.

analysis in parallel, using the notion of virtual cost to take into account the informational rents. Under costless consolidation, A_1 learns for free A_2's type at date $\boldsymbol{t} = 3$, while under delegation, A_1 stays uninformed. The sub-contracting game is thus different in the two cases. In the costless consolidation case, A_1 does not have to let a rent to A_2 and just reimburses A_2 for its true cost ($\underline{\theta}$ or $\bar{\theta}$). Under delegation, A_1 must design an incentive compatible sub-contract and A_2 keeps an informational rent. In that case A_1 has to reimburse A_2 for its virtual costs ($\underline{\theta}$ or $\bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta$) which are bigger than true costs. Anticipating this, A_1 behaves in the grand-contract as if its true costs were $(2\underline{\theta}, \underline{\theta}+\bar{\theta}, \bar{\theta}+\underline{\theta}, 2\bar{\theta})$ under costless consolidation, or $(2\underline{\theta}, \underline{\theta}+\bar{\theta}+\frac{\nu}{1-\nu}\Delta\theta, \bar{\theta}+\underline{\theta}, 2\bar{\theta}+\frac{\nu}{1-\nu}\Delta\theta)$ under delegation. Apart from this difference, the two maximization programs faced by P at date $\boldsymbol{t} = 1$ are similar. As usual in the screening literature, (coalition) incentive constraints are binding upwards. The relevant ones prevent a coalition from pretending being less efficient than what it is. The binding participation constraint is for the less efficient coalition. Finally, mixed coalitions are not screened apart: $(\hat{u}_1, \hat{q}_1) = (\hat{u}_2, \hat{q}_2) = (\hat{u}, \hat{q})$.

Under costless consolidation, the optimal grand-contract is thus solution to:

$$(\mathcal{P}^C): \max_{\{(\underline{u},\underline{q});(\hat{u},\hat{q});(\bar{u},\bar{q})\}} \nu^2(S(\underline{q})-2\underline{\theta}\underline{q}-\underline{u})+2\nu(1-\nu)(S(\hat{q})-(\underline{\theta}+\bar{\theta})\hat{q}-\hat{u})+(1-\nu)^2(S(\bar{q})-2\bar{\theta}\bar{q}-\bar{u})$$

subject to

$$\underline{u} \geq \hat{u} + \Delta\theta\hat{q}, \tag{1}$$

$$\hat{u} \geq \bar{u} + \Delta\theta\bar{q}. \tag{2}$$

$$\bar{U}_1 = \nu\hat{u} + (1-\nu)\bar{u} \geq 0. \tag{3}$$

The costless consolidation analyzed above slightly differs from the model used in the literature in terms of the timing for its formation. As one can see by comparing the participation constraints (3) and (5), this difference allows us to keep the delegation and consolidation organizations as close as possible.

Under delegation, the principal's problem becomes:

$$(\mathcal{P}^D): \max_{\{(\underline{u},\underline{q});(\hat{u},\hat{q});(\bar{u},\bar{q})\}} \nu^2(S(\underline{q})-2\underline{\theta}\underline{q}-\underline{u})+2\nu(1-\nu)(S(\hat{q})-(\underline{\theta}+\bar{\theta})\hat{q}-\hat{u})+(1-\nu)^2(S(\bar{q})-2\bar{\theta}\bar{q}-\bar{u})$$

subject to

$$\underline{u} \geq \hat{u} + \Delta\theta\hat{q},$$

$$\hat{u} \geq \bar{u} + \Delta\theta\bar{q} + \frac{\nu}{1-\nu}\Delta\theta(\hat{q}-\bar{q}). \tag{4}$$

$$\bar{U}_1 = \nu(\hat{u} - \Delta\theta\bar{q}) + (1-\nu)\bar{u} \geq 0. \tag{5}$$

The solution to these programs is straightforward and summarized in the next proposition.

Proposition 1 • *The optimal grand-contract under* **costless consolidation** *entails:*

Only the efficient agent A_1 gets a positive interim information rent

$$\underline{U}_1^C = \Delta\theta(\nu\hat{q}^C + (1-\nu)\bar{q}^C) > 0 \text{ and } \bar{U}_1^C = 0; \tag{6}$$

There is no output distortion with respect to the first-best for a $(\theta_1 = \underline{\theta}, \theta_2 = \underline{\theta})$-coalition, $\underline{q}^C = q^*(2\underline{\theta})$ and a downward distortion below the first-best otherwise;

$$S'(\hat{q}^C) = \underline{\theta} + \bar{\theta} + \frac{\nu}{2(1-\nu)}\Delta\theta, \quad \text{and} \quad S'(\bar{q}^C) = 2\bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta. \tag{7}$$

• *The optimal grand-contract under* **delegation** *entails:*

Only the efficient agents A_1 and A_2 get a positive information rent which is the same:

$$\underline{U}_i^D = \Delta\theta(\nu\hat{q}^D + (1-\nu)\bar{q}^D), \text{ and } \bar{U}_i^D = 0 \text{ for } i = 1, 2; \tag{8}$$

The optimal schedule of outputs yields the same first-best level of production as under consolidation for a $(\theta_1 = \underline{\theta}, \theta_2 = \underline{\theta})$-coalition $\underline{q}^D = \underline{q}^C = q^*(2\underline{\theta})$ and downward distortions compared to consolidation for the other quantities $\hat{q}^D < \hat{q}^C$ and $\bar{q}^D < \bar{q}^C$ with:

$$S'(\hat{q}^D) = \underline{\theta} + \bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta \quad \text{and} \quad S'(\bar{q}^D) = 2\bar{\theta} + \frac{2\nu}{1-\nu}\Delta\theta. \tag{9}$$

• *If the principal could directly contract and communicate with both suppliers, he would optimally implement the same quantities and rents as under delegation.*

It is straightforward to observe that delegation yields a lower profit to the principal than what he can get through a costless consolidation. Indeed, denoting by $V^C(\underline{q}, \hat{q}, \bar{q})$ and $V^D(\underline{q}, \hat{q}, \bar{q})$ the corresponding expected profits for any given schedule of outputs $(\underline{q}, \hat{q}, \bar{q})$,

$$V^C(\underline{q}, \hat{q}, \bar{q}) = V^*(\underline{q}, \hat{q}, \bar{q}) - \nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q}),$$

$$V^D(\underline{q}, \hat{q}, \bar{q}) = V^*(\underline{q}, \hat{q}, \bar{q}) - 2\nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q}),$$

where $V^*(\underline{q}, \hat{q}, \bar{q}) = \nu^2(S(\underline{q}) - 2\underline{\theta}\underline{q}) + 2\nu(1-\nu)(S(\hat{q}) - (\underline{\theta}+\bar{\theta})\hat{q}) + (1-\nu)^2(S(\bar{q}) - 2\bar{\theta}\bar{q})$ is the complete information profit achieved by the principal when implementing any schedule of outputs $(\underline{q}, \hat{q}, \bar{q})$. For further references, we denote by $V^C = V^C(\underline{q}^C, \hat{q}^C, \bar{q}^C)$ and $V^D = V^D(\underline{q}^D, \hat{q}^D, \bar{q}^D)$ the optimal profits achieved under costless consolidation and delegation. The second terms in the expression of $V^C(\underline{q}, \hat{q}, \bar{q})$ and $V^D(\underline{q}, \hat{q}, \bar{q})$ represent the agency costs incurred in both cases. Agency costs under delegation are twice those under consolidation. Intuitively, the fact that both agents retain now private information duplicates information rents and hardens the trade-off between efficiency and rent extraction.[7]

The consolidated organization can be viewed as having a high degree of integration.[8] Complete information between suppliers facilitates the coordination of their production decisions and creates scope for efficiency gains along the supply chain. Coming back to the introductory example of Toyota, some of the *kyoryoku kai* are actively organized by direct suppliers such as Denso or Aisin and regroup their own sub-contractors. Consolidation is then achieved (at least partially) through audits, meetings with top management or informal coordination activities. On the other side, delegation in our model refers to what is generally called arm's length contracting and corresponds to more traditional contractor-subcontractor relationships with the contractor being less involved in coordination activities and productive information being transmitted only indirectly through incentive schemes.

3 Comparison between Organizational Forms

Let us now suppose that A_1 can directly acquire perfect information on A_2's type at a finite cost c.[9] This cost of monitoring may represent time and resources spent to perform accounting audits, to organize top-management information sharing or to undertake technological intelligence. It is likely to vary across industries and is certainly the higher the more technologically different the activities of the contractor and the subcontractor are. At the time of incurring this cost, A_1 anticipates that direct monitoring saves on the information rent that he would have to leave to A_2 if he chose instead to stay at arm's length with the sub-contractor. When A_1 is efficient (resp. inefficient), monitoring A_2 costs c instead of the information rent $\nu\Delta\theta\hat{q}$ (resp. $\nu\Delta\theta\bar{q}$) that A_1 must give up to A_2 under arm's length contracting. A_1's choice between monitoring or contracting at arm's length with A_2 depends thus on his own type since this efficiency parameter affects pro-

[7] The reader will have recognized the similarity of this argument with the double-marginalization effect of the I.O. literature (see Spengler (1950)).

[8] The literature on vertical integration has often pointed out that integration changes the informational structure, arguing that information is better obtained between integrated units rather than on market relationships. See Arrow (1975) and Riordan (1990). This literature has nevertheless focused on the incentives to make or buy in a two-agent context whereas the present paper is more interested in the incentive problem related to that choice when it affects a principal.

[9] The fact that acquired information is perfect distinguishes our model from the standard auditing literature (see Mookherjee and Png, 1989 or Strausz, 1997).

duction and thus the information rent left to A_2. This is the *efficiency effect.* Note also that, no matter how information on A_2 is learned by A_1, this is the principal who in fine bears the cost of learning. This cost affects the overall agency cost that the principal has to pay when he wants to implement a given production schedule $(\underline{q}, \hat{q}, \bar{q})$.[10] Noticing that the cost of truthful revelation by A_1 is, as under delegation, the expected information rent left to him, namely $\nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q})$, the overall agency cost can thus a priori take four different expressions:

Case 1: Suppose that A_1 always monitors A_2, the agency cost borne by the principal for implementing outputs $(\underline{q}, \hat{q}, \bar{q})$ is $c + \nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q})$.

Case 2: If A_1 monitors A_2 only when he is efficient $\theta_1 = \underline{\theta}$, the agency cost becomes $\nu(c + (1-\nu)\Delta\theta\bar{q}) + \nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q})$.

Case 3: If A_1 always contracts at arm's length with A_2, the agency cost is the same as under delegation, namely $2\nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q})$.

Case 4: The remaining possibility would be that only the inefficient contractor monitors. As we show below, the principal never chooses to induce monitoring only by the inefficient contractor. The intuition is straightforward. Since A_2's information rent is greater when the output is greater, A_1 is more willing to monitor his peer when he is efficient since the coalition must produce more and the rent left to A_2 is larger. Thus, if A_1 monitors his peer when he is inefficient, he necessarily also monitors when he is efficient.

Gathering the three relevant cases, the total agency cost is finally written as:

$$C(\underline{q}, \hat{q}, \bar{q}) = \nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q}) + \min\{c, \nu(c + (1-\nu)\Delta\theta\bar{q}), \nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q})\}. \quad (10)$$

Although this derivation of the agency cost is rather intuitive, it is not fully rigorous. We must verify that the interplay between the moral hazard and the adverse selection dimensions of the incentive problem does lead to such an agency cost.

Consider for instance Case 1 where the principal wants to induce monitoring by agent A_1. Coalition incentive constraints are then written as under complete information. Participation constraints must nevertheless be adapted to take into account that A_1 incurs a fixed-cost of monitoring and must be reimbursed for doing so by the principal. The inefficient agent A_1's participation constraint is now:

$$\bar{U}_1 = \nu\hat{u} + (1-\nu)\bar{u} - c \geq 0. \quad (11)$$

When accepting the contract proposed by the principal, the contractor must expect a positive payoff anticipating that he will monitor the sub-contractor.

[10] Restricting attention to quantity schedules such that $\hat{q}_1 = \hat{q}_2$ is without loss of generality as we prove in the Appendix.

Whatever his own type, A_1 must also be induced to monitor A_2. This yields the following type-by-type moral hazard constraints of A_1: for the efficient type $\theta_1 = \underline{\theta}$,

$$\nu\underline{u} + (1-\nu)\hat{u} - c \geq \tag{12}$$

$$\max_{\{\Phi(\underline{\theta},\underline{\theta});\Phi(\underline{\theta},\bar{\theta})\}} \left\{ \nu \left(t(\Phi(\underline{\theta},\underline{\theta})) - 2\underline{\theta} q(\Phi(\underline{\theta},\underline{\theta})) \right) + (1-\nu) \left(t(\Phi(\underline{\theta},\bar{\theta})) - \left(\underline{\theta} + \bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta \right) q(\Phi(\underline{\theta},\bar{\theta})) \right) \right\},$$

and for the inefficient type $\theta_1 = \bar{\theta}$,

$$\nu\hat{u} + (1-\nu)\bar{u} - c \geq \tag{13}$$

$$\max_{\{\Phi(\bar{\theta},\underline{\theta});\Phi(\bar{\theta},\bar{\theta})\}} \left\{ \nu \left(t(\Phi(\bar{\theta},\underline{\theta})) - (\underline{\theta} + \bar{\theta}) q(\Phi(\bar{\theta},\underline{\theta})) \right) + (1-\nu) \left(t(\Phi(\bar{\theta},\bar{\theta})) - \left(2\bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta \right) q(\Phi(\bar{\theta},\bar{\theta})) \right) \right\}.$$

The r.h.s. of (12) and (13) represent what A_1 obtains if he does not monitor A_2. In this case, A_1 has to learn indirectly A_2's type through arm's length contracting and he must give up to A_2 some information rent to do so. To evaluate the benefit of reaching an optimal sub-contract with A_2 under these circumstances, one must replace the cost of A_2 by its virtual cost in the r.h.s. of (12) and (13).[11] Consider first the moral hazard constraint for an inefficient agent A_1 and let us try to simplify it. To do so, we must figure out what are the productions chosen under sub-contracting if A_1 does not monitor A_2, given that the principal anticipates that monitoring actually occurs. Remember that, in that case, the coalition produces efficiently and large volumes are expected. If an inefficient agent A_1 does not monitor, the optimal arm's length contract maximizes the virtual surplus of the coalition which is lower than its true surplus. A revealed preference argument shows immediately that, under asymmetric information, the $(\bar{\theta},\bar{\theta})$-coalition is certainly not willing to produce more than $\bar{q}$. Hence, $\Phi^*(\bar{\theta},\bar{\theta}) = (\bar{\theta},\bar{\theta})$ is the best manipulation out of the equilibrium path. When A_1 faces an efficient agent A_2, the virtual surplus of the coalition is instead equal to the true surplus. The production is the same in state $(\bar{\theta},\underline{\theta})$ whether there is complete or asymmetric information within the coalition: $\Phi^*(\bar{\theta},\underline{\theta}) = (\bar{\theta},\underline{\theta})$. For an inefficient agent A_1, the best manipulation out of the equilibrium remains to truthfully report the cost vector to the principal. The incentive constraint (13) is then satisfied when the rent left to an efficient agent A_2, namely $\Delta\theta q(\Phi^*(\bar{\theta},\bar{\theta})) = \Delta\theta\bar{q}$, is greater than the fixed cost of monitoring:

$$c \leq \nu\Delta\theta\bar{q} \tag{14}$$

[11]When deriving (12) and (13), an extensive use is made of Maskin and Tirole (1990). The subcontracting stage following any monitoring decision by A_1 can be analyzed as an informed principal problem in a setting with private values and quasi-linear preferences. We know from the work of these authors that the continuation equilibrium following any monitoring decision by A_1 does not depend on A_2's beliefs on A_1's type. It is thus equivalent to consider that A_2 knows perfectly A_1's type in the subcontracting game. This allows us to derive the moral hazard constraints without worrying about the out-of-equilibrium beliefs that A_2 may hold following an unexpected action of A_1.

Comparing (12) and (13), the most stringent moral hazard constraint is (13) which concerns the inefficient contractor. Surplus losses from arm's length contracting being greater when A_1 is efficient, (12) is easier to satisfy than (14) so that the only moral hazard problem comes from an inefficient A_1. This also eliminates Case 4.

Finally, a contract inducing consolidation implements a schedule of outputs $(\underline{q}, \hat{q}, \bar{q})$ at a cost $C_1(\underline{q}, \hat{q}, \bar{q})$ which is solution to:[12]

$$C_1(\underline{q}, \hat{q}, \bar{q}) = \min_{\{\underline{u}, \hat{u}, \bar{u}\}} \nu^2 \underline{u} + 2\nu(1-\nu)\hat{u} + (1-\nu)^2 \bar{u}$$

subject to (1), (2), (11) and (14).

We show in the Appendix that this cost function can in fact be written as:

$$C_1(\underline{q}, \hat{q}, \bar{q}) = c + \nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q})$$

under the restriction $c \leq \nu\Delta\theta\bar{q}$.

Everything happens thus as if A_1 is reimbursed a fixed amount c for the fixed-cost of monitoring and, on top of that, receives the same expected rent as under costless consolidation. In the Appendix, we verify that the agency cost postulated for the Cases 2 and 3 are also correct.

Finally, the organizational problem faced by the principal consists in finding the least costly way of gathering information in the organization:

$$(\mathcal{P}) : \max_{\{(\underline{q}, \hat{q}, \bar{q})\}} V^*(\underline{q}, \hat{q}, \bar{q}) - C(\underline{q}, \hat{q}, \bar{q}).$$

The presence of a fixed-cost of monitoring introduces several non-convexities in $(\mathcal{P})$. This leads a priori to various regimes with either Case 1, 2 or 3 being optimal, depending on the size of this fixed-cost. The minimal cost paid for implementing a given schedule of outputs $(\underline{q}, \hat{q}, \bar{q})$ results in fact from a choice among several technologies for getting information. These technologies differ with respect to the fixed- and the marginal cost involved. The fixed-cost is the cost of monitoring, the marginal cost is related to the information rent left to the agents. If P wants to induce consolidation, he actually chooses a technology with a high fixed-cost and a low marginal cost. Instead, if P wants arm's length contracting, he chooses a technology without any fixed-cost but with a high marginal cost.

In Case 1, outputs are the same as under costless consolidation. The corresponding profit is nevertheless translated downwards because monitoring is now costly. Similarly, the outputs and the principal's profits in Case 3 are the same as under delegation. The

[12] As usual we write only the relevant constraints. Other incentive and participation constraints are satisfied at the optimum as it can be checked ex post

optimal outputs in Case 2 (referred thereafter as *partial consolidation*) are respectively $q^*(2\underline{\theta})$, $\hat{q}^C$ and $\bar{q}^D$ and the principal's profit is $V^*(q^*(2\underline{\theta}), \hat{q}^C, \bar{q}^D) - \nu\Delta\theta(\nu\hat{q}^C + 2(1-\nu)\bar{q}^D) - \nu c$. It is easy to show that $V^D < V^{PC} = V^*(q^*(2\underline{\theta}), \hat{q}^C, \bar{q}^D) - \nu\Delta\theta(\nu\hat{q}^C + 2(1-\nu)\bar{q}^D) < V^C$. We have drawn on Figure 3 the different expressions of the principal's expected profit depending on c. The principal's profit is the upper envelope of the three different linear parts corresponding to his expected profit in the three different cases above.

Theorem 1 *: There are three possible regimes for the principal's problem:*

- *When $c \in [0, c_1^*]$ where $c_1^* = \frac{V^C - V^{PC}}{1-\nu}$ the principal induces a consolidation between the agents, the vector of outputs is thus $(q^*(2\underline{\theta}), \hat{q}^C, \bar{q}^C)$;*

- *When $c \in [c_2^*, +\infty[$ where $c_2^* = \frac{V^{PC} - V^D}{\nu} > c_1^*$ the principal prefers arm's length contracting between the agents, the vector of outputs is $(q^*(2\underline{\theta}), \hat{q}^D, \bar{q}^D)$;*

- *When $c \in [c_1^*, c_2^*]$, the principal chooses to induce monitoring by A_1 only when the latter is efficient. There is partial consolidation, the vector of outputs is $(q^*(2\underline{\theta}), \hat{q}^C, \bar{q}^D)$.*

- *Even though outputs are discontinuous functions of c, the principal's payoff is continuously decreasing with c.*

Theorem 1 shows that, as the fixed-cost of monitoring increases, the principal's choice goes from a consolidated organization to a partially consolidated one where only the efficient agent A_1 learns directly A_2's type to, finally arm's length contracting. Actually, the existence of the two extreme regimes is hardly surprising and the interest of Theorem 1 lies instead in the identification of the intermediate regime and more specifically in the fact that efficient contractors are first to find it beneficial to consolidate. Intuitively, those contractors produce at higher scales and enjoy more of the efficiency gains that consolidation generates: this is the *efficiency effect*. This intermediate regime is also the main predictive content of our theory of supplier networks. Qualitatively, its existence has the following interpretation: more efficient contractors should monitor more intensively their subcontractors even though efficiency in the productive activity has no direct impact on efficiency in the monitoring activity.[13] From an empirical point of view, the question is thus not that much "do we observe partially consolidated networks ?" but rather "can we explain differences in the monitoring intensities by differences in the productive efficiency parameters of the contractors?"

Another important result of the optimization is that whereas optimal quantities are constant within each regime, organizations do not adapt smoothly to improvements in monitoring technologies but instead will jump discontinuously between three possibilities.

[13]We conjecture that this feature of our model is very robust to continuous generalizations of the discrete assumptions (a continuum of production cost parameters or a continuum of monitoring efforts).

As a result, the optimal organizational choices may not be robust to small perturbations in the various parameters of the model. This suggests that even small improvements in monitoring technology, maybe by means of new information technology or other organizational innovations which facilitate information exchanges, might also be accompanied by radical changes in prices, production and organizational forms. In particular, as monitoring costs decrease, the chain of command is flattened.

Allocative efficiency is only one aspect of the overall comparison between both organizational forms. They also differ in the distribution of rents that they induce. This has clearly an impact on incentives to make specific investments in the relationships in a framework where those investments are non-verifiable and are only driven by the prospects of getting information rents. Under consolidation, A_1 is completely informed on the subcontractor and captures the whole surplus of the coalition. Only A_1 may get a positive rent and has thus incentives to make such a specific investment. Under delegation, both agents receive instead the same positive rents and this more even distribution of rents restores incentives for both. This suggests that our analysis may be possibly biased in favor of consolidation when information structures are not endogenized.

To illustrate, consider the following bare-boned version of the model where $S(q) = S.q$ for some positive S and the possible output q can be either 0 or 1.[14]

By adapting our continuous formulation to that discrete case, it is straightforward to show that one may find numerical values of S, $\underline{\theta}$ and $\bar{\theta}$ such that $\underline{q}^D = \hat{q}^D = 1$ and $\bar{q}^D = 0$ under delegation and $\underline{q}^C = \hat{q}^C = 1$ and $\bar{q}^C = 0$ under consolidation. Suppose now that by investing an amount I ex ante, i.e., before contracting, agents improve the probability of being efficient from ν to $\nu_1 = \nu + \Delta\nu$. These investments are non-verifiable and non-observable and the only incentives for investment come from the agents' prospects for getting an information rent. Under delegation, both agents secure an expected rent $\nu_1.\nu_1\Delta\theta$. At a symmetric Nash equilibrium, they both invest whenever

$$\nu_1\Delta\nu\Delta\theta > I. \tag{15}$$

Instead, under consolidation, the bottom agent gets no rent and never invests, keeping the probability of being efficient at ν. The top agent invests whenever

$$\Delta\nu(\nu\Delta\theta - c) > I. \tag{16}$$

Under delegation, the firm's expected profit can be written as:

$$V^D = \nu_1(2 - \nu_1)(S - 2\underline{\theta}) - 2\nu_1\Delta\theta.$$

[14]This can be viewed as a modeling of the case where the top firm has to build up one large scale project.

Under consolidation, we have instead:

$$V^C = (\nu + \nu_1 - \nu\nu_1)(S - 2\underline{\theta}) - \nu_1 c - (\nu + \nu_1 - \nu\nu_1)\Delta\theta.$$

There exist values of c and I such that (15), (16) and $V^D - V^C > 0$ are verified. This means that delegation can finally be optimal when information structures are endogenous and affected by ex ante investments.

This small model predicts also that consolidated organizations, when they emerge, are characterized by a very asymmetric pattern of investments and by a decreasing expected efficiency along the supply chain. Arm's length contracting is characterized by a much more symmetric distribution of costs between the different layers of the supply chain.

4 Comparison with Centralization

In this section, we compare the results obtained in Theorem 1 with what would occur in a more centralized organization where P could contract with both suppliers directly. Our aim here is first to get a better understanding of the losses and the possible systematic bias coming from the delegation of the monitoring task. In particular, we will show that when the monitoring activity is delegated, the optimal organization is tilted towards too much arm's length contracting. The second objective is to justify our focus on a three-layer hierarchy in the first place. We show in Proposition 4 that once P is constrained to delegate the monitoring task, there is no loss of control in delegating to A_1 control on A_2 as soon as both suppliers would perfectly collude in a centralized organization. The three-layer hierarchy performs as well as a centralized organization where the top firm contracts directly with both suppliers.

- **Centralization and Monitoring by the Principal**

Suppose that the principal can perform the monitoring task by himself instead of delegating this task to the contractor. We model thus a centralized organization where the principal can directly monitor, communicate, and contract with the sub-contractor.

To keep the model as close as possible as before and in particular to allow for monitoring conditional on A_1's type, the grand-contract works sequentially so that the principal decides to monitor A_2 or not as a function of A_1's report (which remains truthful in equilibrium).

The optimal degree of monitoring is quite comparable to what is achieved with delegated monitoring even though the cut-offs between the three regimes may change.

Proposition 2 *: Suppose that the principal can monitor A_2 at cost c and directly communicates and contracts with this supplier. Then, the principal is strictly better off than* *if he delegates monitoring to A_1 as long as the monitoring technology is used (see Figure 3). There exist again two threshold values c_1^p, and c_2^p delimiting three intervals.*

- *When $c \leq c_1^p = c_1^*$, the principal always monitors A_2.*

- *When $c_1^p \leq c \leq c_2^p$ with $c_2^p > c_2^*$ the principal monitors A_2 if and only if A_1 is efficient.*

- *Finally, when $c \geq c_2^p$ the principal acquires information indirectly by letting A_2 report his type directly to him. The same outcome as under delegation is implemented.*

We already saw that, when the principal cannot retain control of the monitoring task, choosing consolidation leaves to the contractor the possibility to better manipulate reports to the principal on both his cost but also that of the sub-contractor. That joint manipulation is no longer an issue when the sub-contractor is directly monitored by the principal. This makes monitoring more attractive in the centralized organization than when monitoring is delegated.

Note also that whether P has control over the monitoring task or not does not change the threshold value between consolidation and partial consolidation. Indeed, the new screening opportunities that that direct control offers to the principal (which concern the mixed coalitions) do not influence the optimal quantity $\bar{q}$, and determine the same cut-off $c_1^* = c_1^P$.

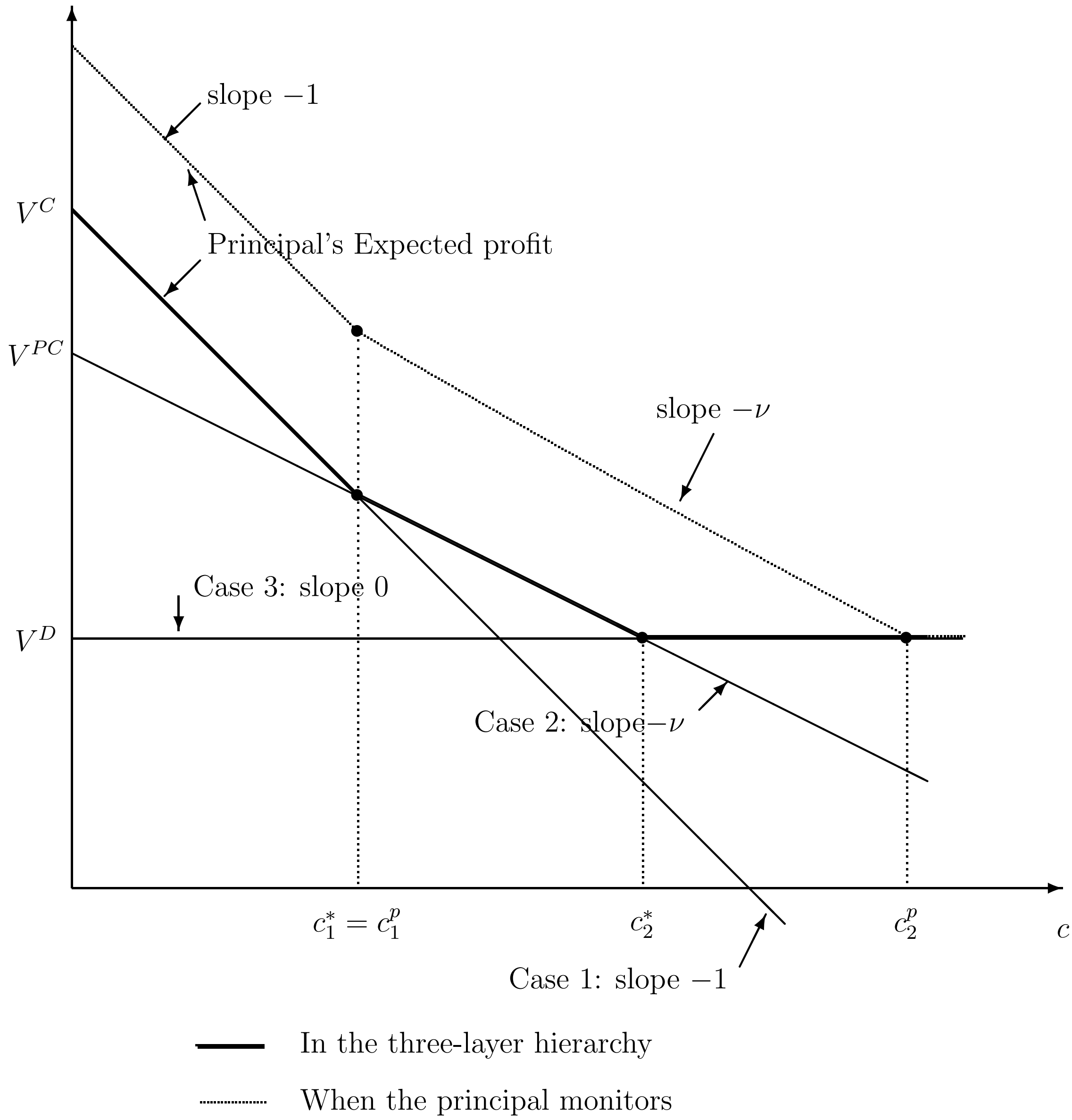

Figure 3: The principal's payoff.

However, the contractor may have a comparative advantage in monitoring other suppliers so that the benefit of delegating that task still outweighs the cost of opening to the contractor new strategic opportunities for manipulating information. The correct comparison is then between delegated monitoring with an efficient technology or keeping a centralized organization but an inefficient monitoring device.

Nevertheless, Proposition 2 suggests also that the principal may also benefit from splitting the task of monitoring suppliers and production. Such separation reduces agency costs by ensuring that the supervisor is unable to internalize all gains from manipulating information. In the case of a productive supervisor, such gains always exist as our analysis shows. As a matter of fact, that supervision and production should be split gives a rationale to a large part of the literature on three-tier hierarchies which, following Tirole (1986), assumes that supervisors do not produce. However, exactly as the contractor has a comparative advantage over the top firm to perform monitoring, he is also likely to have a comparative advantage over any other third-party not involved in the production process. In the case of supplier networks, monitoring is facilitated by technological proximity, justifying the fact that this task cannot be beneficially delegated to an outside agent.

• **Centralization and Delegated Monitoring**

Let us keep the assumption that the principal can communicate and contract with the sub-contractor and let us now again assume that the task of monitoring is necessarily delegated to the contractor. We are particularly interested in the possibility for the two suppliers to collude to promote their collective profits against the top firm. Let us start with a simple benchmark where this collusion does not take place.

Proposition 3 *If the top firm can contract with both agents and prevent collusion between them, then delegating monitoring to the contractor A_1 is costless. The principal's payoff is the same as when he can monitor the other supplier A_2 by himself.*

In such an environment, the principal finds it costless to delegate monitoring to agent A_1. Indeed, the information obtained through monitoring, even though it is non-verifiable, is shared by the two suppliers who do not cooperate. It is thus possible to build a revelation mechanism[15] to exploit this non-cooperative behavior and extract this piece information costlessly. It is enough to have both suppliers reporting information on the monitored agent's cost and to compare those reports. If reports conflict, both agents are heavily punished. With such a scheme, A_2's information can be costlessly extracted by the principal just as if he had monitored himself. Moreover, the fact that the agents can be punished for conflicting reports also provides costless incentives for A_1 to exert monitoring.

When collusion is an issue, such harsh punishments are not available to the principal and delegating the monitoring task may be costly. To model this, suppose that at the collusion stage, A_1 has all bargaining power in offering an enforceable side-contract to A_2.[16] The contractor faces almost the same design problem as in the case of the sub-contract except that now, the reservation utility of agent A_2 may not be zero but is instead fixed by the initial grand-contract offered by the principal. While he had no influence on the sub-contracting game examined in Sections 2 and 3, the principal might now use the

[15]See Maskin (1999).

[16]This collusion proposal takes place at date 4 in the timing of Figure 1 with date 1 corresponding to a grand-contract offered to both agents. Note that keeping the same allocation of the bargaining power in this collusion and in the official contract envisioned in the previous sections of the paper does not introduce any bias in the analysis.

grand-contract to distort side-contracting in his favor by playing on the sub-contractor's status quo payoff if he refuses to collude with the contractor.

Increasing the share of the coalitional surplus that the sub-contractor can always guarantee himself reduces the contractor's incentives to monitor because he can no longer extract all the coalitional rent. As long as the principal wants some form of consolidation, raising the sub-contractor's status quo payoff should thus be costly. Building a centralized mechanism which is collusion-proof and leaves the contractor with zero payoff should be the best that can be done by the principal.

To confirm this intuition, we consider the same game as depicted in Section 2, except that now, agent A_1 must ensure that the sub-contractor gets utility $U_r(\theta_2)$ where $U_r(\theta_2) \geq 0$ at the sub-contracting stage. This utility level is guaranteed by the principal through some initial contract signed with the sub-contractor. The sub-contract must now satisfy, the interim participation constraints of the sub-contractor:

$$y(\theta_1, \underline{\theta}) - \underline{\theta} q(\Phi(\theta_1, \underline{\theta})) \geq U_r(\underline{\theta}),$$

$$y(\theta_1, \bar{\theta}) - \bar{\theta} q(\Phi(\theta_1, \bar{\theta}))) \geq U_r(\bar{\theta}),$$

in addition to the usual incentive and participation constraints.

Proposition 4 *Suppose that $U_r(\theta_2)$ can be chosen by P. Then he optimally sets $U_r(\theta_2) \equiv 0$ and gets the same payoff as in the three-layer hierarchy model.*

This proposition is important because it validates our focus on a three-layer hierarchy to analyze the incentives to consolidate. This seemingly decentralized organizational form is nothing else that a possible implementation of the optimal centralized arrangement when collusion is an issue.[17]

5 Conclusion

Our study of supplier networks outlines the importance of the monitoring aspect of the contractor-subcontractor relationship. Information flows between suppliers, necessary to enhance coordination, cannot be obtained without an active involvement by the contractor. As a consequence, the top firm must simultaneously shape the incentives for efficient production and for efficient information sharing. It turns out that the intensity of monitoring depends of course on the monitoring costs (high intensity or consolidation for low costs, low intensity or arm's length contracting for high costs) but more interestingly

[17]On that issue, see also Faure-Grimaud, Laffont and Martimort (2003) who obtain a similar result in a model where supervisory information is exogenous and not endogenized as here.

also on the productive efficiency of the contractor. Through the *efficiency effect* a more efficient contractor has more incentives to monitor.

By endogenizing these incentives, our model gives a unified treatment of a whole literature on organizational design which either assumes that agents exogenously share information and contract efficiently or do not share information and contract at arm's length under asymmetric information.

Several extensions of our work should be worth to pursue both on the empirical and theoretical sides.

First, our analysis is certainly amenable to several empirical tests. The main testable prediction of our model is certainly linked to the *efficiency effect* and the fact that the intensity of monitoring should be positively correlated with the contractor's efficiency and profit. Consolidation should emerge for high production volumes whereas delegation correspond to smaller scales of activities. It is also worth noticing that if one wants to recover the distribution of the monitoring costs from observing organizational structures, not taking into account the intermediate regime where consolidation may or may not arise depending on the contractor's efficiency would introduce a significant bias. Lastly, extending our model to take into account specific investments which improve costs also predicts that organizations where delegation appears will have a more symmetric pattern of investments and efficiency whereas strong asymmetry may arise in consolidated organizations.

On the theory side, it would be interesting to introduce more than two suppliers and analyze the determinants of the clusters of consolidated activities. That extension could be useful as a first step towards an incentive theory of networks.[18] Adding multiple contractors and sub-contractors may be useful to discuss the issue of competition at each layer and how it affects networking.

Finally, while much of the early literature on organizational design, like our model, concentrates on Leontieff production functions, some recent papers (Mookherjee and Tsumagari (2003) and Severinov (2003) most noticeably), stress the central role of this hypothesis for the results concerning the superiority of consolidation and the equivalence between delegation and direct contracting.[19] Considering substitutes may change these properties. Our main results are nevertheless robust to a weakening of the Leontieff hypothesis. To explain this, note, first, that in our model the fact that costless consolidation performs better than delegation simply comes from the fact that coalitional virtual costs are higher than coalitional true costs. When the principal faces a consolidation he must take

[18]See Kranton and Minehart (2000 and 2001) for models of network formation and its consequence on specific investments. The issue of information and incentives is not modeled in those papers.

[19]See also Dana (1993) for a weakening of the perfect complement hypothesis.

into account true costs while he must take into account virtual costs under delegation. As it brings him higher profits to face lower costs (in the sense of first-order stochastic dominance), the principal prefers consolidation. Then, with a more general production function, the qualitative features of the optimal contracting arrangement would remain the same: consolidation for low monitoring costs, partial consolidation for an intermediate range of values and delegation for high costs; however proving these results would involve much more technicalities.

We hope to investigate some of these issues in future research.

References

Arrow, K., 1975, "Vertical Integration and Communication," *Bell Journal of Economics*, 6: 173-183.

Baiman, S. and M. Rajan, 2002, "The Role of Information and Opportunism on the Choice of Buyer-Seller Relationship," *Journal of Accounting Research*, 40(2): 247-278.

Bala, V. and S. Goyal, 2000, "A Non-Cooperative Model of Network Formation," *Econometrica*, 66: 743-768.

Baron, D. and D. Besanko, 1992, "Information, Control, and Organizational Structure," *Journal of Economics and Management Strategy*, 1: 237-275.

Baron, D. and D. Besanko, 1999, "Informational Alliances," *Review of Economic Studies*, 66: 743-768.

Dana, J., 1993, "The Organization and Scope of Agents: Regulating Multiproduct Industries," *Journal of Economic Theory*, 59: 282-310.

Faure-Grimaud, A., J.J. Laffont and and D. Martimort, 2003,"Collusion, Delegation and Supervision with Soft Information," *Review of Economic Studies*, 70(2): 253-280.

Gilbert, R. and M. Riordan, 1995, "Regulating Complementary Products: A Comparative Institutional Analysis," *RAND Journal of Economics*, 26: 243-256.

Hines, P. and N. Rich, 1998, "Outsourcing Competitive Advantage: the Use of Supplier Associations," *International Journal of Physical Distribution and Logistics Management*, 28(7): 524-546.

Jackson, M. and A. Wolinski, 1996, "A Strategic Model of Economic and Social Networks," *Journal of Economic Theory,* 71: 44-74.

Kofman, F. and J. Lawarrée, 1993, "Collusion in Hierarchical Agency," *Econometrica*, 61: 629-656.

Kranton, R. and D. Minehart, 2000, "Networks versus Vertical Integration," *RAND Journal of Economics*, 31: 570-601.

Kranton, R. and D. Minehart, 2001, "A Theory of Buyer-Seller Networks," *American Economic Review*, 91: 485-508.

Laffont, J.-J. and D. Martimort, 1997, "Collusion under Asymmetric Information," *Econometrica*, 65: 875-911.

Laffont, J.-J. and D. Martimort, 1998, "Collusion and Delegation," *RAND Journal of Economics*, 29: 280-305.

Laffont, J.-J. and J. Tirole, 1991, "The Politics of Government Decision-Making: a Theory of Regulatory Capture," *Quarterly Journal of Economics*, 106: 1089-1127.

Maskin, E., 1999, "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 229: 929-948.

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

Melumad, N., Mookherjee, D. and S. Reichelstein, 1995, "Hierarchical Decentralization of Incentive Contracts," *RAND Journal of Economics*, 26: 654-672.

Mookherjee, D. and I. Png, 1989, "Optimal Auditing, Insurance and Redistribution," *Quarterly Journal of Economics*, 102: 399-415.

Mookherjee, D. and S. Reichelstein, 2001, "Incentives and Coordination in Hierarchies," *Advances in Theoretical Economics:* Vol. 1 (article 4), http://www.bepress.com/beje/advances/vol1/iss1/art4.

Mookherjee, D. and M. Tsumagari, 2003, "The Organization of Supplier Networks: Effects of Mergers and Delegation," *forthcoming Econometrica.*

Perrow, C., 1992, "Small-Firm Networks," in *Networks and Organizations*, N. Nohria and R. Eclles (eds.), Harvard Business School Press.

Riordan, M., 1990, "What is Vertical Integration," in M. Aoki, B. Gustaffson and O. Williamson (eds.), *The Firm as a Nexus of Contracts*, Free Press.

Severinov, S., 2003, "Optimal Organization: Centralization, Decentralization or Delegation ?," *mimeo.*

Spengler, J., 1950, "Vertical Integration and Antitrust Policy," *Journal of Political Economy,* 58: 347-352.

Strausz, R., 1997, "Delegation of Monitoring in a Principal-Agent Relationship," *Review of Economic Studies*, 64: 337-357.

Tirole, J., 1986, "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics and Organization*, 2: 181-214.

Yun, M., 1999, "Subcontracting Relations in the Korean Automotive Industry: Risk Sharing and Technological Capabilities," *International Journal of Industrial Organization*, 17: 81-108.

Appendix

• **Proof of Proposition 1: Costless Consolidation**:Under complete information within the coalition, the following coalition incentive constraints must be satisfied to induce information revelation:

$$\underline{u} \geq \max\{\hat{u}_1 + \Delta\theta\hat{q}_1; \hat{u}_2 + \Delta\theta\hat{q}_2; \bar{u} + 2\Delta\theta\bar{q}\}, \tag{17}$$

$$\hat{u}_1 \geq \max\{\hat{u}_2; \underline{u} - \Delta\theta\underline{q}; \bar{u} + \Delta\theta\bar{q}\}, \tag{18}$$

$$\hat{u}_2 \geq \max\{\hat{u}_1; \underline{u} - \Delta\theta\underline{q}; \bar{u} + \Delta\theta\bar{q}\}, \tag{19}$$

$$\bar{u} \geq \max\{\hat{u}_1 - \Delta\theta\hat{q}_1; \hat{u}_2 - \Delta\theta\hat{q}_2; \underline{u} - 2\Delta\theta\underline{q}\}. \tag{20}$$

Clearly (18) and (19) imply that the two mixed coalitions cannot be screened apart both in terms of their aggregate payoffs $\hat{u}_1 = \hat{u}_2 = \hat{u}$, but also in terms of their respective outputs[20] $\hat{q}_1 = \hat{q}_2 = \hat{q}$. The two upward coalition incentive constraints give thus (1) and (2). Participation constraints are written at the interim stage, knowing that agent A_1 will capture all the coalitional rent, as (3) when A_1 is inefficient and $\nu\underline{u} + (1-\nu)\hat{u} \geq 0$ when he is efficient. First, note that (1), (2) and (3) are binding since the principal wants to reduce $\underline{u}$, $\hat{u}$ and $\bar{u}$ as much as possible. This yields the following expressions for the rents: $\underline{u} = \Delta\theta\hat{q} + \nu\Delta\theta\bar{q}$, $\hat{u} = \nu\Delta\theta\bar{q}$, $\bar{u} = -(1-\nu)\Delta\theta\bar{q}$. With those expressions, we directly obtain (6) and $\underline{U}^C > 0$ since outputs are positive. Inserting those expressions into the objective function and optimizing yields $\underline{q}^C = q^*(2\underline{\theta})$ and (7). We can now check that the remaining coalition incentive constraints are satisfied.

Delegation: Consider the design of the optimal sub-contract. Fix A_1's type θ_1. From Maskin and Tirole (1990), everything happens as if θ_1 were known by A_2 at the sub-contracting stage. Fixing P's grand-contract, an intermediate agent A_1 with type θ_1 finds the optimal sub-contract as a solution to:

$$(\mathcal{P}^{SC}): \quad \max_{\{(\underline{U}_2(\theta_1),\Phi(\theta_1,\underline{\theta}));(\bar{U}_2(\theta_1),\Phi(\theta_1,\bar{\theta}))\}} \nu(t(\Phi(\theta_1,\underline{\theta})) - (\theta_1+\underline{\theta})q(\Phi(\theta_1,\underline{\theta})) - \underline{U}_2(\theta_1))$$
$$+(1-\nu)(t(\Phi(\theta_1,\bar{\theta})) - (\theta_1+\bar{\theta})q(\Phi(\theta_1,\bar{\theta})) - \bar{U}_2(\theta_1))$$

[20]To see this last point, note that $\hat{q}_1$ and $\hat{q}_2$ play a symmetric role in the incentive constraints (the participation constraints are still missing but one can verify that $\hat{q}_1$ and $\hat{q}_2$ also play a symmetric role in these constraints). If the contract $(\underline{u}, \hat{u}, \hat{u}, \bar{u}, \underline{q}, \hat{q}_1, \hat{q}_2, \bar{q})$ satisfies these constraints, so does the contract $(\underline{u}, \hat{u}, \hat{u}, \bar{u}, \underline{q}, \hat{q}_2, \hat{q}_1, \bar{q})$ which brings exactly the same payoff. The principal's objective function being concave w.r.t. outputs, the principal can weakly increase his expected payoff by offering $(\underline{u}, \hat{u}, \hat{u}, \bar{u}, \underline{q}, \frac{\hat{q}_1+\hat{q}_2}{2}, \frac{\hat{q}_1+\hat{q}_2}{2}, \bar{q})$.

subject to

$$\underline{U}_2(\theta_1) = y(\theta_1, \underline{\theta}) - \underline{\theta} q(\Phi(\theta_1, \underline{\theta})) \geq y(\theta_1, \bar{\theta}) - \underline{\theta} q(\Phi(\theta_1, \bar{\theta})), \tag{21}$$

$$\bar{U}_2(\theta_1) = y(\theta_1, \bar{\theta}) - \bar{\theta} q(\Phi(\theta_1, \bar{\theta})) \geq 0, \tag{22}$$

where $\underline{U}_2(\theta_1)$ (resp. $\bar{U}_2(\theta_1)$) is A_2's information rent when he is efficient (resp. inefficient) and $\Phi(\cdot)$ denotes the report function chosen by A_1. Writing the conditions for optimality of the null manipulation of reports ($\Phi^*(\theta_1, \theta_2) = (\theta_1, \theta_2)$ for all (θ_1, θ_2)) yields the following coalition incentive constraints:

$$\underline{u} \geq \max\{\hat{u}_1 + \Delta\theta \hat{q}_1; \hat{u}_2 + \Delta\theta \hat{q}_2; \bar{u} + 2\Delta\theta \bar{q}\}, \tag{23}$$

$$\hat{u}_1 - \frac{\nu}{1-\nu}\Delta\theta \hat{q}_1 \geq \max\left\{\hat{u}_2 - \frac{\nu}{1-\nu}\Delta\theta \hat{q}_2; \underline{u} - \Delta\theta \underline{q} - \frac{\nu}{1-\nu}\Delta\theta \underline{q}; \bar{u} + \Delta\theta \bar{q} - \frac{\nu}{1-\nu}\Delta\theta \bar{q}\right\}, \tag{24}$$

$$\hat{u}_2 \geq \max\{\hat{u}_1; \underline{u} - \Delta\theta \underline{q}; \bar{u} + \Delta\theta \bar{q}\}, \tag{25}$$

$$\bar{u} - \frac{\nu}{1-\nu}\Delta\theta \bar{q} \geq \max\left\{\hat{u}_1 - \Delta\theta \hat{q}_1 - \frac{\nu}{1-\nu}\Delta\theta \hat{q}_1; \hat{u}_2 - \Delta\theta \hat{q}_2 - \frac{\nu}{1-\nu}\Delta\theta \hat{q}_2; \underline{u} - 2\Delta\theta \underline{q} - \frac{\nu}{1-\nu}\Delta\theta \underline{q}\right\}. \tag{26}$$

Since he anticipates that he will have to leave some information rent to A_2 if the latter is efficient, A_1's participation constraints at the interim stage are:

$$\underline{U}_1 = \nu(\underline{u} - \Delta\theta \hat{q}_1) + (1-\nu)\hat{u}_1 \geq 0, \tag{27}$$

$$\bar{U}_1 = \nu(\hat{u}_2 - \Delta\theta \bar{q}) + (1-\nu)\bar{u} \geq 0. \tag{28}$$

Consider the relaxed maximization program of the principal:

$$(\mathcal{P}^{D'}) : \quad \max_{\{(\underline{u},\underline{q});(\hat{u}_1,\hat{q}_1);(\hat{u}_2,\hat{q}_2);(\bar{u},\bar{q})\}} \nu^2(S(\underline{q}) - 2\underline{\theta}\underline{q} - \underline{u}) + \nu(1-\nu)(S(\hat{q}_1) - (\underline{\theta} + \bar{\theta})\hat{q}_1 - \hat{u}_1)$$

$$+\nu(1-\nu)(S(\hat{q}_2) - (\underline{\theta} + \bar{\theta})\hat{q}_2 - \hat{u}_2) + (1-\nu)^2(S(\bar{q}) - 2\bar{\theta}\bar{q} - \bar{u})$$

subject to $\underline{u} \geq \hat{u}_2 + \Delta\theta \hat{q}_2$, $\hat{u}_2 \leq \hat{u}_1$, $\hat{u}_1 - \frac{\nu}{1-\nu}\Delta\theta \hat{q}_1 \geq \bar{u} + \Delta\theta \bar{q} - \frac{\nu}{1-\nu}\Delta\theta \bar{q}$, and (28).

Optimizing yields $\hat{q}_1^D = \hat{q}_2^D = \hat{q}^D$, $\underline{q}^D = q^*(2\underline{\theta})$ and (9). We can verify now that all the remaining constraints are satisfied and that we identified the solution of the principal's program.

• Let us prove now that this contract is also optimal when P can contract with both suppliers. If we denote by $(\underline{u}^i, \hat{u}_1^i, \hat{u}_2^i, \bar{u}^i)$ the individual rent schedule of agent i in that situation (with obvious notations), we know that the following upward incentive constraints are binding

$$\nu\underline{u}^i + (1-\nu)\hat{u}_i^i \geq \nu(\hat{u}_j^i + \Delta\theta\hat{q}_j) + (1-\nu)(\bar{u}^i + \Delta\theta\bar{q}). \tag{29}$$

The inefficient type's participation constraint also is binding

$$\nu\hat{u}_j^i + (1-\nu)\bar{u}^i \geq 0. \tag{30}$$

The sum of the expected rents left by the principal is thus $\nu(\nu\Delta\theta(\hat{q}_1+\hat{q}_2))+2(1-\nu)\nu\Delta\theta\bar{q}$, which is the same value as in the delegation case. This leads to the same optimal quantity schedule. Moreover, because (29) and (30) are binding, we can verify that the rent schedules are also the same in both cases.

• **Proof of Theorem 1**

Derivation of $C_1(\underline{q}, \hat{q}, \bar{q})$: First, note that a grand-contract must satisfy (17) to (20) to induce information revelation from A_1 once he has learnt A_2's type. Let us denote by $\widetilde{\Phi}(\bar{\theta}, \cdot)$ the maximand on the right-hand side of (13) and assume for the time being that the implemented quantity and rent schedules are symmetric (i.e. $\hat{q}_1 = \hat{q}_2 = \hat{q}$ and $\hat{u}_1 = \hat{u}_2 = \hat{u}$). Using (19), we get $\widetilde{\Phi}(\bar{\theta}, \underline{\theta}) = (\bar{\theta}, \underline{\theta})$. Using (20), we obtain

$$\bar{u} \geq \hat{u} - \Delta\theta\hat{q} \geq \hat{u} - \Delta\theta\hat{q} - \frac{\nu}{1-\nu}\Delta\theta(\hat{q} - \bar{q})$$

since $\bar{q} \leq \hat{q}$. Similarly, $\bar{u} \geq \underline{u} - 2\Delta\theta\underline{q} \geq \underline{u} - 2\Delta\theta\underline{q} - \frac{\nu}{1-\nu}\Delta\theta(\underline{q} - \bar{q})$ since $\bar{q} \leq \underline{q}$. Henceforth, $\widetilde{\Phi}(\bar{\theta}, \bar{\theta}) = (\bar{\theta}, \bar{\theta})$. (13) rewrites as $\nu\hat{u} + (1-\nu)\bar{u} - c \geq \nu(\hat{u} - \Delta\theta\bar{q}) + (1-\nu)\bar{u}$ and simplifying yields condition (14).

Second, let us compute the maximands on the right-hand side of (12). We denote by $\widetilde{\Phi}(\underline{\theta}, \cdot)$ those maximands. From (17), $\widetilde{\Phi}(\underline{\theta}, \underline{\theta}) = (\underline{\theta}, \underline{\theta})$. Consider now a $(\theta_1 = \underline{\theta}, \theta_2 = \bar{\theta})$-coalition. Note that $\hat{u} - \frac{\nu}{1-\nu}\Delta\theta\hat{q} \geq \underline{u} - \Delta\theta\underline{q} - \frac{\nu}{1-\nu}\Delta\theta\underline{q}$, since $\hat{u} \geq \underline{u} - \Delta\theta\underline{q}$ and $\underline{q} \geq \hat{q}$. Hence the only issue is to determine whether this mixed coalition says the truth or lies upward (i.e., announces $(\bar{\theta}, \bar{\theta})$). Equivalently to (12), we can write two moral hazard constraints, each one corresponding to one report of this coalition: when the $(\theta_1 = \underline{\theta}, \theta_2 = \bar{\theta})$-coalition tells the truth, the moral hazard constraint rewrites as $\nu\underline{u} + (1-\nu)\hat{u} - c \geq \nu(\underline{u} - \Delta\theta\hat{q}) + (1-\nu)\hat{u}$ or

$$c \leq \nu\Delta\theta\hat{q}, \tag{31}$$

when the $(\theta_1 = \underline{\theta}, \theta_2 = \bar{\theta})$-coalition lies upward, the moral hazard constraint rewrites as $\nu\underline{u} + (1-\nu)\hat{u} - c \geq \nu(\underline{u} - \Delta\theta\bar{q}) + (1-\nu)\bar{u}$, or

$$c \leq \nu\Delta\theta\bar{q} + (1-\nu)(\hat{u} - \bar{u} - \bar{q}). \tag{32}$$

That is to say (12) implies (31) and (32), and (31) and (32) imply (12). Because $\hat{q} \geq \bar{q}$ and (2) holds, (31) and (32) are always verified when (14) holds.

Let us now derive $C_1(\underline{q}, \hat{q}, \bar{q})$. As long as (12) is not binding, (1), (2) and (11) are all binding to minimize C_1. Therefore, $C_1(\underline{q}, \hat{q}, \bar{q}) = c + \nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q})$ when $c \leq \nu\Delta\theta\bar{q}$. Moreover, as long as we can neglect the moral hazard constraints (i.e. as long as $c \leq \nu\Delta\theta\bar{q}^C$) we know from the study of the costless consolidation case that the principal cannot do better with an asymmetric contract.

Derivation of $C_2(\underline{q}, \hat{q}, \bar{q})$: Any incentive compatible grand-contract inducing monitoring only by the efficient type of A_1 must be such that (17) and (18) (coalitions where A_1 is efficient) and (25) and (26) (coalitions where A_1 is inefficient) hold together. Using (18) and (25), we obtain immediately that $\hat{u}_1 = \hat{u}_2 = \hat{u}$. As $\hat{q}_1$ and $\hat{q}_2$ play a symmetric role in all the relevant constraints of the program, there is no loss of generality in looking for the optimal contract in the class of symmetric contracts such that $\hat{q}_1 = \hat{q}_2$.

Let us turn now to the moral hazard constraints. A_1 must be induced to learn directly information on A_2 if and only if he is efficient. For the efficient type $\theta_1 = \underline{\theta}$, the moral hazard constraint is still (12). For the inefficient type $\theta_1 = \bar{\theta}$, it is now:

$$\nu\hat{u}_2 + (1-\nu)\bar{u} - \nu\Delta\theta\bar{q} \geq \tag{33}$$

$$\max_{\{\Phi(\bar{\theta},\underline{\theta});\Phi(\bar{\theta},\bar{\theta})\}} \{\nu(t(\Phi(\bar{\theta},\underline{\theta})) - (\underline{\theta}+\bar{\theta})q(\Phi(\bar{\theta},\underline{\theta}))) + (1-\nu)(t(\Phi(\bar{\theta},\bar{\theta})) - 2\bar{\theta}q(\Phi(\bar{\theta},\bar{\theta}))) - c\}.$$

At this stage, we cannot guess whether $\Phi^*(\underline{\theta},\bar{\theta})$ is equal to $(\underline{\theta},\bar{\theta})$ or to $(\bar{\theta},\bar{\theta})$. The fact that the grand-contract is incentive compatible is not sufficient to conclude that the contractor behaves truthfully even out of the equilibrium path. This lack of determination only allows us to transform equation (12) into:

$$c \leq \min\{\nu\Delta\theta\hat{q}; \nu\Delta\theta\bar{q} + (1-\nu)(\hat{u} - \bar{u} - \Delta\theta\bar{q})\}. \tag{34}$$

We now have to show that (33) can be written as

$$c \geq \nu\Delta\theta\bar{q} + \max\{0; (1-\nu)(\hat{u} - \bar{u} - \Delta\theta\hat{q})\}. \tag{35}$$

Let us consider the r.h.s. of (33) and denote by $\widetilde{\Phi}(\bar{\theta},\cdot)$ its maximand. Due to the symmetry of the optimal contract and to (18), we have necessarily $\widetilde{\Phi}(\bar{\theta},\underline{\theta}) = (\bar{\theta},\underline{\theta})$. A coalition made of an inefficient A_1 and an efficient A_2 finds it optimal to tell the truth even outside the equilibrium path. $\widetilde{\Phi}(\bar{\theta},\bar{\theta})$ can instead take one of two values. As (24) holds and $\hat{q} \leq \underline{q}$, we have $\hat{u} - \Delta\theta\hat{q} - \frac{\nu}{1-\nu}\Delta\theta\hat{q} \geq \underline{u} - 2\Delta\theta\underline{q} - \frac{\nu}{1-\nu}\Delta\theta\underline{q}$ and the $(\bar{\theta},\bar{\theta})$ coalition prefers to report $(\bar{\theta},\underline{\theta})$ than $(\underline{\theta},\underline{\theta})$. The issue is then to know whether this coalition lies downward and "locally" (i.e., says $(\bar{\theta},\underline{\theta})$) or tells the truth. To each case corresponds one moral hazard constraint. Combining the two constraints obtained gives (35) .

Let us now compute $C_2(\underline{q}, \hat{q}, \bar{q})$. First note that as soon as $c \neq \nu\Delta\theta\bar{q}$, one of the moral hazard constraints (34) or (35) must be binding, because if none is binding, one can check that (1) is binding at the optimum and then (34) and (35) imply that $c = \nu\Delta\theta\bar{q}$. When $\nu\Delta\theta\hat{q} > c > \nu\Delta\theta\bar{q}$, (1) cannot be binding anymore and, as the principal wants to reduce $\hat{u}$ as much as possible, (34) must be the binding moral hazard constraint (if (35) were binding, the principal could increase $\bar{u}$ and decrease $\hat{u}$ in such a way that the implementation would be less costly for him). Thus when minimizing C_2, the binding constraints are (1), (5) and (34) (one can check ex post that all the other incentive and participation constraints are satisfied): $C_2(\underline{q}, \hat{q}, \bar{q}) = \nu c + \nu\Delta\theta(\nu\hat{q} + 2(1-\nu)\bar{q})$.

Derivation of $C_3(\underline{q}, \hat{q}, \bar{q})$: A grand-contract that prevents monitoring must satisfy (23) to (26) to be delegation-proof. The moral hazard constraints are (33) for an inefficient agent $\theta_1 = \bar{\theta}$ and

$$\nu\underline{u} + (1-\nu)\hat{u}_1 - \nu\Delta\theta\hat{q}_1 \geq \tag{36}$$

$$\max_{\{\Phi(\underline{\theta},\underline{\theta});\Phi(\underline{\theta},\bar{\theta})\}} \{\nu(t(\Phi(\underline{\theta},\underline{\theta})) - 2\underline{\theta}q(\Phi(\underline{\theta},\underline{\theta}))) + (1-\nu)(t(\Phi(\underline{\theta},\bar{\theta})) - (\underline{\theta}+\bar{\theta})q(\Phi(\underline{\theta},\bar{\theta}))) - c\}$$

for an efficient contractor $\theta_1 = \underline{\theta}$.

The r.h.s. of equation (36) corresponds to what an efficient contractor obtains if he does learn directly A_2's type, minus the cost of this direct learning. Let us denote by $\widetilde{\Phi}(\underline{\theta}, .)$ the maximand of its r.h.s.. We infer from (23) that $\widetilde{\Phi}(\underline{\theta},\underline{\theta}) = (\underline{\theta},\underline{\theta})$, and from (25) that $\widetilde{\Phi}(\underline{\theta},\bar{\theta}) = (\bar{\theta},\underline{\theta})$. Hence (36) can be written $\nu\underline{u} + (1-\nu)\hat{u}_1 - \nu\Delta\theta\hat{q}_1 \geq \nu\underline{u} + (1-\nu)\hat{u}_2 - c$, or after simplification $c \geq \nu\Delta\theta\hat{q}_1 + (1-\nu)(\hat{u}_2 - \hat{u}_1)$.

Consider now equation (33). We know from the derivation of $C_2(\underline{q}, \hat{q}, \bar{q})$ that it is equivalent to the following constraints:

$$c \geq \max\{\nu\Delta\theta\bar{q}; \nu\Delta\theta\bar{q} + (1-\nu)(\hat{u}_1 - \Delta\theta\hat{q}_1 - \bar{u}); \nu\Delta\theta\bar{q} + (1-\nu)(\hat{u}_2 - \Delta\theta\hat{q}_2 - \bar{u})\}.$$

We want to show that when $c \geq \nu\Delta\theta\hat{q}_1 + (1-\nu)(\hat{u}_2 - \hat{u}_1)$, this constraint is automatically satisfied. For the first term, it comes from the fact that due to (25), $\nu\Delta\theta\hat{q}_1 + (1-\nu)(\hat{u}_2 - \hat{u}_1) \geq \nu\Delta\theta\hat{q}_1$ and due to (24) and (26), $\nu\Delta\theta\hat{q}_1 \geq \nu\Delta\theta\bar{q}$. For the second one, we already know that $c \geq \nu\Delta\theta\hat{q}_1 + (1-\nu)(\hat{u}_2 - \hat{u}_1)$ implies that c is higher than $\nu\Delta\theta\hat{q}_1$. But due to constraint (26), we have (it is just another way to write the constraint): $\nu\Delta\theta\hat{q}_1 \geq \nu\Delta\theta\bar{q} + (1-\nu)(\hat{u}_1 - \Delta\theta\hat{q}_1 - \bar{u})$. For the third one, we can use the fact that $\hat{q}_2 \geq \hat{q}_1$ and constraint (26) to obtain $\bar{u} - \frac{\nu}{1-\nu}\Delta\theta\bar{q} \geq \hat{u}_1 - \Delta\theta\hat{q}_2 - \frac{\nu}{1-\nu}\Delta\theta\hat{q}_1$ or equivalently $\nu\Delta\theta\hat{q}_1 + (1-\nu)(\hat{u}_2 - \hat{u}_1) \geq \nu\Delta\theta\bar{q} + (1-\nu)(\hat{u}_2 - \Delta\theta\hat{q}_2 - \bar{u})$. Hence, we have shown that once the agent is given the incentives not to learn directly information when he is efficient, he has also no incentives to learn directly information when he is inefficient.

The condition $c \geq \nu\Delta\theta\hat{q}_1 + (1-\nu)(\hat{u}_2 - \hat{u}_1)$ ensures that no moral hazard constraint is binding and thus the agency cost is exactly the same as under delegation and given by $C_3(\underline{q}, \hat{q}, \bar{q}) = 2\nu\Delta\theta(\nu\hat{q} + (1-\nu)\bar{q})$.

We are now ready to prove the proposition and the theorem. The only thing we have to check is that the conditions on c used for the characterization of the different agency costs are indeed verified when the principal optimally chooses between the three types of organizations and contracts. Consider the situation where the principal induces direct learning for both types of A_1; the expression of $C_1(\underline{q}, \hat{q}, \bar{q})$ valid if $c \leq \nu\Delta\theta\bar{q}^C$. Hence we have to check that $c_1^* \leq \nu\Delta\theta\bar{q}^C$. After a quick reasoning, it is sufficient to show that the following relations hold $\nu\Delta\theta\bar{q}^D \leq c_1^* \leq \nu\Delta\theta\bar{q}^C$, $\nu\Delta\theta\hat{q}^D \leq c_2^* \leq \nu\Delta\theta\hat{q}^C$ and $c_1^* \leq c_2^*$.

Let us start by the last one. As $\bar{q}^C$ verifies $S'(\bar{q}^C) = \underline{\theta} + \bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta$ and $\hat{q}^D$ verifies $S'(\hat{q}^D) = 2\bar{\theta} + \frac{\nu}{1-\nu}\Delta\theta$, for all values of the parameters $\bar{q}^C \leq \hat{q}^D$; and the relation $c_1^* \leq c_2^*$ is just a consequence of the two other relations.

Consider now the relation $\nu\Delta\theta\bar{q}^D \leq c_1^* \leq \nu\Delta\theta\bar{q}^C$. To prove it we will proceed as follows: when c is equal to $\nu\Delta\theta\bar{q}^D$, we will prove that there exists a contract that satisfies all the consolidation constraints[21] and that brings to the principal the same payoff as the optimal partial consolidation contract associated with this value of c. Hence it will show that when $c = \nu\Delta\theta\bar{q}^D$, the principal prefers to induce monitoring by both types and $\nu\Delta\theta\bar{q}^D \leq c_1^*$. Consider the following rent and quantity schedules:

$$\begin{cases} \underline{q} = \underline{q}^* & \underline{u} = \Delta\theta(\hat{q}^C + \bar{q}^D) \\ \hat{q} = \hat{q}^C & \hat{u} = \Delta\theta\bar{q}^D \\ \bar{q} = \bar{q}^D & \bar{u} = 0. \end{cases}$$

It is the optimal schedule under partial consolidation and one can check that all the consolidation constraints are satisfied (for $c = \nu\Delta\theta\bar{q}^D$). As it is not the optimal contract under consolidation, we can conclude that the principal can do better if he implements monitoring by both types of A_1. Consider the case $c = \nu\Delta\theta\bar{q}^C$ and the following rent and quantity schedules:

$$\begin{cases} \underline{q} = \underline{q}^* & \underline{u} = \Delta\theta(\hat{q}^C + \bar{q}^C) \\ \hat{q} = \hat{q}^C & \hat{u} = \Delta\theta\bar{q}^C \\ \bar{q} = \bar{q}^C & \bar{u} = 0. \end{cases}$$

It is the optimal contract under consolidation and it satisfies all the partial consolidation constraints. We deduce that when $c = \nu\Delta\theta\bar{q}^C$, the principal prefers to induce monitoring only if A_1 is efficient. We proved so far that $\nu\Delta\theta\bar{q}^D \leq c_1^* \leq \nu\Delta\theta\bar{q}^C$. The proof of the relation $\nu\Delta\theta\hat{q}^D \leq c_2^* \leq \nu\Delta\theta\hat{q}^C$ follows exactly the same lines and is left to the reader.

[21] That is to say a contract that is incentive compatible and individually rational in the case where the principal wants to induce monitoring by both types.

Proofs which can be left on the Website

• **Proof of Proposition 2:** Let us use a superscript i to denote A_i's utility. In this setting, there are no moral hazard constraints. Suppose that the principal monitors A_2 whatever A_1's report, the binding constraints are $\nu\hat{u}_2^1 + (1-\nu)\bar{u}^1 \geq 0$, $\nu\underline{u}^1 + (1-\nu)\hat{u}_1^1 \geq \nu\hat{u}_2^1 + (1-\nu)\bar{u}^1 + \nu\Delta\theta\hat{q}_2 + (1-\nu)\Delta\theta\bar{q}$, $\underline{u}^2 \geq 0$, $\hat{u}_1^2 \geq 0$, $\hat{u}_2^2 \geq 0$, and $\bar{u}^2 \geq 0$. The expected rent that the principal has to leave to A_1 only is equal to $\nu\Delta\theta(\nu\hat{q}_2 + (1-\nu)\bar{q})$. One can replace this in the objective function and derive the first-order conditions. This gives the following optimal quantity schedule indexed by a "P" meaning monitoring by the principal: $\underline{q}^P = q^*(2\underline{\theta})$, $\hat{q}_1^P = q^*(\underline{\theta} + \bar{\theta})$, $\hat{q}_2 = \hat{q}^D$, and $\bar{q}^P = \bar{q}^D$. The principal's payoff function is linear in c with slope -1 as the principal has to pay c for the monitoring activity. The efficiency losses are smaller than when P delegates monitoring as can be seen from the expression of the rent left to the agents; the principal's profit is thus translated upward in this new situation (see Figure 3).

Suppose that the principal monitors A_2 if and only if A_1 is $\underline{\theta}$ and reports truthfully to be so. The binding constraints are $\nu\hat{u}_2^1 + (1-\nu)\bar{u}^1 \geq 0$, $\nu\underline{u}^1 + (1-\nu)\hat{u}_1^1 \geq \nu\hat{u}_2^1 + (1-\nu)\bar{u}^1 + \nu\Delta\theta\hat{q}_2 + (1-\nu)\Delta\theta\bar{q}$, $\underline{u}^2 \geq 0$, $\hat{u}_1^2 \geq 0$, $\hat{u}_2^2 \geq \bar{u}^2 + \Delta\theta\bar{q}$, and $\bar{u}^2 \geq 0$. The expected rent that the principal must leave to the agents is now equal to $\nu\Delta\theta(\nu\hat{q}_2 + 2(1-\nu)\bar{q})$ and the optimal quantity schedule is then $\underline{q}^P = q^*(2\underline{\theta})$, $\hat{q}_1^P = q^*(\underline{\theta}+\bar{\theta})$, $\hat{q}_2^P = \hat{q}^C$, and $\bar{q}^P = \bar{q}^D$. The principal's payoff function is linear in c with slope $-\nu$ as the principal has to pay c, with probability ν, in order to monitor A_2. Again, compared to the delegated monitoring case, the principal's profit is translated upward, due to smaller efficiency losses (see Figure 3).

Suppose that the principal acquires information indirectly and relies on A_2's report to do so. Then, we face the centralized structure and according to Proposition **??** the payoff function of the principal is constant when c varies and is worth V^D.

Clearly, with direct monitoring the principal can now distinguish between mixed coalitions. Suppose that the principal wants to monitor A_2 if and only if A_1 reports truthfully $\bar{\theta}$. The binding constraints are $\nu\hat{u}_2^1 + (1-\nu)\bar{u}^1 \geq 0$, $\nu\underline{u}^1 + (1-\nu)\hat{u}_1^1 \geq \nu\hat{u}_2^1 + (1-\nu)\bar{u}^1 + \nu\Delta\theta\hat{q}_2 + (1-\nu)\Delta\theta\bar{q}$, $\underline{u}^2 \geq \hat{u}_1^2 + \Delta\theta\hat{q}_1$, $\hat{u}_1^2 \geq 0$, $\hat{u}_2^2 \geq 0$, and $\bar{u}^2 \geq 0$. The expected rent that the principal must leave to the agents is equal to $\nu\Delta\theta(\nu\hat{q}_2 + \nu\hat{q}_1 + (1-\nu)\bar{q})$ and the optimal quantity schedule is then $\underline{q}^P = q^*(2\underline{\theta})$, $\hat{q}_1^P = \hat{q}_2^P = \hat{q}^D$ and $\bar{q}^P = \bar{q}^C$. The principal's payoff function is linear in c with slope $-(1-\nu)$. One can verify that when $c \leq \nu\Delta\theta\hat{q}^D$, the principal can obtain a higher payoff by acquiring directly information whatever is A_1's type rather than by acquiring directly information when A_1 is $\bar{\theta}$ only.[22] Moreover one can also verify that when $c \geq \nu\Delta\theta\bar{q}^C$, the principal can obtain a higher

[22]Consider the optimal contract when the principal decides to monitor A_2 when A_1 is $\bar{\theta}$. When $c \leq \nu\Delta\theta\hat{q}^D$, if the principal always monitors, he can implement the same quantity schedule, give the same rent to A_1 and put A_2 on his reservation value by setting $\underline{u}^2 = 0$. Doing so, he obtains a higher payoff.

payoff by acquiring indirectly information through contracting whatever is A_1's type. As $\hat{q}^D \geq \bar{q}^C$, we proved that the strategy *"monitor if and only if A_1 is $\bar{\theta}$"* is never optimal.

The same kind of study as in the basic case permits to identify the threshold values. The result $c_2^* < c_2^p$ comes from the fact that the payoff in case delegation is chosen is the same in the two situations (monitoring by A_1 or by P), while the payoff in case of partial consolidation is higher in case of monitoring by the principal. To obtain that $c_1^* = c_1^p$, remark that in case of monitoring by the principal, the consolidation and partial consolidation schedules differ only with respect to the quantity produced by a $(\bar{\theta}, \bar{\theta})$ coalition and that this quantity is the same as the corresponding quantity in case of monitoring by agent A_1.

• **Proof of Proposition 3:** As long as the principal does not want to use monitoring himself (i.e. $c \geq c_2^p$), he can guarantee himself the delegation payoff even if he had no control on the monitoring activity. To do so, let us consider the following (symmetric) transfers: $\underline{t} = \underline{\theta} q^*(2\underline{\theta}) + \Delta\theta \hat{q}^D$, $\hat{t}_1 = \underline{\theta}\hat{q}^D + \Delta\theta\bar{q}^D$, $\hat{t}_2 = \bar{\theta}\hat{q}^D$, and $\bar{t} = \bar{\theta}\bar{q}^D$ where $\hat{t}_1$ (resp. $\hat{t}_2$) is offered to the efficient (resp. inefficient) agent in a mixed-coalition. It is easy to check that those transfers and the outputs $(q^*(2\underline{\theta}), \hat{q}^D, \bar{q}^D)$ yield rents $\underline{U}^D$ and $\bar{U}^D = 0$ to both agents. Moreover, the mechanism is dominant strategy incentive compatible, thus A_1 does not have any incentive to monitor A_2.

Let us consider the more interesting case where A_1 with type θ_1 has to monitor. We denote by $t^i(\hat{\theta}_1, \hat{\theta}_2, \tilde{\theta}_2)$ the transfer received by $A_i (i = 1, 2)$ when A_1 reports the whole vector of types $(\hat{\theta}_1, \hat{\theta}_2)$ and A_2 reports $\tilde{\theta}_2$. The same notations follows for outputs. When $\hat{\theta}_2 = \tilde{\theta}_2$ we use the usual (equilibrium) definitions of transfers and outputs. Note first that given that he monitors agent A_1 prefers to reports the truth on A_2 when $t^1(\hat{\theta}_1, \hat{\theta}_2, \tilde{\theta}_2) = -\infty$ for $\hat{\theta}_2 \neq \tilde{\theta}_2$. Similarly, with $t^2(\hat{\theta}_1, \hat{\theta}_2, \tilde{\theta}_2) = -\infty$ for $\hat{\theta}_2 \neq \tilde{\theta}_2$, A_2 reports the truth on his type at no cost for the principal. A_2's incentive constraints are trivially satisfied and his participation constraints only remain. Everything happens thus as if the principal had direct information on θ_2. Moreover, if A_1 does not monitor, either always saying always $\tilde{\theta}_2 = \underline{\theta}$ or $\tilde{\theta}_2 = \bar{\theta}$ exposes A_1 to the risk that A_2 has a different type. Given that infinite punishments follow in that case, the moral hazard constraint is trivially satisfied and remain only the participation constraints $\nu(\underline{t}^1 - \underline{\theta}\underline{q}) + (1 - \nu)(\hat{t}_1^1 - \underline{\theta}\hat{q}_1) - c \geq 0$ for a $\underline{\theta}$-agent A_1 and $\nu(\hat{t}_2^1 - \underline{\theta}\hat{q}_2) + (1 - \nu)(\bar{t}^1 - \bar{\theta}\bar{q}) - c \geq 0$ for a $\bar{\theta}$-agent A_1. Everything happens as if monitoring was verifiable by the principal.

• **Proof of Proposition 4:** Remark first that as $U_r(\theta_2)$ is a reservation utility obtained by A_2 by playing optimally in a certain game, we must have $U_r(\bar{\theta}) \leq U_r(\underline{\theta})$. Suppose that the principal wants to induce monitoring whatever A_1's type. In that case, on the equilibrium path the coalition will still reason under complete information and manipulating the reservation utility of A_2 will just change the participation constraints by making them more stringent. To see this point, let us denote by $C_1(U_r)(\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q})$ the implementation cost corresponding to the quantity schedule $(\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q})$. This cost minimizes the expected rent of the coalition of agents, subject to (at least) the following constraints:

$$\nu(\hat{u}_2 - U_r(\underline{\theta})) + (1-\nu)(\bar{u} - U_r(\bar{\theta})) - c \geq 0 \tag{37}$$

which says that a $\bar{\theta}$-A_1 must expect a positive rent knowing that he will have to pay c for the monitoring activity and let $U_r(\theta_2)$ to A_2, and the coalition incentive constraints (17) to (20), which are written under complete information and are exactly the same as when $U_r \equiv 0$. As (37) is more stringent than (3) as soon as $U_r(\underline{\theta}) > 0$ or $U_r(\bar{\theta}) > 0$, we can already deduce that $C_1(U_r) \geq C_1(0) = C_1$ when $c \leq c_1^*$ because in that case constraints (17) to (20) and (3) are the only relevant constraints when computing $C_1(0)$. If P wants to induce monitoring and $c \leq c_1^*$ then he can optimally set $U_r \equiv 0$.

Let us turn to the situation in which P wants to induce partial monitoring (monitoring if and only if A_1 is efficient) and reason over $C_2(U_r)(\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q})$ the implementation cost corresponding to the quantity schedule $(\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q})$. There will be different cases depending on whether $U_r(\underline{\theta}) - U_r(\bar{\theta}) > \Delta\theta\bar{q}$ or not.

Suppose $U_r(\underline{\theta}) - U_r(\bar{\theta}) \leq \Delta\theta\bar{q}$, the binding constraints in the program of agent A_1 when proposing a sub-contract to A_2 are the same as when $U_r \equiv 0$: virtual cost parameters are thus the same. If $\theta_1 = \underline{\theta}$, coalition incentive constraints are written under complete information (see constraints (17) and (18)) and if $\theta_1 = \bar{\theta}$, the reservation utility profile of agent A_2 ensures that there are no countervailing incentive effects and coalition incentive constraints take into account the standard virtual cost (see constraints (25) and (26)). From (18)) and (25) we can obtain that $\hat{u}_1 = \hat{u}_2 = \hat{u}$ and from (17) that the rent schedule must satisfy

$$\underline{u} \geq \hat{u} + \Delta\theta \max\{\hat{q}_1; \hat{q}_2\}. \tag{38}$$

We must also consider the participation constraints coming from this subcontracting game: in particular, because an inefficient A_1 must expect a positive rent, we must have:

$$\nu(\hat{u} - \Delta\theta\bar{q}) + (1-\nu)(\bar{u} - U_r(\bar{\theta})) \geq 0, \tag{39}$$

instead of (28) with the former being more stringent. When $U_r \equiv 0$ and $c_1^* \leq c \leq c_2^*$, C_2 was computed by taking into account these coalition incentive and participation constraints, plus a moral hazard constraint corresponding to the fact that a $\underline{\theta}$-A_1 must not prefer to save on the monitoring cost and rely on indirect learning. After such a deviation, A_1 could in particular lie and set $\Phi(\underline{\theta}, \underline{\theta}) = (\underline{\theta}, \underline{\theta})$ and $\Phi(\underline{\theta}, \bar{\theta}) = (\bar{\theta}, \bar{\theta})$. This gives the following constraint:

$$c \leq \nu(U_r(\bar{\theta}) - U_r(\underline{\theta})) + \nu\Delta\theta\bar{q} + (1-\nu)(\hat{u} - \bar{u} - \Delta\theta\bar{q}), \tag{40}$$

which corresponds to the only relevant moral hazard constraint when $U_r \equiv 0$. Combining (38), (39) and (40) we can deduce a lower bound on the expected rent left to the agents and thus on the implementation cost $C_2(U_r)$:

$$C_2(U_r)((\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q}) \geq \nu c + \nu\Delta\theta(\nu \max\{\hat{q}_1; \hat{q}_2\} + 2(1-\nu)\bar{q}) + \nu^2 U_r(\underline{\theta}) + (1-\nu-\nu^2)U_r(\bar{\theta}),$$

and as $U_r(\underline{\theta}) \geq U_r(\bar{\theta})$ there is no way to fix U_r in such a way that the implementation cost is lower than when $U_r \equiv 0$. This analysis holds whenever $c_1^* \leq c \leq c_2^*$.

Suppose now that $U_r(\underline{\theta}) - U_r(\bar{\theta}) > \Delta\theta\bar{q}$. In that case, virtual cost parameters may have changed. However, we can still write a lower bound on the implementation cost. First remark that a $\bar{\theta}$-agent A_1 must expect a positive rent :

$$\nu\hat{u}_2 + (1-\nu)\bar{u} \geq \nu U_r(\underline{\theta}) + (1-\nu)U_r(\bar{\theta}) > \nu\Delta\theta\bar{q}.$$

Next, a $\underline{\theta}$-A_1 must prefer to invest c and tell the truth rather than propose an incentive contract to A_2 and lie. In particular, we must have :

$$\nu\underline{u} + (1-\nu)\hat{u}_1 - c - (\nu U_r(\underline{\theta}) + (1-\nu)U_r(\bar{\theta})) \geq$$

$$\nu\Delta\theta\hat{q}_2 + (1-\nu)\Delta\theta\bar{q} + \nu\hat{u}_2 + (1-\nu)\bar{u} - (\nu U_r(\underline{\theta}) + (1-\nu)U_r(\bar{\theta})),$$

where the right hand side is what A_1 obtains if he sets $\Phi(\underline{\theta}, \underline{\theta}) = (\bar{\theta}, \underline{\theta})$, $\Phi(\underline{\theta}, \bar{\theta}) = (\bar{\theta}, \bar{\theta})$ and rely on arm's length contracting, and

$$\nu\underline{u} + (1-\nu)\hat{u}_1 - c - (\nu U_r(\underline{\theta}) + (1-\nu)U_r(\bar{\theta})) \geq$$

$$\nu\Delta\theta\hat{q}_1 + (1-\nu)\Delta\theta\bar{q} + \nu\hat{u}_1 + (1-\nu)\bar{u} - (\nu U_r(\underline{\theta}) + (1-\nu)U_r(\bar{\theta})),$$

where the right hand side is what A_1 obtains if he announces $(\underline{\theta}, \bar{\theta})$ instead of $(\underline{\theta}, \underline{\theta})$ and $(\bar{\theta}, \bar{\theta})$ instead of $(\underline{\theta}, \bar{\theta})$. Now, using the fact that due to the incentive constraints $\hat{u}_1 \geq \hat{u}_2$, we can write:

$$\nu\underline{u} + (1-\nu)\hat{u}_1 \geq c + \nu U_r(\underline{\theta}) + (1-\nu)U_r(\bar{\theta}) + \nu\Delta\theta \max\{\hat{q}_1, \hat{q}_2\} + (1-\nu)\bar{q},$$

so that the implementation cost $C_2(U_r)$ (which is the overall rent that P must let to the agents) verifies

$$C_2(U_r)((\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q}) \geq \nu c + \nu\Delta\theta(\max\{\hat{q}_1, \hat{q}_2\} + 2(1-\nu)\bar{q}).$$

Choosing $U_r \not\equiv 0$ is dominated.

Let us turn to the situation where P wants to induce arm's length contracting. We denote $C_3(U_r)(\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q})$ the implementation cost corresponding to the quantity schedule $(\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q})$. Suppose first that $U_r(\underline{\theta}) - U_r(\bar{\theta}) \leq \Delta\theta\bar{q}$. In that case, as argued before, the virtual costs that must be taken into account in the coalition incentive constraints are the same as if $U_r \equiv 0$ because there are no countervailing incentive effects. The initial contract must satisfy (23) to (26) and an inefficient agent A_1 must expect a positive rent: $\nu(\hat{u}_2 - \Delta\theta\bar{q}) + (1-\nu)(\bar{u} - U_r(\bar{\theta})) \geq 0$ (which is relaxed when one decreases $U_r(\bar{\theta})$). When $U_r \equiv 0$, we know that these are the only relevant constraints to determine C_3, as long as $c \geq c_2^*$. Thus we can conclude that $U_r \not\equiv 0$ cannot reduce the implementation cost and does not allow the principal to get a higher profit than V^D (the profit in the delegation case).

Suppose now that $\Delta\theta\hat{q}_1 \geq U_r(\underline{\theta}) - U_r(\bar{\theta}) > \Delta\theta\bar{q}$. In that case, there are no countervailing incentives when A_1 is efficient: constraints (23) and (24) must hold. There are some countervailing incentives when A_1 is inefficient, but we can nevertheless write that such an agent must expect a positive rent, i.e. $\nu\hat{u}_2 + (1-\nu)\bar{u} \geq \nu U_r(\underline{\theta}) + (1-\nu)U_r(\bar{\theta}) > \Delta\theta\bar{q}$. From (23) we know that $\underline{u} \geq \hat{u}_2 + \Delta\theta\hat{q}_2$, and from (24) that $\hat{u}_1 \geq \bar{u} + \Delta\theta\bar{q} + \frac{\nu}{1-\nu}\Delta\theta(\hat{q}_1 - \bar{q})$. Hence we can write a lower bound for the implementation cost:

$$C_3(U_r)(\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q}) \geq \nu\Delta\theta(\nu\hat{q}_1 + \nu\hat{q}_2 + 2(1-\nu)\bar{q}).$$

Hence $C_3(U_r) \geq C_3(0)$.

Finally, suppose that $U_r(\underline{\theta}) - U_r(\bar{\theta}) > \Delta\theta\hat{q}_1$. In that case there are countervailing incentives whatever A_1's type. From the fact that an inefficient A_1 must expect a positive rent we know that $\nu\hat{u}_2 + (1-\nu)\bar{u} \geq \nu U_r(\underline{\theta}) + (1-\nu)U_r(\bar{\theta}) > \Delta\theta\hat{q}_1$. Now consider an efficient A_1. If he sets $\Phi(\underline{\theta}, \underline{\theta}) = (\bar{\theta}, \underline{\theta})$, $\Phi(\underline{\theta}, \bar{\theta}) = (\bar{\theta}, \bar{\theta})$, he will just have to leave the rent schedule U_r to A_2 and thus obtain the rent $\nu(\hat{u}_2 + \Delta\theta\hat{q}_2 - U_r(\underline{\theta})) + (1-\nu)(\bar{u} + \Delta\theta\bar{q} - U_r(\bar{\theta}))$. In order to prevent that deviation, P must guarantee

$$\nu\underline{u} + (1-\nu)\hat{u}_1 - \nu U_r(\underline{\theta}) - (1-\nu)U_r(\bar{\theta}) \geq \nu(\hat{u}_2 + \Delta\theta\hat{q}_2 - U_r(\underline{\theta})) + (1-\nu)(\bar{u} + \Delta\theta\bar{q} - U_r(\bar{\theta})),$$

combining this inequality with the rent he must left to an inefficient A_1 gives a lower bound on the implementation cost:

$$C_3(U_r)(\underline{q}, \hat{q}_1, \hat{q}_2, \bar{q}) \geq \nu\Delta\theta(\nu\hat{q}_1 + \nu\hat{q}_2 + (1-\nu)(\hat{q}_1 + \bar{q})).$$

Because $\hat{q}_1 \geq \bar{q}$, we obtain the desired result.

Gathering all the cases studied yields Proposition 4.