The Organization of Supplier Networks: Effects of Delegation and Intermediation

Dilip Mookherjee
Department of Economics, Boston University; email:dilipm@bu.edu

Masatoshi Tsumagari
Department of Economics, Keio University; email:tsuma@econ.keio.ac.jp

Forthcoming, *Econometrica*

Final version: August 9 2003

Abstract

In a one principal two-agent model with adverse selection and collusion among agents, we show that delegating to one agent the right to subcontract with the other agent always earns lower profit for the principal compared with centralized contracting. Delegation to an intermediary is also not in the principal's interest if the agents supply substitutes. It can be beneficial if the agents produce complements and the intermediary is well informed.

Earlier versions of this paper have previously been circulated under different titles: "The Organization of Supplier Networks: Effects of Mergers and Delegation", and "Delegated Management and Contract Intermediation". We have benefited from discussions with Jean-Jacques Laffont and Gorkem Celik, detailed suggestions of three anonymous referees, and comments of seminar participants at Boston University, Cambridge University, University of Chicago, Keio University, Kyoto University, University of North Carolina, Stockholm School of Economics and the April 2001 Decentralization Conference at Northwestern University.

1 Introduction

We study the value of alternative ways of organizing production or supply relationships for a Principal (called P). There are two agents A_1 and A_2 supplying distinct inputs that combine to form an output valued by P, according to a quasiconcave CRS production function. The agents are privately informed about their own supply costs, which are distributed according to independent, continuous one dimensional random variables. A_1 and A_2 can collude with one another, whereby A_1 offers an enforceable side contract to A_2 which is unobserved by P. The side contract coordinates cost reports, reallocates production assignments and payments between the agents. In the case where the agents supply perfect substitutes, this reduces to the design of a procurement auction with collusion and resale among bidders.

Anticipating collusion among the agents, P may consider contracting with A_1 alone, delegating to him the authority to contract with A_2. Or P may consider delegating this authority to an intermediary M not directly involved in production, but possessing superior information about agents' costs. That decentralization may be an optimal response to collusion is a theme which has been explored recently by a number of authors (Baliga and Sjostrom (1998), Faure-Grimaud, Laffont and Martimort (2003)). We address this question in the context described above by comparing the following three contracting arrangements. In *Centralization with Collusion (CC)*, P personally contracts with both suppliers, who can collude with one another. *Delegation to a Supplier (DS)* is a setting where P contracts with A_1 alone, and delegates to A_1 the authority to contract with A_2. And in *Delegation to Middleman (DM)*, P delegates to an intermediary M authority to contract with A_1 and A_2. In a procurement setting, these correspond to P contracting with a 'prime' supplier who subcontracts with other suppliers in DS, with a middleman with specialized information in DM, and directly with all suppliers in CC. In the context of a firm owned by P, CC corresponds to the firm managed by its owner, DS by one of the workers, and DM by a professional manager possessing specialized information about production costs but otherwise uninvolved in production.

Our first main result is that the presence of collusion does *not* justify delegation to a supplier: DS is strictly dominated by CC in general. By contracting with both suppliers, the Principal can affect their outside options when they bargain over a side contract. This reduces the severity of the problem of *double marginalization of rents* inherent in vertical contracting relationships (Tirole (1988, Ch. 4)). The problem arises from the monopsony power exercised by A_1 over A_2 in the subcontracting relationship in DS, causing a markup in A_1's subcontracting cost over A_2's production cost and consequent underprocurement. Moreover, A_1 is privately informed about his subcontract cost while contracting with P, causing a double markup in P's effective cost of procuring from A_2. The inflation of procurement costs owing to the cascading of information rents across vertical layers causes P to procure too little of the final good, besides distorting allocation of inputs across the two agents. Under centralization P can offer a contract to both agents which provides a higher outside option to A_2 while bargaining with A_1 over the side contract. This reduces A_1's monopsony power in the subcontracting relationship, easing the extent of double marginalization of rents.

Delegation to a middleman M who is informed about supplier costs has the potential advantage of limiting rents earned by A_1 and A_2, and avoiding the input allocation distortions that DS entails. This enhances efficiency in the relationship between the middle and bottom tiers of the organization. On the other hand the agency problem between the top and middle tier is accentuated, since P now deals with a single consolidated agent M who controls the supply of both inputs and is privately informed about the cost of each. Such consolidation limits 'competition' but facilitates 'coordination' among suppliers. We show that delegation to M is never worthwhile for P when the two inputs are substitutes, owing to the dominance of the competition-suppressing effect. However delegated intermediation is profitable for P if the two inputs are complements, and some distributional conditions (that limit the 'loss of control' attending delegation) happen to be satisfied.

Moreover, DM is not dominated by a larger centralized arrangement where P retains the right to contract personally with M and the two suppliers, if M colludes with the suppliers against P and is perfectly informed about agents' supply costs. If M is not subject to asymmetric information *vis-a-vis* the suppliers, there is no problem of double marginalization of rents in this delegation arrangement that P can rectify by contracting directly with all the agents.

The potential benefits of intermediation do rely on the specialized information possessed by the intermediary about supplier costs. If M has the same information as P, we show that DM is worse than either of the other two alternatives DS or CC. In that case P contracts with an agent (M) that is less informed than A_1, implying that the coalition that P deals with is subject to higher cost in DM compared with DS. This exemplifies a general principle that P prefers to deal with a more 'internally efficient' coalition, defined by the shadow cost to the coalition of delivering the final output to the Principal. However greater internal efficiency need not correspond to lower informational distortions within the coalition. For instance, in CC the coalition is subject to asymmetric information, whereas it is not in DM. The comparison between them can go either way, depending on whether the inputs are complements or substitutes. On the other hand, if P has to choose between delegating authority between the middleman, and either of the suppliers, then it is better to delegate to the better informed agent. In general, the shadow cost of the coalition depends on a number of other parameters besides information structure, such as allocation of bargaining power and complementarity or substitutability of the inputs.

In summary, we find that delegation is justified only to intermediaries or middlemen, in circumstances involving technological complementarities and where the intermediary is sufficiently well informed. Our model differs from much of the existing literature on collusion in organizations by allowing a general production function, and supplier cost shocks represented by continuum rather than two-point random variables. The relation of our results to those of existing literature on delegation and collusion (such as Baron and Besanko (1992, 1999), Baliga and Sjostrom (1998), Celik (2002), Faure-Grimaud *et al* (2003) and Laffont and Martimort (1998)) is discussed in detail in Section 6.

Section 2 presents the model. Section 3 describes the CC structure. Section 4 then describes DS and compares it to CC. Section 5 is devoted to results concerning DM. Section 6 discusses relation to existing literature, while Section 7 concludes by discussing extensions, shortcomings and future directions. All proofs are gathered in the Appendix.

2 Model

A principal (P) values an output jointly produced by two suppliers A_1 and A_2 according to a production function $q = Q(q_1, q_2)$, where q_i denotes the product delivered by A_i. Q is homogeneous of degree one, quasi-concave and strictly increasing in each argument. P's return from final product q is $V(q)$, where V is a twice continuously differentiable and strictly concave function satisfying Inada conditions. We shall assume that both agents are *essential* in the production process: $q_i = 0$ implies $Q = 0$, i.e., the production isoquants do not touch the axes. The interiority of production assignments simplifies the analysis, but is not essential to the results. The results extend to the case of perfect substitutability (as in a procurement auction), where

$q = q_1 + q_2$.[1]

A_i's cost of supplying q_i is $\theta_i q_i$. θ_i is observed by A_i, but not by P or $A_j, j \neq i$. It is common knowledge that P and A_j share common beliefs over the realization of θ_i, represented by distribution function $F_i(\theta_i)$ and density $f_i(\theta_i)$ on the interval $[\underline{\theta}_i, \bar{\theta}_i]$. θ_1 and θ_2 are independently distributed. The density function f_i is assumed to be continuous, bounded and everywhere positive on its support, with a monotone hazard rate: $\frac{F_i(\theta_i)}{f_i(\theta_i)}$ is non-decreasing in θ_i.[2]

M is a fourth-party intermediary who plays no role in production, but may be better informed than P about the realization of supplier costs. We shall consider two polar cases, one where M has perfect information regarding θ_1, θ_2, another where she has exactly the same information as P. In Section 7 we discuss how our results are likely to extend to the case where M is imperfectly informed about supplier cost.

P and all the agents are risk neutral. P's objective is to maximize the expected value of V, less expected payments to the agents. M's objective is to maximize expected transfers received, and supplier A_i seeks to maximize expected transfers received, less expected production costs. M and the two suppliers have outside options equal to 0.

Before proceeding to consider the different contracting networks it is useful to establish a simple result concerning ranking of different agency problems according to the distribution of unit costs. Since the suppliers and intermediary collude with one another, the Principal's problem in any given setting can be viewed as contracting with the supplier-intermediary coalition which collectively delivers the final output at a particular (shadow) unit cost, determined by the outcome of side contracting within the coalition. The following result states that different coalitional structures can be ranked by their respective shadow costs.

Proposition 1 *Consider the problem of P contracting with a single agent who delivers final output q at a unit cost of either c_1 or c_2, where: (a) $c_i, i = 1, 2$ is distributed over some nondegenerate interval with a continuous positive density function g_i (with associated distribution function G_i), and (b) the distribution of c_1 first order stochastically dominates that of c_2. Then*

[1] This follows from the fact that such a production function can be obtained as the limit of a sequence of CES production functions with elasticity of substitution converging to ∞.

[2] The case where the density is zero at either end-point can also be accommodated, with the understanding that the hazard rate is set at any such endpoint to ensure that the hazard rate is continuous over the support.

P's expected profit when the agent's cost is c_2 is at least as high as when it is c_1 (and is strictly higher if $G_2(c) > G_1(c)$ for (almost) all c in the support of c_1).

3 Centralization

In the absence of collusion between the suppliers, analysis of the centralized regime is standard (analogous to the analysis of auctions by Myerson (1981)). By the Revelation Principle, P can confine attention to the class of revelation mechanisms, in which both suppliers agree to participate and report their costs truthfully. Such a mechanism is depicted as $q_i(\theta), X_i(\theta), i = 1, 2$, where $\theta \equiv (\theta_1, \theta_2)$, q_i denotes the production assignment and X_i the payment to A_i. Using the approach of Myerson, the optimal revelation mechanism can be shown to be found by first selecting production assignments $q_i^C(\theta), i = 1, 2$ to maximize pointwise

$$V(Q(q_1, q_2)) - \sum_{i=1}^{2} h_i(\theta_i)q_i \tag{1}$$

the difference between principal's revenue and *virtual* costs. The virtual unit cost of q_i is $h_i(\theta_i) \equiv \theta_i + \frac{F_i(\theta_i)}{f_i(\theta_i)}$, the sum of production cost (θ_i) and informational rent $\frac{F_i(\theta_i)}{f_i(\theta_i)}$ which accrues to A_i. The incentive to limit such rents causes P to procure less than in a first-best perfect information setting from each supplier. Letting $\pi(\theta) \equiv \max_{q_1,q_2}[V(Q(q_1, q_2)) - \sum_{i=1}^{2} \theta_i q_i]$ denote the first-best profit function, the Revelation Principle implies that the expected profit of P in this setting $\Pi^C \equiv E_\theta \pi(h_1(\theta_1), h_2(\theta_2))$ represents an upper bound to profits in any mechanism that does not involve M.

We now introduce collusion between the suppliers. We adopt the standard formulation in the literature following Tirole (1986) that collusion consists of an enforceable side-contract between the agents, where they can coordinate cost reports to P, reallocate production assignments and payments received from P. The side contract cannot be observed by P, nor can P verify what payments or production reallocations occurred between the two suppliers. P can however verify aggregate output q delivered by the coalition, so that production reallocations can consist only of changing the relative contributions of the two suppliers to the final output. In the context of an auction where $Q = q_1 + q_2$, the side contract amounts to a resale contract.

We make a specific assumption concerning allocation of bargaining power within the coalition: A_1 makes a take-it-or-leave-it offer of a side contract to

A_2. If A_2 refuses, they play the contract offered by P (which we shall refer to as the *grand contract (GC)*) noncooperatively. Hence the grand contract serves as the outside option for the two suppliers when they negotiate a side contract. The side contract is subject to asymmetric information within the coalition, since A_1 does not know A_2's cost. Accordingly, the side contract involves exchange of messages between the suppliers, which leads to a joint decision concerning the reports they send to P, and a reallocation of the payments and production assignments mandated by the grand contract corresponding to their joint report.

Formally, the nature of contracts and timing of moves in the CC game is as follows:

(CC1) P offers grand contracts $q_i(m_1, m_2), X_i(m_1, m_2), i = 1, 2$ to both suppliers, where m_i is a message submitted by A_i, consisting of a decision whether or not to accept the grand contract, and a cost report in case of acceptance. If either of the two agents does not accept, then payments and production assignments for both are identically zero.

(CC2) A_1 offers a side-contract $m(e_1, e_2), t_2(e_1, e_2), \tilde{q}_i(e_1, e_2)$ to A_2, satisfying the constraint $Q(q(m(e))) = Q(\tilde{q}(e))$ for all e, where $e \equiv (e_1, e_2)$ denotes a vector of cost reports exchanged internally within the coalition, $\tilde{q} \equiv (\tilde{q}_1, \tilde{q}_2)$ is a production assignment decided by the coalition, $m \equiv (m_1, m_2)$ is the set of cost reports submitted to P, and t_2 is a side payment from A_1 to A_2.

(CC3) A_2 decides whether to participate in the side contract. If not, agents play the mechanism designed by P noncooperatively. Otherwise the game continues.

(CC4) A_i observes realization of θ_i.

(CC5) A_1 and A_2 reconsider whether to participate in the side-contract. If either of them decides not to, they play the mechanism designed by P noncooperatively. Otherwise they play according to the side contract and exchange cost messages e_1 and e_2.

Centralization in the absence of collusion among suppliers (denoted by C) corresponds to the case where stages CC2 and CC3 do not appear, while stage CC5 is modified to independent reporting of costs and participation decisions by the two suppliers to P. In that case P can verify the input supplied by each agent.

Note the following features of this formulation, many of which are shared by other organizational variants to be studied subsequently:

(a) The side contract is offered *ex ante* by A_1, before either supplier has learnt his cost. One interpretation is that A_1 and A_2 are two firms that participate in many different procurement auctions organized by different purchasers. So the side contract represents a long-term arrangement between the suppliers specifying how they collude in a large variety of future procurement settings. The convenience of this formulation is that 'informed principal' issues do not arise in the analysis of side contracts.

(b) The suppliers can postpone responding to P's contract offer until after they have communicated with one another. This pertains not only to the bids they submit but also whether they decide to participate in the auction.[3] This seems natural: when agents collude, it would be hard for P to prevent them from communicating with one another before they respond with a participation decision to the GC. Collusion thus allows participation decisions of the agents to be postponed from the *interim* to the *ex post* stage, compounding agency problems. Note also that coalition members have an *ex ante* as well as *interim* opportunity to decide whether to participate in the side contract. This is natural given that the side contract is negotiated *ex ante* as a mutually profitable long-term agreement, in which participation is voluntary. In most situations *interim* participation implies *ex ante* participation, so the latter is redundant. Nevertheless, it simplifies our arguments in some places so we prefer to include the *ex ante* participation decision explicitly.

(c) The enforceability of the contracts offered by P depends on whether the agents collude, since collusion alters the information about inputs supplied by different agents that is verifiable. In general, verifiable information includes signals q_1^s, q_2^s of the respective contributions q_1, q_2 of the two agents (e.g., based on accounting information), and their joint output q (based on physical evidence). The grand contract specifies targets for the contributions of each agent, based on the realized signals. Collusion permits the agents to costlessly manipulate the signals of their respective contributions in a way that is consistent with their true joint output (i.e., so that $Q(q_1^s, q_2^s) = Q(q_1, q_2) = q$). One agent can thus claim to produce more than she actually did, only if the other agent correspondingly claims to have produced less. In the case of an auction, for instance, this takes the form of one firm transferring some of its output clandestinely to the other

[3] In this respect our model differs from Faure-Grimaud *et al* (2003), where participation decisions are made at the *interim* stage, and bids at the *ex post* stage.

agent in exchange for a side payment. In the absence of such collusion, the observed signal of each agent's production is the true production level: $q_i^s \equiv q_i, i = 1, 2$.[4]

(d) All the bargaining power in side contracting has been assigned to A_1 (as it will be in the DS regime). The purpose has been to ensure that the structure of the side contracting game does not vary between centralization and delegation, and this structure seems natural in the case that P delegates subcontracting authority to A_1. A more general formulation may involve side contracts designed by an arbitrator acting on behalf of the agents, who assigns a given set of welfare weights to the utilities of the two agents. Our formulation corresponds to the case where a zero welfare weight has been assigned to A_2.

4 Delegation to a Supplier

The DS game consists of the following stages:

(DS1) P offers a contract $q_1(m), q_2(m), X_1(m)$ to A_1, where m is a message sent by A_1 which specifies if A_1 agrees to participate, and if so can consist of a report concerning supply costs.[5] In the event of non-participation, $q_1 = q_2 = X_1 = 0$.

(DS2) A_1 offers a side contract $m(e_1, e_2), t_2(e_1, e_2)$, $\tilde{q}_i(e_1, e_2)$ to A_2, satisfying the constraint $Q(q(m(e))) = Q(\tilde{q}(e))$ for all e, where e_i denotes a cost message to be submitted by A_i to the other supplier, m denotes the report that A_1 will subsequently submit to P, and t_2 is a transfer

[4] It would not be equivalent for P to instead offer a single collective grand contract to the coalition, specifying aggregate payment and collective output of the coalition corresponding to their joint reports. This presumes that the two suppliers must necessarily side contract with one another, in order to assign their relative production assignments and payments within the coalition. In this case, if the suppliers fail to agree on a side contract, there will be no production or payments, and CC will reduce to the DS regime. As we shall show below, it is essential in CC that P offers a set of grand contracts the noncooperative play of which represents a fallback option for the agents should they fail to agree on the side contract. This option affects relative bargaining power of the two agents and thus (despite never being utilized in equilibrium) represents the key advantage of centralization over delegation.

[5] In this regime, the contract with A_1 can equivalently be specified in terms of a target for the joint output q rather than the detailed contributions of each agent, since there is no 'fallback' noncooperative game that the suppliers can play in the event they fail to agree on a side contract.

from A_1 to A_2.[6] In the event that A_2 does not participate in the side contract in either (DS3) or (DS5), $t_2 = 0$, A_1 must also decide not to participate in the contract offered by P.

(DS3) A_2 decides whether to participate in the side contract. If not, the game ends here. Otherwise, the game continues.

(DS4) A_i observes the realization of θ_i.

(DS5) A_1 and A_2 decide whether to participate in the side contract. If either decide not to, the game ends here. Otherwise, it continues.

(DS6) A_1, A_2 exchange cost messages e_1, e_2.

Here side contracting takes place under a setting similar to that in centralization with collusion (CC). A_1 has the power to make a take-it-or-leave-it offer to A_2, and A_1 responds to P's prime contract offer only after communicating and contracting with A_2. The only difference is that P does not separately contract with A_2. Refusal of the side-contract offered by A_1 now implies that no production will take place, so A_2's options are now restricted. P can thus trivially achieve the outcome of DS in CC by offering a null contract to A_2 in the latter. Whether CC strictly dominates DS depends on the value of offering A_2 a side contract option as well, in order to augment A_2's bargaining power in the side contracting game.

It is easily seen that DS reduces to the problem in contracting with A_1 alone for delivery of the final output. Given any output target, A_1 will allocate production assignments in order to minimize the sum of his own production cost $\theta_1 q_1$, and the rent-inclusive 'virtual' cost $h_2(\theta_2)q_2$ of procuring from A_2. If $c(\theta_1, \theta_2) \equiv \min\{\theta_1 q_1 + \theta_2 q_2 | Q(q_1, q_2) = 1\}$ denotes the minimum unit cost of delivering one unit of final output when input q_i costs θ_i, A_1's effective cost of delivering one unit of the final output in DS equals $c(\theta_1, h_2(\theta_2))$. The maximum profit that P can attain in DS is thus represented by the solution to a single agent problem who delivers the final output at unit cost $c(\theta_1, h_2(\theta_2))$, and is privately informed regarding the realization of this cost.

This can be solved using the techniques of Baron and Myerson (1982). Without loss of generality, A_1 will be asked to report the realization of his

[6]The assumption that the side contract is negotiated before cost realizations are known implies that the side contract offer by A_1 does not communicate any information to A_2. Dropping this assumption (i.e., allowing the side contract to be offered after costs are observed) would make no difference to the results, by virtue of the results of Maskin and Tirole (1990), and the fact that utilities are quasilinear.

unit cost $c(\theta_1, h_2(\theta_2))$, and the output target will be a nonincreasing function of this reported cost (henceforth denoted $q^{DS}(c)$). A_1 will be able to exploit his private information about the realization of both θ_1 and $h_2(\theta_2)$, and the fact that she can decide whether or not to participate in the prime contract *ex post*. It follows that the optimal payment to A_1 following report of unit cost c will be

$$X^{DS}(c) = cq^{DS}(c) + \int_c^{\bar{c}} q^{DS}(y)dy \tag{2}$$

where $\bar{c}$ denotes the upper endpoint of the support of $c(\theta_1, h_2(\theta_2))$.

It is well known that under these conditions, DS will result in a strictly lower profit for P compared to C, centralization in the absence of collusion (as explained in more detail in Melumad, Mookherjee and Reichelstein (1995, Theorems 1,3)). The following decomposition of X^{DS} makes this clear:

$$\begin{aligned} X^{DS} &= E_\theta[c(\theta_1, h_2(\theta_2))q^{DS}(c(\theta_1, h_2(\theta_2))) + \int_{c(\theta_1,h_2(\theta_2))}^{c(\bar{\theta}_1,h_2(\bar{\theta}_2))} q^{DS}(y)dy] \\ &= E_\theta[\theta_1 q_1^{DS} + h_2(\theta_2)q_2^{DS} + \int_{\theta_1}^{\bar{\theta}_1} q_1^{DS}(y, h_2(\theta_2))dy + \int_{h_2(\theta_2)}^{h_2(\bar{\theta}_2)} q_2^{DS}(\bar{\theta}_1, h)dh] \\ &= E_\theta[h_1(\theta_1)q_1^{DS} + h_2(\theta_2)q_2^{DS} + \int_{h_2(\theta_2)}^{h_2(\bar{\theta}_2)} q_2^{DS}(\bar{\theta}_1, h)dh] \end{aligned} \tag{3}$$

where

$$q_i^{DS} \equiv c_i(\theta_1, h_2(\theta_2))q^{DS}(c(\theta_1, h_2(\theta_2)))$$

denotes the production assignments in DS. Since each agent's assignment q_i^{DS} is nonincreasing in θ_i, these production assignments can be implemented in C. The minimum expected cost of implementing them in C will be

$$E_\theta[h_1(\theta_1)q_1^{DS} + h_2(\theta_2)q_2^{DS}] \tag{4}$$

Comparing (4) with (3) it follows that optimal assignments in DS can be implemented in C at lower cost.

The difference is represented by the last term on the right-hand side of (3). This reflects the double marginalization of rents, arising from rents earned by A_1 owing to privacy of his information *vis-a-vis* P with regard to the magnitude of the rents paid to A_2. If A_1 could be compelled to respond to the prime contract with his participation decision at the *interim* stage before communicating with A_2, these 'double' rents could be taxed away by P upfront in the prime contract. The fact that A_1 responds *ex post* prevents P from extracting these rents. Limited liability constraints (or risk aversion) of A_1 would produce the same effect, as emphasized by McAfee and McMillan (1995).

The double marginalization results in two kinds of distortions. The first and more fundamental one is reduced procurement of the final good, owing to the increased procurement cost resulting from cascading of rents across two vertical layers. This distortion applies irrespective of the nature of the production function.

There is a supplementary distortion in the allocation of inputs between the two agents, in the case of a technology which admits some substitutability. The double markup applies to inputs procured from A_2, whereas only a single markup applies to the input supplied by A_1. This results in a bias in production assignments in A_1's favor, which takes the form of insufficient 'outsourcing' by A_1 to A_2. The second-best production assignments (in the optimal centralized contract) involves the unit cost function $c(h_1(\theta_1), h_2(\theta_2))$. Accordingly relative contributions of the two suppliers $\frac{q_1}{q_2}$ which should optimally equal $\frac{c_1(h_1(\theta_1),h_2(\theta_2))}{c_2(h_1(\theta_1),h_2(\theta_2))}$ are set by A_1 equal to $\frac{c_1(\theta_1,h_2(\theta_2))}{c_2(\theta_1,h_2(\theta_2))}$ instead.

Nevertheless, the main question of interest for us pertains to comparison of DS with centralization in the presence of collusion. The problem of double marginalization of rents arises in CC as well, where A_1 makes a take-it-or-leave-it offer of a collusive side contract to A_2. The relevant question is the *relative* magnitude of the double rent problem in the two settings. As we have explained above, the only difference between DS and CC is that in the latter P has the opportunity of contracting with both suppliers, thus manipulating A_2's outside option when they bargain over a side contract.

Our first main result is that this opportunity to manipulate outside options of colluding suppliers is *always* valuable to P, so delegation to a supplier continues to be dominated by centralization even in the presence of collusion.

Proposition 2 *There exists a grand contract which if offered by P induces a (perfect Bayesian) equilibrium in the continuation game in CC which generates a higher expected profit for P than any mechanism in DS.*

The basic idea is that increasing the outside option of an agent who earns informational rents can cause increased efficiency. This is familiar from models of moral hazard with limited liability (e.g., see Mookherjee (1997) or Banerjee, Gertler and Ghatak (2002)), and adverse selection models with type dependent reservation utilities (Lewis and Sappington (1989) or Jullien (2000)). Forcing the contract designer (A_1) to offer higher rents to a supplying agent (A_2) reduces the productive distortion resulting from the designer's (A_1's) incentive to minimize informational rents of the agent (A_2). It causes A_1 to move closer to maximizing the sum of the rents of the two suppliers, rather than expand his own rent at the expense of A_2.

To explain this point, consider a designer denoted C who wishes to procure quantity q of a good valued at $V(q)$ from an agent denoted A, privately informed about her production cost θq, where θ is distributed on an interval $[\underline{\theta}, \bar{\theta}]$ with density f, distribution F, and corresponding virtual cost $h(\theta) = \theta + \frac{F(\theta)}{f(\theta)}$. Suppose the agent has an interim outside option of 0 for all types. Then the optimal procurement $q(\theta)$ maximizes $E[V(q(\theta)) - h(\theta)q(\theta)]$ subject to a monotonicity constraint on q. Ignoring the monotonicity constraint, $q(\theta)$ is chosen to pointwise maximimize $V(q) - h(\theta)q$, so the rent-inclusive virtual cost $h(\theta)$ is the shadow cost of procurement.

Now suppose A has a supplementary *ex ante* outside option of $u > 0$, while her *interim* options are unchanged. Given the magnitude of informational rents implied by incentive compatibility, C's problem is subject to the additional participation constraint $E[\frac{F(\theta)}{f(\theta)}q(\theta)] + \bar{U} \geq u$, where $\bar{U}$ denotes the interim rent of type $\bar{\theta}$ of agent A. Letting λ denote the shadow price of the *ex ante* participation constraint, and again ignoring the monotonicity constraint, the optimal procurement $q(\theta)$ now pointwise maximizes $V(q) - h(\theta)q + \lambda[F(\theta)/f(\theta)]q$. Hence the shadow cost drops from $h(\theta)$ to $h(\theta) - \lambda[F(\theta)/f(\theta)]$ whenever λ is strictly positive, i.e., when the *ex ante* participation constraint binds.

The original distortion arose from C's incentive to underprocure in order to limit the informational rents paid to A. If A's outside option is raised then there is no need to underprocure as much as previously. Given the original underprocurement, raising the amount procured (rather than paying more lumpsum) is the efficient way of awarding A higher rents. This corresponds to a lower shadow cost of procurement, as raising q by one unit relaxes the participation constraint, which is valued at the corresponding Kuhn-Tucker multiplier of the constraint (appropriately normalized). If we identify C with A_1, A with A_2, q, θ with q_2, θ_2 respectively, it is evident that A_1's shadow cost of procuring from A_2 falls from $h_2(\theta_2)$ to

$$z_{1-\lambda}(\theta_2) = \theta_2 + (1-\lambda)\frac{F_2(\theta_2)}{f_2(\theta_2)}. \tag{5}$$

This causes A_1's shadow cost of delivering the final output to drop from $c(\theta_1, h_2(\theta_2))$ to $c(\theta_1, z_{1-\lambda}(\theta_2))$, and Proposition 1 can then be applied.

More specifically, the proof of Proposition 2 is based on the following set of contracts offered by P to the two agents under centralization. These contracts mimic what the outcome of DS would be if A_1's shadow cost of procuring from A_2 were to hypothetically equal

$$z_\alpha(\theta_2) = \theta_2 + \alpha \frac{F_2(\theta_2)}{f_2(\theta_2)} \tag{6}$$

where α is a parameter lying between 0 and 1. The case of $\alpha < 1$ corresponds to a lower procurement cost than the outcome under DS (which corresponds to $\alpha = 1$). Accordingly it corresponds to higher procurement and rents for A_2, relative to the outcome of DS.

With these contracts serving as the status quo for collusion, the proof shows that it is optimal for A_1 to not offer any side-contract at all, i.e., to propose that they play P's mechanism noncooperatively. In other words, with outside options formed by the hypothetical DS mechanism corresponding to procurement cost z_α, the optimal side-contract for A_1 mimics the outcome of this mechanism exactly: the shadow price λ in the optimal side contract equals $1 - \alpha$.[7]

Note that this construction implements in CC an outcome which is one possible equilibrium of the CC game following offer of the constructed grand contract.[8] We have not been able to establish if this is the unique equilibrium of the continuation game. Note however that there cannot exist an alternative equilibrium which *ex ante* Pareto dominates (from the agents' standpoint) the selected equilibrium — for if there were, it would have been in A_1's interest to propose a side contract which plays the alternate equilibrium. Moreover, alternative (but more complicated) grand contracts can be constructed in CC with a unique noncooperative equilibrium in undominated strategies, the collusive outcome of which strictly dominates DS.[9]

---

[7]The structure of the argument is as follows. The side contracting problem is complicated by type dependent interim participation constraints, in addition to a higher *ex ante* participation constraint for A_2. We deal with this by first relaxing the interim participation constraints to a zero outside option for each type, and preserve only the *ex ante* constraint. Then the shadow price of this constraint λ has to be at least $1 - \alpha$, otherwise the amount procured from A_2 is too little to attain the required *ex ante* rent, necessitating an additional lumpsum transfer. It is cheaper to procure more and reduce the lumpsum payment. On the other hand, if λ exceeds $1 - \alpha$ then A_2 is given more rents than is necessary. So in the optimal side-contract λ must exactly equal $1 - \alpha$. This means that the side contract replicates the contract offered by P without any collusive manipulation. Hence ignoring the type-dependent interim participation constraints we obtain a solution which automatically satisfies those constraints, since the interim outside options are given by the interim rents from noncooperative play of P's offered mechanism.

[8]The approach of Baliga and Sjostrom (1998) is similar in this respect.

[9]The proof of this is available on request.

The generality of Proposition 2 is notable: it applies irrespective of the nature of the technology or information structure. The suboptimality of delegation to a supplier is in contrast to the results of Baliga and Sjostrom (1998) and Faure-Grimaud *et al* (2003), and in line with those of Celik (2002). One question raised by a referee concerns the role of the assumption that P cannot monitor the inputs of the two agents separately. In the case of perfect complementarity between the two agents' tasks, note that such monitoring has no value, and yet Proposition 2 applies in that case. Intuitively, monitoring inputs can possibly alleviate the distortion in input assignments that arises when there is some substitutability between them. But it cannot alleviate the distortion in the quantity of the final good procured by P, which arises irrespective of the nature of the production function. We suspect therefore that the result extends also to the case where inputs are monitored. But we have been unable so far to confirm this in general, owing to the complexity of the multidimensional incentive problem in DS with input monitoring. However we have been able to verify that the result is valid with input monitoring in the case where the cost shocks of the two agents are identically and exponentially distributed with a lower bound of 0.

5 Delegation to a Middleman or Manager

Now suppose P contracts only with an intermediary or manager M, and delegates to M the authority to contract with the two suppliers. Given P's inability to monitor inputs, P contracts with M only over the final output. Given such a contract, M offers contracts to the two suppliers which determine their production assignments and payments as a function of cost reports they submit to M. Formally, the DM game consists of the following stages.

(DM1) P offers a contract $q_1(m), q_2(m), X_M(m)$ to M, where m is a message to be sent by M which includes a decision whether to participate, and in that event a cost report. In the event of non-participation, $q_1 = q_2 = X_M = 0$.

(DM2) M offers side contracts $t_i(e_M, e_1, e_2), \tilde{q}_i(e_M, e_1, e_2)$ to A_1, A_2, along with a planned participation decision and cost report $m(e_M, e_1, e_2)$ to be submitted eventually to P, where e_i denotes a cost message to be submitted by A_i, and e_M by M. The side-contract must satisfy the constraint that $Q(q(m(e))) = Q(\tilde{q}(e))$ for all $e \equiv (e_M, e_1, e_2)$. In the event that one of the suppliers declines to participate, $\tilde{q}_1 = \tilde{q}_2 = t_1 =$

$t_2 = 0$, M must also decide not to participate in the contract offered by P.

(DM3) A_1 and A_2 decide whether to participate in the side contract. If either decides not to, the game ends here. Otherwise it continues.

(DM4) A_i observes the realization of θ_i and $s_i, i = 1, 2$; M observes the realization of $s_i, i = 1, 2$ where s_i is a signal of θ_i, which is either completely informative ($s_i = \theta_i, i = 1, 2$) or completely uninformative (in which case M has the same beliefs as P over θ_i).

(DM5) M and $A_i, i = 1, 2$ reconsider whether to participate in the side contract. If any of them decides not to, the game ends here. Otherwise the side contract is played: A_i submits report e_i and M submits e_M.

In delegating to M, P does not contract with the suppliers at all. Accordingly, M has the power to make a take-it-or-leave-it offer of subcontracts to the suppliers, who have no opportunity to produce anything if they happen to turn down the subcontracts offered by M. This is analagous to DS where A_2 does not have the opportunity to produce if he turns down the subcontract offered by A_1.

The version of centralization with collusion that corresponds to DM is one where P personally contracts with all three agents M, A_1 and A_2, with M subsequently proposing a side contract to the two suppliers. In this case the suppliers have the option of refusing the side contract offered by M and playing the mechanism proposed by P noncooperatively. This organizational variant we refer to as *Centralization with M and with Collusion (CMC)*. In effect, M is treated as a supervisor who is asked to report what he knows about the realization of the supplier costs to P, who can use this information in designing supply contracts for A_1 and A_2. But M can collude with the suppliers with regard to the reports she submits, limiting what P can achieve from extending the centralized mode to incorporate M. This game consists of the following stages.

(CMC1) P offers a contract to all three agents:

$$(X_M(m), X_1(m), X_2(m), q_1(m), q_2(m)),$$

where $m \equiv (m_M, m_1, m_2)$ denote their respective cost messages. In the event that any agent $A_i, i = 1, 2$ or M decides not to participate, $X_M = X_i = q_i = 0, i = 1, 2$.

(CMC2) M offers a side contract $m(e_M, e_1, e_2), t_i(e_M, e_1, e_2), \tilde{q}_i(e_M, e_1, e_2), i = 1, 2$ to A_1 and A_2, where e_i denotes a cost message to be submitted by A_i, and e_M a message to be submitted by M. The side-contract must satisfy the constraint that $Q(q(m(e))) = Q(\tilde{q}(e))$ for all $e \equiv (e_M, e_1, e_2)$.

(CMC3) $A_i, i = 1, 2$ decide whether to participate in the side-contract offered by M. If either of them decides not to, they play P's mechanism noncooperatively. Otherwise the game continues.

(CMC4) M and A_i observe the realization of their signals s_i and θ_i, i=1,2, respectively.

(CMC5) M and $A_i, i = 1, 2$ decide whether to participate in the side-contract offered by M. If either of them decide not to, they play P's mechanism noncooperatively. Otherwise the game continues.

(CMC6) M, $A_i, i = 1, 2$ exchange cost messages e_M, e_1, e_2.

By construction the structure of the side contracting game in DM and CMC are the same, with the exception of the outside options additionally available to the suppliers in CMC of playing the grand contract offered by P noncooperatively. It is evident that P can achieve in CMC anything that she can achieve in DM, since she always has the option of offering a null contract to the suppliers in the grand contract. Hence (as in the comparison between DM and CC) delegation cannot perform better than centralization. The relevant question is whether delegation can achieve the optimal centralized outcome, given the presence of collusion between M and the suppliers.

5.1 Where M is Perfectly Informed

When M is perfectly informed, DM can indeed replicate the best centralized outcome of CMC. The intuitive reason is simple: the coalition is no longer subject to any internal asymmetric information when M is perfectly informed. In DM, the unit cost incurred by M in delivering the output to P is the first-best cost $c(\theta_1, \theta_2)$, since M can procure each input at its true cost, and production allocation across the suppliers will be efficient. Hence the relevant coalitions in centralization and delegation are subject to the same shadow cost of delivering output to P.

Proposition 3 [10] *When M is perfectly informed, P can attain the same expected profit in DM as in any equilibrium outcome of CMC.*

Increasing the outside options of the suppliers by offering them a fallback contract now do not reduce delivery cost because there are no productive distortions in DM, unlike the case of delegation to A_1 in DS. It merely has the effect of redistributing rents within the coalition, owing to the operation of the Coase Theorem.

The principle of ranking different regimes by the shadow costs of the associated agent coalitions (Proposition 1) allows us to compare the benefits of delegating to M rather than A_1. The DM regime corresponds to a one agent problem with *ex post* unit cost $c(\theta_1, \theta_2)$, while the DS regime corresponds to the higher unit cost $c(\theta_1, h_2(\theta_2))$ owing to the need for A_1 to pay informational rents to A_2. Hence it pays P to delegate to the better informed agent.

Proposition 4 *If M is perfectly informed, P obtains a higher expected profit in DM than in DS.*

Since DS is dominated by both DM and CC, it remains for us to compare DM with CC. Bringing M into the organization enables rents of suppliers to be eliminated, at the cost of allowing M to earn rents *vis-a-vis* the Principal. The regime DM effectively consolidates the two agents in CC into a single agent that supplies both inputs q_1, q_2 at a total cost of $\theta_1 q_1 + \theta_2 q_2$. Whereas in CC, P deals with two separate agents A_1 and A_2 that collude under conditions of asymmetric information.

It is useful to first compare DM with C, centralization in the absence of collusion, and then examine the effects of introducing collusion within centralization. The comparison between DM and C has been evaluated under special conditions by Baron and Besanko (1992), Gilbert and Riordan (1995), and Severinov (1999). We generalize their results below.

The comparison between DM and C turns out to depend on whether the two inputs q_1, q_2 are substitutes or complements, as defined below. Let $q_i^C(\theta)$ denote the second-best production assignments (defined in (1)), and $q_i^{DM}(\theta)$ the production assignments resulting in the optimal solution to DM.

[10] The proof employs an assumption of 'passive' beliefs off the equilibrium path, i.e., that such deviations do not result in any updating of priors. This helps avoid possible incentives for agents to refuse to participate in the side contract at the interim stage, in the anticipation that such deviations trigger other agents to change their beliefs in a manner that enhances their continuation payoffs in the ensuing noncooperative play of the grand contract.

Clearly, $q_i^{DM}(\theta) \equiv c_i(\theta) q^{DM}(c(\theta))$, where $q^{DM}(c)$ is the optimal output in DM as a function of the unit cost c reported by M.[11]

Definition 1 *The two inputs are said to be* **substitutes** *if* $q_i^{DM}(\theta_1, \theta_2)$ *is increasing in* $\theta_j, j \neq i = 1, 2$. *They are said to be* **complements** *if* $q_i^C(\theta_1, \theta_2)$ *is decreasing in* $\theta_j, j \neq i = 1, 2$.

The definition of substitutes refers to the optimal input assignments in DM, whereas that of complements refers to the centralized solution. These definitions do not directly refer to the parameters of the model. So it is more appropriate to view them as notions of 'strategic' substitutes or complements. These turn out to correspond to the common technical notions of substitutability and complementarity, as expressed by the following Lemma.

Lemma 1 *(1) The two inputs are substitutes if either of the following conditions hold:*

(1a) q_1 *and* q_2 *are perfect substitutes, i.e.,* $Q = q_1 + q_2$.

(1b) the production function Q *has constant elasticity of substitution which is sufficiently large, and* $c + G(c)/g(c)$ *is non-decreasing in* c, *where* $G(c)$ *denotes the distribution function of* $c(\theta_1, \theta_2)$.

(2) The two inputs are complements if for all (q_1, q_2):

$$\frac{\partial^2 V(Q(q_1, q_2))}{\partial q_1 \partial q_2} > 0$$

The following result generalizes Baron-Besanko (1992), Gilbert-Riordan (1995) and Severinov (1999).

Proposition 5 *Suppose that* M *is perfectly informed.*

(i) C *generates higher expected profits for* P *than DM if the two inputs are substitutes.*

(ii) DM generates a higher expected profit for P *than* C *if the two inputs are complements, and in addition* θ_1, θ_2 *are identically and exponentially distributed with* $\underline{\theta}_i = 0$.

[11] q^{DM} is the solution to the one-agent problem where the agent has unit cost $c(\theta)$, and is privately informed about this cost.

Consolidating two noncolluding suppliers into a single composite supplier has two effects. First, there is an effect of *internalization of bidding externalities*. The effect on P's profit depends on whether the two inputs are substitutes or complements. In the case of substitutes, a lower cost report submitted by one supplier reduces the quantity of input procured from the other supplier, lowering the latter's rents. Internalization of this externality with consolidation causes cost reports to be raised upward (on average), to P's disadvantage. This is essentially the effect of eliminating competition between the two suppliers. If instead the two inputs are complements, the effect goes in the opposite direction: cost underreporting by one supplier *raises* the input level procured from (and hence the rent earned by) the other supplier. Then consolidation lowers cost reports on average. In this case the enhanced coordination between the two supplying units is beneficial to P.

The second effect of consolidation is that P faces a single merged agent with two dimensional private information, who supplies two inputs rather than one. Replacement of a two dimensional incentive problem by a pair of separate one dimensional ones creates a *loss of control* for the Principal. This control-loss is represented by a more stringent set of incentive compatibility constraints, arising from the enhanced control over reporting and input assignments exercised by the consolidated agent. Any set of input assignments that can feasibly be implemented in DM (by some payment schedule) is also implementable in C, but the converse is not true. Implementability in C is represented simply by the requirement that the demand for each input is nonincreasing in the cost reported by the corresponding supplier. With consolidation it additionally requires the Jacobian of the input demand function be a symmetric, negative semidefinite matrix.[12] In addition relative input proportions will be selected by the merged agent on the basis of proportionality of marginal products to their relative (true) costs, whereas the Principal faces no such constraint in C. Hence implementability in DM requires an additional set of conditions, over and above the constraint that input demands be nonincreasing functions of own-cost. The loss of control by the Principal is a cost of merging the two suppliers.

When the two inputs are substitutes, it is clear that on both counts consolidation is disadvantageous to the Principal: it suppresses competition *and* accentuates incentive problems owing to control loss. So in general C is better in the substitutes case. When the inputs are complements instead, consolidation is advantageous to P through the internalization of bidding

[12]See McAfee and McMillan (1988) for a detailed explanation.

externalities across the separate suppliers, but could prove disadvantageous owing to control loss. In general this tradeoff is difficult to assess. In the special case where cost shocks follow identical, exponential distributions with a lower bound equal to zero, the increased stringency of the DM incentive constraints happens not to bite at the optimal solution to C: virtual costs are linear, symmetric functions of the true costs, so relative virtual costs are proportional to relative true costs. Hence relative factor proportions are not subject to additional distortions in DM. Moreover, the scale of total output in the solution to C is a function of a one dimensional sufficient statistic (the *ex post* minimum unit cost of delivering a unit of final output) of the costs of the two agents. The solution to C can thus be implemented in DM with a one-dimensional incentive scheme.[13] The control loss then does not bite, and DM comes out ahead on account of its ability to encourage coordination across the suppliers.

We now examine the implications of allowing collusion in the centralized regime. Since collusion among suppliers cannot improve P's profits, it follows that DM dominates CC whenever it happens to dominate C. Less straightforward is the case of substitute inputs where C dominates DM. We saw above that this is owed to the fact that the DM regime suppresses competition between the suppliers, unlike C where the suppliers cannot coordinate their reports. When suppliers collude, this benefit of the centralized regime may be substantially lost. Nevertheless we are able to show that the previous result extends even in the presence of collusion.

Proposition 6 [14] *The result of Proposition 5 extends even when suppliers collude under centralization (i.e., C is replaced by CC).*

In the case of substitutes, therefore, P benefits from dealing with a coalition that is more subject to problems of asymmetric information within the coalition. In CC, the presence of asymmetric information between the two suppliers leads to inefficient productive assignments within the coalition. Since the side contract is designed by A_1 it is prone to the same kind of inefficiency as in DS, owing to A_1's incentive to reallocate production assignments in his own favor at the expense of A_2. This inefficiency can be limited by P offering a grand contract which offers generous terms to A_2.

[13] We have been able to extend this result to some cases where the costs of the two agents are nonidentical but follow exponential distributions.

[14] The proof that CC strictly dominates DM in the case of substitutes employs an assumption of 'passive' beliefs off the equilibrium path, i.e., that such deviations do not result in any updating of priors. This assumption is not required to show the corresponding weak inequality result. See the proof for further details.

The proof that CC can replicate the profits achieved in DM in the case of substitutes is a simple modification of the argument used to establish Proposition 2. As explained above, the construction used in that argument with any parameter α less than one induces an optimal side contract with Kuhn-Tucker multiplier $1-\alpha$ assigned by A_1 to the rents of A_2. When the two inputs are substitutes and $\alpha = 0$, the resulting shadow cost for A_1 in procuring from A_2 turns out to be $z_0(\theta_2) = \theta_2$, the same as incurred by M in the arrangement DM. In CC, A_1 can thus be motivated to behave the same way as a perfectly informed intermediary M in DM in the substitutes case.

Unfortunately the argument cannot be modified to demonstrate that CC achieves a strictly higher profit. The strict inequality is established by an entirely different grand contract, which employs a similar idea. This grand contract lowers A_1's shadow cost of procuring q_2 strictly below A_2's true cost θ_2 when the two inputs are substitutes. The delivery cost of the supplier coalition in centralization then falls below the delivery cost of M, permitting centralization to strictly dominate DM.

5.2 When M is Not Better Informed Than P

The preceding results are substantially altered when M is not better informed than P. In that case the delivery cost of M reduces to the second-best unit cost function $c(h_1(\theta_1), h_2(\theta_2))$, which is uniformly higher than the delivery cost $c(\theta_1, h_2(\theta_2))$ of A_1 in the DS regime. Applying Proposition 1 again yields the following result.

Proposition 7 *If M has the same information as P, DM generates lower expected profit to P compared with DS.*

Since Proposition 2 implies that DS in turn is inferior to CC, it follows that in this case DM is the worst of the three organizational alternatives.

6 Relation to Existing Literature

Comparisons between variants of DS and C have been the subject of a substantial literature, which includes Baron and Besanko (1992), Melumad, Mookherjee and Reichelstein (1992, 1995, 1997), Mookherjee and Reichelstein (1997, 2001), and Severinov (1999). Reasons why DM performs poorly relative to C have been explored in McAfee and McMillan (1995) and Faure-Grimaud and Martimort (2001). These models therefore cannot explain why DM may dominate centralization or delegation to suppliers. As Radner (1993) and van Zandt (1996) have persuasively argued, there is a need to explain the widespread phenomenon of delegation of control to managers or intermediaries who play no direct role in the productive process. Moreover, this literature does not consider versions of centralization with collusion.

The papers most closely related to ours deal with the question of centralization versus decentralization in the presence of collusion. Laffont and Martimort (1998) deal with this question in an adverse selection setting with two suppliers that produce perfectly complementary inputs, when each supplier's costs take two possible values. They compare the DS regime with a formulation of CC that differs from ours. In centralization they assume that the supplier coalition is organized by a fourth party that cares symmetrically about the utilities of both suppliers, so the side contract is designed to maximize the sum of their *ex ante* utilities. Hence the allocation of bargaining power changes exogenously in favor of A_2 when the organization switches from decentralization to centralization. In our formulation the structure of the side contracting game is the same in both organizational variants, and the allocation of bargaining power changes endogenously. In their model Laffont and Martimort (1998) find that delegation to a supplier achieves the same outcome as centralization, a result which owes partly to their assumption of two possible cost types for each supplier, and perfect complementarity of inputs.[15]

Faure-Grimaud *et al* (2003) consider a version of DM where there is one supplier, and a (possibly risk-averse) supervisor M who is imperfectly informed about the supplier's cost. The supervisor's signal is positively correlated with the supplier's cost, and each can take two possible values. They compare delegation to M with centralization where P contracts with both agents (which corresponds to CMC here). In both variants the supervisor and supplier collude in a similar way, with the supervisor making a take-it-or-leave-it side contract offer to the supplier. However one difference in their formulation is that participation decisions of the supplier and supervisor in the grand contract offered by P are made at the *interim* stage, before the two agents have had an opportunity to communicate with one another. This feature limits the severity of double marginalization of rents that delegation is prone to. Their main result is that delegation and centralization achieve equivalent outcomes, mirroring our result that DM and CMC are equivalent.

[15]They also find that delegation achieves a superior outcome when communication costs in centralization restrict centralized contracts to be 'anonymous', wherein the two suppliers must be treated symmetrically.

However, their results differ in one crucial respect: in their context coalitional agreements are inefficient since the supervisor is imperfectly informed about the agent's cost. Despite this inefficiency, it does not pay for P to contract with the supplier at the same time. The particular information structure in Faure Grimaud *et al* plays a special role in this regard.[16] This is argued by Celik (2002), who finds that delegation to the supervisor is dominated by centralization with a different information structure. In Celik's model, the agent's cost has an arbitrary finite support and the supervisor's information is represented by a connected partition over the cost space. With such a structure, only 'downward' incentive constraints bind, and in that context it pays the principal to create 'countervailing' incentives by contracting with the supplier also in order to raise his outside option. This is analogous to the argument underlying Proposition 2 in this paper.

Baliga and Sjostrom (1998) consider a setting with two suppliers that collectively produce an output for the Principal, under conditions of moral hazard with limited liability rather than adverse selection. The suppliers work sequentially, with the downstream supplier able to monitor the effort of the upstream supplier. Collusion between the suppliers is not subject to any asymmetric information. However, limited liability constraints may prevent coalitional agreements from achieving efficiency. In this setting they find delegation equivalent to centralization (i.e., their analogues of DS and CC are equivalent) for some parameter values, while for others delegation is strictly dominated. In the latter cases, the distortion appears to be similar to a double marginalization of rent problem. Nevertheless the fact that for some parameter values they find DS and CC to be equivalent is in contrast to our finding in Proposition 2. Conceivably this could owe either to different assumptions concerning the underlying source of an incentive problem (moral hazard rather than adverse selection), or different range of actions available to agents (i.e., whether these are discrete or continuous).

Baron and Besanko (1999) consider an adverse selection setting where two suppliers produce perfectly complementary inputs. They treat as endogenous the question whether the two suppliers consolidate into an 'information alliance' similar to DM. In an alliance a fourth party gains access to information about each supplier's cost, and operates as a single agent

[16]Specifically, 'upward' rather than 'downward' incentive constraints bind, wherein the supervisor-supplier coalition is tempted to report a high cost signal when the low signal is observed by the supervisor (and the actual cost type of the supplier is high) and thereby qualify for a higher production target. In such a context countervailing incentives generated by higher outside options of the supplier do not benefit the Principal. We thank Gorkem Celik for offering us this explanation.

vis-a-vis the Principal. At the first stage of the game, the two suppliers decide whether to remain separate (as in C, centralization without collusion) or form an alliance (create DM). The Principal then reacts with a contract offer (to the pair of suppliers separately if they decided to remain separate, or to the consolidated entity if they form an alliance). They produce examples where such alliances form endogenously, and results in a Pareto improvement. Their formulation thus incorporates endogenous selection of organizational form by the agents themselves, to which the Principal reacts passively. Our formulation in contrast presumes that the Principal can select the organizational form (e.g., DM versus CC or C), to which the agents react passively. Our formulation seems more appropriate for Principals that are 'big players' such as government procurement or regulatory agencies, who have the power to mandate how regulated entities should be constituted. Baron and Besanko's formulation appears more appropriate for settings where the Principal is a purchasing firm in a market setting that has to adapt to whichever organizational form that suppliers create in their own interest. Nevertheless, their result that informational alliances can benefit the Principal is in line with the result that DM dominates C, centralization without collusion, from P's point of view when the inputs supplied are complementary. Our results suggest that the Pareto-improving character of alliances will not extend to the case where the inputs are substitutes.

A different literature on incentive effects of delegation introduces contract renegotiation or incomplete contracts. Papers focusing on contract renegotiation include Dessein (2002) and Poitevin (1995, 2000), while Aghion and Tirole (1997) use an incomplete contract approach. Moreover, all these papers deal with the question of delegation of authority between a Principal and a single agent, and do not address questions pertaining to delegation of authority to third party managers or intermediaries possessing specialized information.

7 Concluding Comments

The main conclusion of this paper is that decentralization cannot generally be justified as an optimal organizational response to the presence of collusion among agents. Retaining control with regard to contracting with every relevant agent in the organization enables the Principal to limit problems of double marginalization of rents inherent in vertical side contracting relationships among agents. For instance, centralization dominates delegation to one of the suppliers or to intermediaries with no informational advantage over the Principal. Only in certain circumstances can delegation be justified (e.g., when authority is delegated to a well-informed intermediary and the inputs supplied are complementary). The theory thus predicts circumstances (defined by complementarity or substitutability of activities, and the dispersion of information among agents) where delegation arrangements are likely to be more prevalent.

Our analysis also provided some results concerning choice of who to delegate authority to (assuming that P has chosen to delegate to someone). The general principle is to choose a coalition with the lowest shadow cost of delivering the final output. Accordingly, if the intermediary is better (resp. more poorly) informed than any given supplier, then it is best for the Principal to delegate to the intermediary (resp. the supplier). But it is not generally true that coalitions less subject to internal asymmetric information have a lower shadow cost. When the inputs are substitutes, for instance, P is better off in CC with a coalition consisting of two asymmetrically informed suppliers, compared with DM where the coalition is not subject to any asymmetric information at all.

We considered the polar cases where the intermediary M is either perfectly informed about agent costs, or has the same information as P. What happens in the intermediate case where M is better informed than P, yet is imperfectly informed about supplier costs? Here the precise information structure between M and the suppliers plays an important role. Suppose that each supplier A_i observe the signals s_i observed by M.[17] Then the intermediary has no private information within the coalition, while suppliers have better information regarding their own costs than M. The resulting coalitional side contract will involve distortions, with each supplier earning information rents. The shadow cost of M will now be higher than the case where M is perfectly informed, lowering the value of DM to P. The value of DM to P will typically increase continuously as M becomes better informed. We thus expect that the result of Propositions 4 and 6 will continue to apply if M is sufficiently better informed than P, whereas the result of Proposition 7 will apply if M is not much better informed than P. Intermediaries will thus be relied on only if they have sufficient informational expertise relative to P or suppliers.

Our model considered the case where all the bargaining power in side-contracting rests with one of the agents. This matters for the outcomes

[17]In the other case, M and the agents observe correlated signals, whence Cremer-McLean mechanisms can be designed by M within the coalition to extract all information rents from the agents. The results will then be the same as when M is perfectly informed (assuming that M and the suppliers are risk-neutral).

of DS and CC where there is asymmetric information within the coalition, while the outcome of DM would be unchanged if M is perfectly informed. It would be interesting to extend the analysis to the case of a more general allocation of bargaining power within the coalition, e.g., where side contracts are designed by an arbitrator who assigns arbitrary welfare weights to different agents.

Our analysis abstracted from the possibility that the middleman/manager may be risk averse. We also did not allow for possible collusion-within-collusion, which would affect coalitions with more than two members. For instance in DM we did not allow the suppliers to collude against M, which further restricts the design of side contracts. In the context where M is perfectly informed, however, this is unlikely to alter our results: since coalitional agreements are internally efficient, there would be no further room for subcoalitions to form. In more general situations, however, the effects of subcoalitions on side contracts would need to be explored further.

We assumed throughout the cost uncertainties of the two suppliers are uncorrelated. In the presence of correlation, centralization possesses some additional advantages, wherein the Principal can extract all the rents of the suppliers in the absence of any collusion (i.e., by using Cremer-McLean (1988) type of mechanisms). In the presence of collusion, however, the feasibility of such rent extraction mechanisms is likely to be considerably restricted (see, for instance, Laffont and Martimort (2000)). The comparison between delegation and centralization in such a context needs to be studied in future research.

We assumed that the principal could only monitor the quantity of final output delivered, but not the allocation of inputs supplied by the agents. This widened the scope of collusion to reallocate production assignments among suppliers. If the principal could monitor inputs then one of the distortions associated with delegation to a supplier could be avoided, wherein the 'managing' supplier procures too little from the other supplier. But it would not alleviate the more general problem of double marginalization of rents which causes the principal to procure too little of the final good. In the case of i.i.d. exponentially distributed cost shocks (with a lower bound of zero) DS continues to be dominated by CC for any technology. We suspect the result extends to more general classes of distributions. Concerning the comparison of DM and CC, input monitoring by the principal would typically enhance the value of DM because it would enable the principal to control the intermediary better. In the case where the principal's benefit is additively separable in the two inputs, it can be shown that DM always dominates CC.[18] This is consistent with the result of Dana (1993) which shows that consolidation of two agents is beneficial with input monitoring if their cost shocks are independent or negatively correlated. In the case of nonseparabilities in production, the competition and coordination effects studied in this paper would additionally come into play.

The model studied in this paper could be developed further to accommodate a wider variety of questions regarding the design of organizational structure. One of these deal with effects of changes in information technology that improve the Principal's information relative to intermediaries or suppliers. Another deals with the grouping of activities within multiproduct firms. We hope to address these questions in future work.

References

[1] Aghion, P. and J. Tirole (1997), "Formal and Real Authority in Organizations," *Journal of Political Economy*,105(1):1-29.

[2] Baliga, S. and T. Sjostrom (1998), "Decentalization and Collusion," *Journal of Economic Theory,* 83:196-232.

[3] Banerjee A., P. Gertler and M. Ghatak (2002), "Empowerment and Efficiency: Tenancy Reform in West Bengal," *Journal of Political Economy* 110(2):239-280.

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

[5] Baron, D. P. and D. Besanko (1999), "Informational Alliances," *Review of Economic Studies,* 66:743-768.

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

[7] Celik G. (2002), "Mechanism Design with Collusive Supervision," mimeo, December 2002, Department of Economics, University of British Columbia, Vancouver.

[18] This result was included in a previous version of the paper, and is available on request.

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

[9] Dana, J. (1993), "The Organization and Scope of Agents: Regulating Multiproduct Industries," *Journal of Economic Theory,* 59(2), 288–310.

[10] Dessein, W. (2002), "Authority and Communication in Organizations", *Review of Economic Studies*, 69, 811–838.

[11] Eaton, M. L. (1987), "Lectures on Topics in Probability Inequalities," *Centrum voor Wiskunde en Infomatica,* Amsterdam, Netherlands.

[12] Faure-Grimaud, A. and D. Martimort (2001), "Some Agency Costs of Intermediated Contracting," *Economics Letters,* 71(1):75-82.

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

[14] Gilbert, R. J. and M. H. Riordan (1995), "Regulating Complementary Products:A Comparative Institutional Analysis," *Rand Journal of Economics,* 26:243-256.

[15] Jullien, B. (2000), "Participation Constraints in Adverse Selection Models", *Journal of Economic Theory,* 93(1):1-47.

[16] Laffont, J-J. and D. Martimort (1998), "Collusion and Delegation," *Rand Journal of Economics,* 29:280-305.

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

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

[19] McAfee, R.P. and J. McMillan (1988), "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory,* 46:335-354.

[20] McAfee, R.P. and J. McMillan (1995), "Organizational Diseconomies of Scale," *Journal of Economics and Management Strategy,* 4:399-426.

[21] Maskin, E. and J. Tirole (1990), "The Principal-Agent Relationship with an Informed Principal. I:Private Values." *Economerica,* 58:379-410.

[22] Melumad, N. D., D. Mookherjee and S. Reichelstein (1992), "A Theory of Responsibility Centers," *Journal of Accounting and Economics,* 15:445-484.

[23] Melumad, N. D., D. Mookherjee and S. Reichelstein (1995), "Hierarchical Decentralization of Incentive Contracts," *Rand Journal of Economics,* 26:654-672.

[24] Melumad, N. D., D. Mookherjee and S. Reichelstein (1997), "Contract Complexity, Incentives, and the Value of Delegation," *Journal of Economics and Management Strategy,* 6:257-289.

[25] Mookherjee, D. (1997), "Informational Rents and Property Rights in Land,"in J. Roemer(Edited) *Property Relations, Incentives and Welfare*, London:McMillan (for International Economic Association).

[26] Mookherjee, D. and S. Reichelstein (1997), "Budgeting and Hierarchical Control," *Journal of Accounting Research,* 35(2): 129-55.

[27] Mookherjee, D. and S. Reichelstein (2001), "Incentives and Coordination in Hierarchies," *Advances in Theoretical Economics,* 1(1).

[28] Myerson, R. B.(1981), "Optimal Auction Design,", *Mathematics of Operation Research,* 6:58-73.

[29] Poitevin, M. (1995), "Contract Renegotiation and Organizational Design,", mimeo, CIRANO Working Paper No. 95-3, Montreal.

[30] Poitevin, M. (2000), "Can the Theory of Incentives Explain Decentralization?," *Canadian Journal of Economics,* 33(4):878-906.

[31] Radner, R. (1993), "The Organization of Decentralized Information Processing," *Econometrica*, 61:1109-1146.

[32] Severinov, S. (1999), "Optimal Organization: Centralization, Decentralization or Delegation?," *mimeo,* Department of Economics, University of Wisconsin, Madison.

[33] Tirole, J. (1986), "Hierarchies and Bureaucracies: on the Role of Collusion in Organizations," *Jounral of Law, Economics and Organization,* 2(2):181-214.

[34] Tirole, J. (1988), *The Theory of Industrial Organization*, Cambridge, MA: MIT Press.

[35] Van-Zandt, T. (1996), "Decentralized Information Processing in the Theory of Organizations," Sertel, M. ed. *Contemporary Economic Development Reviewed, Volume 4: The Enterprise and its Environment*, MacMillan, London.

8 Appendix:Proofs

Proof of Proposition 1: Using the techniques of Baron and Myerson (1982), the optimal solution to the problem of contracting with an agent with unit cost c_1 can be expressed as follows. The optimal output $q(c)$ corresponding to cost report c solves $V'(q) = z_1(c)$, where $z_1(.)$ is a (rent-inclusive) cost function defined over the support $[\underline{c}_1, \bar{c}_2]$ of c_1 which is continuous, nondecreasing, strictly increasing over a set of positive (c_1) measure, and such that $z_1(c_1) \geq c_1$ with strict inequality holding almost everywhere. Moreover the optimal payment function is

$$X(c_1) = c_1 q(c_1) + \int_{c_1}^{\bar{c}_1} q(c)dc. \tag{7}$$

It follows that $X(c)$ and $q(c)$ are both continuous, nonincreasing functions, which are strictly decreasing over sets of positive (c_1) measure.

Now suppose P contracts instead with an agent with unit cost c_2 which is first-order stochastically dominated by c_1. If the support of c_2 is $[\underline{c}_2, \bar{c}_2]$ then it is evident that $\underline{c}_2 \leq \underline{c}_1$ and $\bar{c}_2 \leq \bar{c}_1$. Let P offer the following output and payment schedule (over the range $[\underline{c}_2, \bar{c}_1]$):

$$q_2(c) = \begin{cases} q(\underline{c}_1) & \text{if } c \in [\underline{c}_2, \underline{c}_1], \\ q(c) & \text{if } c \in (\underline{c}_1, \bar{c}_1]. \end{cases} \tag{8}$$

$$X_2(c) = \begin{cases} X(\underline{c}_1) & \text{if } c \in [\underline{c}_2, \underline{c}_1], \\ X(c) & \text{if } c \in (\underline{c}_1, \bar{c}_1]. \end{cases} \tag{9}$$

Clearly, q_2 and X_2 are both continuous nonincreasing functions. This contract satisfies incentive compatibility and interim participation constraints for the agent with cost c_2. Define P's *ex post* profit function $p(c) \equiv V(q_2(c)) - X_2(c)$ over the range $[\underline{c}_2, \bar{c}_1]$. This function is continuous and differentiable a.e. because $q_2(.)$ and $X_2(.)$ have these properties. At any point of differentiability, $p'(c) = [V'(q_2(c)) - c]q_2'(c) = [z_1(c) - c]q_2'(c)$, which is everywhere nonpositive and strictly negative over a set of positive (c_1) measure. Since the expected profit of P with the c_i-cost agent is $p(\bar{c}_1) - \int_{\underline{c}_2}^{\bar{c}_1} G_i(c)p'(c)dc$, it is no lower with the c_2-cost agent, and strictly higher if $G_2(c) > G_1(c)$ for (almost) all c in the support of c_1. ■

Proof of Proposition 2: Let $q(c)$ denote the optimal output function in DS, where c denotes the unit cost $c = c(\theta_1, h_2(\theta_2))$ of A_1 for delivering output to P. Incentive compatibility implies this is a nonincreasing function, and is implemented at minimum cost by the payment function $X(c) = cq(c) + \int_c^{\bar{c}} q(y)dy$ where $\bar{c}$ denotes the upper bound of c, i.e. $\bar{c} = c(\bar{\theta}_1, h_2(\bar{\theta}_2))$.

For $\alpha \in [0,1]$ we define the following functions: modified virtual cost $z_\alpha(\theta_2) \equiv \theta_2 + \alpha\frac{F_2(\theta_2)}{f_2(\theta_2)}$; associated production assignments $q^\alpha(\theta) \equiv q(c(\theta_1, z_\alpha(\theta_2))), q_i^\alpha(\theta) \equiv c_i(\theta_1, z_\alpha(\theta_2))q^\alpha(\theta)$ (which is also sometimes denoted by $q_i(\theta_1, z_\alpha(\theta_2))$);

$$X_\alpha^C(\theta) \equiv \theta_1 q_1^\alpha(\theta) + \theta_2 q_2^\alpha(\theta) + \int_{\theta_1}^{\bar{\theta}_1} q_1^\alpha(y, \theta_2)dy + \int_{\theta_2}^{\bar{\theta}_2} q_2^\alpha(\theta_1, y)dy$$

the cost of implementing these assignments in C; $X^C_\alpha \equiv E[X^C_\alpha(\theta)] = E[h_1(\theta_1)q^\alpha_1 + h_2(\theta_2)q^\alpha_2]$ the associated *ex ante* expected cost of implementing these assignments in C;

$$X^D_\alpha(\theta) \equiv c(\theta_1, z_\alpha(\theta_2))q(c(\theta_1, z_\alpha(\theta_2))) + \int_{c(\theta_1, z_\alpha(\theta_2))}^{c(\bar{\theta}_1, z_\alpha(\bar{\theta}_2))} q(c)dc$$

the transfer function that implements the output function $q(c)$ with a single agent that delivers this output at unit cost $c(\theta_1, z_\alpha(\theta_2))$; and $X^D_\alpha \equiv E[X^D_\alpha(\theta)]$ the associated *ex ante* expected cost. Note that with $\alpha = 1$, this is exactly the expected payment made by P in the optimal mechanism in DS: $X^D_1 = E[X(c(\theta_1, h_2(\theta_2)))]$.

Claim: For any $\alpha \in [0, 1]$, there exists a grand contract which P can offer in the centralized regime which:

(a) has a (noncooperative) Bayesian equilibrium in which both agents always participate and tell the truth, resulting in the production assignments q^α_i at an expected cost of $\max\{X^C_\alpha, X^D_\alpha\}$ to P;

(b) is collusion-proof, in the sense that it is optimal for A_1 to offer a null side-contract whereby they play the (noncooperative) Bayesian equilibrium described above.

This Claim will establish the proof. To see this, note that at $\alpha = 1$ the outcome of the truthful Bayesian equilibrium of such a mechanism corresponds to the optimal outcome in DS, since $X^D_1 > X^C_1$ by the argument described in the text leading upto (4). Moreover, for α close enough to 1, it will be the case that $X^D_\alpha > X^C_\alpha$. By (a) above, the outcome of CC given the grand contract described enough for any such α will implement the same output function $q(c(\theta_1, z_\alpha(\theta_2)))$ as in the optimal solution to DS at a cost of X^D_α. Since the unit cost of A_1 in DS first order stochastically dominates $c(\theta_1, z_\alpha(\theta_2))$ for any $\alpha < 1$, the argument of Proposition 1 implies that P earns a higher expected profit from the outcome of CC than the optimal outcome in DS.

To prove the claim, let P offer the following grand contract in CC:

$$X_2(\theta) = E_{\theta_1}[\theta_2 q^\alpha_2(\theta_1, \theta_2) + \int_{\theta_2}^{\bar{\theta}_2} q^\alpha_2(\theta_1, y)dy] + z_\alpha(\theta_2)[q^\alpha_2(\theta_1, \theta_2) - E_{\theta_1} q^\alpha_2(\theta_1, \theta_2)] \quad (10)$$

$$X_1(\theta) = X^D_\alpha(\theta) - X_2(\theta) + \max\{X^C_\alpha - X^D_\alpha, 0\} \quad (11)$$

along with the production assignments $q^\alpha_i(\theta)$.

To show that (a) is true for this mechanism, note that

$$X_1(\theta) = \theta_1 q^\alpha_1(\theta_1, \theta_2) + \int_{\theta_1}^{\bar{\theta}_1} q^\alpha_1(y, \theta_2)dy + N(\theta_2) \quad (12)$$

where

$$
\begin{aligned}
N(\theta_2) \equiv \quad & \int_{z_\alpha(\theta_2)}^{z_\alpha(\bar{\theta}_2)} q_2(\bar{\theta}_1, z)dz \\
& - E_{\theta_1}[\theta_2 q_2^\alpha(\theta_1,\theta_2) + \int_{\theta_2}^{\bar{\theta}_2} q_2^\alpha(\theta_1, y)dy] \\
& + z_\alpha(\theta_2) E_{\theta_1} q_2^\alpha(\theta_1,\theta_2) + \max\{X_\alpha^C - X_\alpha^D, 0\}
\end{aligned}
$$

so it is clear that conditional on participating A_1 has a dominant strategy of telling the truth. And if A_2 always participates and tells the truth, then since

$$X_\alpha^C - X_\alpha^D = E[(h_2(\theta_2) - z_\alpha(\theta_2))q_2^\alpha(\theta_1,\theta_2) - \int_{z_\alpha(\theta_2)}^{z_\alpha(\bar{\theta}_2)} q_2(\bar{\theta}_1, z)dz] \qquad (13)$$

it follows that $E_{\theta_2} N(\theta_2) \geq 0$ and A_1's interim payoff from participating is always nonnegative.

Conversely, given that A_1 always participates and tells the truth, A_2's interim payoff is

$$E_{\theta_1}[X_2(\theta) - \theta_2 q_2^\alpha(\theta_1,\theta_2)] = E_{\theta_1} \int_{\theta_2}^{\bar{\theta}_2} q_2^\alpha(\theta_1, y)dy \geq 0,$$

which (combined with the monotonicity of the production assignments) implies it is a best response for A_2 to always participate and tell the truth. This grand contract thus implements q_i^α as a Bayesian equilibrium at expected cost

$$E[X_1(\theta) + X_2(\theta)] = X_\alpha^D + \max\{X_\alpha^C - X_\alpha^D, 0\} = \max\{X_\alpha^C, X_\alpha^D\},$$

which establishes (a).

To establish (b), i.e., collusion-proofness of this equilibrium, we next consider the induced side contract design problem for A_1. This differs from a conventional contract design problem with respect to the participation constraints. If either of the agents decide to exit from the side-contract at the interim stage, then the game is not over, since they subsequently play the grand contract noncooperatively. Accordingly the payoffs from this noncooperative game define the outside options of the two agents. These outside options depend on the types of the agents, as well as the beliefs updated by observation of the exit decision. In turn these beliefs themselves depend on the nature of the side contract (which determines the incentives of the agents to participate in it). So the outside options are endogenously determined by the side contract.

To deal with this problem, we proceed by first noting a version of the Revelation Principle which applies to the side-contracting problem.

Lemma 2 *Consider any given grand contract $X_i(m_1, m_2), q_i(m_1, m_2)$ offered by P. Suppose that A_1 subsequently offers a side contract in which there is a perfect*

Bayesian equilibrium (PBE) (in pure strategies). Then there exists a revelation side-contract and a perfect Bayesian equilibrium of this side contract in which both agents decide to participate at the interim stage, and subsequently exchange cost messages truthfully, which generates an outcome equivalent to the PBE of the original side contract.

While the argument is standard, it is nevertheless useful to outline the argument in order to illustrate the implications of the interim exit options which alter subsequent beliefs of the agents when they exchange cost messages. We use the following notation. Let the original side contract be denoted $(m_1(e), m_2(e), t_2(e), \tilde{q}_1(e), \tilde{q}_2(e))$. Let $P_i \in \{0,1\}$ denote the participation decision of agent A_i at the interim stage (with 0 denoting exit). These decisions may cause agents' beliefs about each others types to be updated, thereby affecting their subsequent choice of cost reports. Let the PBE cost reports be denoted $m^*(\theta, P_1, P_2)$ when they play the grand contract noncooperatively following the participation decisions P_1, P_2 where $P_1 P_2 = 0$, and $e^*(\theta)$ the reports they exchange when they both agree to participate in the side contract. These satisfy the following incentive constraints (where conditional expected utility of each agent is taken with respect to the event corresponding to observed participation decisions) :

(i)

$$\begin{aligned} m_1^*(\theta_1, P_1, P_2) &= \arg\max_{m_1} E_{\theta_2}[X_1(m_1, m_2^*(\theta_2, P_1, P_2)) \\ &- \theta_1 q_1(m_1, m_2^*(\theta_2, P_1, P_2)) \mid P_2(\theta_2) = P_2] \end{aligned}$$

$$\begin{aligned} m_2^*(\theta_2, P_1, P_2) &= \arg\max_{m_2} E_{\theta_1}[X_2(m_1^*(\theta_1, P_1, P_2), m_2) \\ &- \theta_2 q_2(m_1^*(\theta_1, P_1, P_2), m_2) \mid P_1(\theta_1) = P_1] \end{aligned}$$

(ii)

$$\begin{aligned} e_1^*(\theta_1) &\equiv \arg\max_{e_1} E_{\theta_2}[X_1(m(e_1, e_2^*(\theta_2))) \\ &- t_2(e_1, e_2^*(\theta_2)) - \theta_1 \tilde{q}_1(e_1, e_2^*(\theta_2)) \mid P_2(\theta_2) = 1] \end{aligned}$$

$$\begin{aligned} e_2^*(\theta_2) &\equiv \arg\max_{e_2} E_{\theta_1}[X_2(m(e_1^*(\theta_1), e_2)) \\ &+ t_2(e_1^*(\theta_1), e_2) - \theta_2 \tilde{q}_2(e_1^*(\theta_1), e_2) \mid P_1(\theta_1) = 1] \end{aligned}$$

We can define the following revelation side-contract:

(i) $(\hat{m}(\theta), \hat{t}_2(\theta), \hat{q}_i(\theta)) = (m^*(\theta, P_1(\theta_1), P_2(\theta_2)), 0, q_i(m^*(\theta, P_1(\theta_1), P_2(\theta_2))))$ if $P_1(\theta_1)P_2(\theta_2) = 0$.

(ii) $(\hat{m}(\theta), \hat{t}_2(\theta), \hat{q}_i(\theta)) = (m(e^*(\theta))), t_2(e^*(\theta)), \tilde{q}_i(e^*(\theta)))$ if $P_1(\theta_1)P_2(\theta_2) = 1$.

This side-contract replicates for given cost reports the outcome of the PBE of the previous side-contract corresponding to the state of the world where these were the true costs. By standard arguments, there is a PBE of this revelation side-contract where the agents participate and report truthfully in all states. This completes the proof of Lemma 2.

However, the implications of Lemma 2 for representation of the optimal side-contracting problem are non-standard, owing to difficulties in representing interim participation constraints. Note that the incentive constraints associated with truthful reporting are standard: for any θ_1 and $\theta_1^{'}$,

$$\begin{aligned} & E_{\theta_2}[X_1(\hat{m}(\theta)) - \hat{t}_2(\theta) - \theta_1 \hat{q}_1(\theta)] \\ \geq \quad & E_{\theta_2}[X_1(\hat{m}(\theta_1^{'}, \theta_2)) - \hat{t}_2(\theta_1^{'}, \theta_2) - \theta_1 \hat{q}_1(\theta_1^{'}, \theta_2)] \end{aligned}$$

and for any θ_2 and $\theta_2^{'}$,

$$\begin{aligned} & E_{\theta_1}[X_2(\hat{m}(\theta)) + \hat{t}_2(\theta) - \theta_2 \hat{q}_2(\theta)] \\ \geq \quad & E_{\theta_1}[X_2(\hat{m}(\theta_1, \theta_2^{'})) + \hat{t}_2(\theta_1, \theta_2^{'}) - \theta_2 \hat{q}_2(\theta_1, \theta_2^{'})] \end{aligned}$$

The (interim) participation constraints are more complicated: each agent must get at least what she would get in noncooperative play of the grand contract, if she were to exit from the side contract. For A_2 for instance, this inequality states that for any θ_2,

$$\begin{aligned} & E_{\theta_1}[X_2(\hat{m}(\theta)) + \hat{t}_2(\theta) - \theta_2 \hat{q}_2(\theta)] \\ \geq \quad & \Pr(P_1(\theta_1) = 0) E_{\theta_1}[X_2(m^*(\theta, 0, 0)) - \theta_2 q_2(m^*(\theta, 0, 0)) \mid P_1(\theta_1) = 0] \\ + \quad & \Pr(P_1(\theta_1) = 1) E_{\theta_1}[X_2(m^*(\theta, 1, 0)) - \theta_2 q_2(m^*(\theta, 1, 0)) \mid P_1(\theta_1) = 1] \end{aligned} \tag{14}$$

This participation constraint is stated in terms of the original side-contract and PBE; it depends on the beliefs held by agents following observation of interim exit decisions. So Lemma 2 cannot be used to represent the set of possible outcomes from an arbitrary side-contract by a set of revelation side-contracts satisfying a simple set of constraints (that depend only on the revelation contract itself).

We deal with this problem by considering a wider class of revelation side contracts, which satisfy a weaker set of interim participation constraints. Note that in any continuation game following the interim exit/participation decisions in the original PBE of the original side-contract, every type of every agent will receive nonnegative utility. Hence the same will be true in every equivalent revelation side contract:

$$\begin{aligned} E_{\theta_2}[X_1(\hat{m}(\theta)) - \hat{t}_2(\theta) - \theta_1 \hat{q}_1(\theta)] & \geq 0 \\ E_{\theta_1}[X_2(\hat{m}(\theta)) + \hat{t}_2(\theta) - \theta_2 \hat{q}_2(\theta)] & \geq 0 \end{aligned} \tag{15}$$

The converse of course may not be true: any revelation side contract satisfying the truth-telling constraints and the nonnegative interim utility property (15) may not be feasible for A_1, since it may violate the 'true' interim participation constraint which (typically) involves positive outside options for most types of either agent.

Note, however, if A_1 offers the null side contract then the agents always play the grand contract noncooperatively, and learn nothing from observing interim exit/participation decisions. In that case the outside options are defined by the expected payoff from playing the grand contract noncooperatively corresponding to their priors. Since this is exactly the payoff they achieve in equilibrium, it is clear that the null side contract (or the revelation side contract which replicates it) is (trivially) feasible.

In order to show that it is optimal for A_1 to offer the null side contract, it suffices to show that it is optimal within the class of revelation side contracts that are incentive compatible and satisfy the relaxed participation constraints (15). We shall relax the constraints even further, by dropping incentive and participation constraints for A_1 herself.

The (doubly) relaxed side contract design problem corresponds to the following optimization exercise: select joint report $\hat{m}(\theta_1, \theta_2)$, side payment $\hat{t}_2(\theta_1, \theta_2)$ and production assignments $\hat{q}_i(\theta_1, \theta_2)$ to

$$\max E[X_1(\hat{m}(\theta_1, \theta_2)) - \hat{t}_2(\theta_1, \theta_2) - \theta_1 \hat{q}_1(\theta_1, \theta_2)]$$

subject to (for all θ:)

$$\begin{aligned} & E_{\theta_1}[X_2(\hat{m}(\theta_1, \theta_2)) + \hat{t}_2(\theta_1, \theta_2) - \theta_2 \hat{q}_2(\theta_1, \theta_2)] \\ \geq \quad & E_{\theta_1}[X_2(\hat{m}(\theta_1, \theta_2^{'})) + \hat{t}_2(\theta_1, \theta_2^{'}) - \theta_2 \hat{q}_2(\theta_1, \theta_2^{'})] \end{aligned} \tag{16}$$

$$E_{\theta_1}[X_2(\hat{m}(\theta_1, \theta_2)) + \hat{t}_2(\theta_1, \theta_2) - \theta_2 \hat{q}_2(\theta_1, \theta_2)] \geq 0 \tag{17}$$

$$\begin{aligned} & E[X_2(\hat{m}(\theta_1, \theta_2)) + \hat{t}_2(\theta_1, \theta_2) - \theta_2 \hat{q}_2(\theta_1, \theta_2)] \\ \geq \quad & E[X_2(\theta_1, \theta_2) - \theta_2 q_2^{\alpha}(\theta_1, \theta_2)] \end{aligned} \tag{18}$$

and

$$Q(\hat{q}_1(\theta), \hat{q}_2(\theta)) = q^{\alpha}(\hat{m}(\theta)).$$

Here (16) is the incentive compatibility constraint for A_2, (17) the relaxed *interim* participation constraint for A_2, and (18) the corresponding *ex ante* participation constraint.

Define the *interim* payoff $\hat{u}_2(\theta_2)$ for A_2 and the outside option *ex ante* payoff u by

$$\hat{u}_2(\theta_2) \equiv E_{\theta_1}[X_2(\hat{m}(\theta_1, \theta_2)) + \hat{t}_2(\theta_1, \theta_2) - \theta_2 \hat{q}_2(\theta_1, \theta_2)]$$

$$\begin{aligned} u \quad & \equiv \quad E[X_2(\theta_1, \theta_2) - \theta_2 q_2^{\alpha}(\theta_1, \theta_2)] \\ & = \quad E[q_2(\theta_1, z_{\alpha}(\theta_2)) \frac{F_2(\theta_2)}{f_2(\theta_2)}] \end{aligned} \tag{19}$$

The incentive constraint (16) is equivalent to

$$\hat{u}_2(\theta_2) = \hat{u}_2(\bar{\theta}_2) + \int_{\theta_2}^{\bar{\theta}_2} E_{\theta_1}[\hat{q}_2(\theta_1, y)]dy$$

and $E_{\theta_1}[\hat{q}_2(\theta_1, \theta_2)]$ is non-increasing in θ_2. Therefore the participation constraint (17) is equivalent to $\hat{u}_2(\bar{\theta}_2) \geq 0$. If J denotes $\max\{X_\alpha^C - X_\alpha^D, 0\}$, the relaxed problem can then be rewritten as

$$\max E[X_\alpha^D(\hat{m}(\theta)) + J - \theta_1 \hat{q}_1(\theta) - h_2(\theta_2)\hat{q}_2(\theta)] - \hat{u}_2(\bar{\theta}_2) \tag{20}$$

subject to

$$\hat{u}_2(\bar{\theta}_2) \geq 0 \tag{21}$$

$$\hat{u}_2(\bar{\theta}_2) + E[\hat{q}_2(\theta)\frac{F_2(\theta_2)}{f_2(\theta_2)}] \geq u \tag{22}$$

$$Q(\hat{q}_1(\theta), \hat{q}_2(\theta)) = q^\alpha(\hat{m}(\theta)) \tag{23}$$

and $E_{\theta_1}[\hat{q}_2(\theta_1, \theta_2)]$ is non-increasing in θ_2.

For the moment, ignore the constraint that $E_{\theta_1}[\hat{q}_2(\theta_1, \theta_2)]$ is non-increasing in θ_2: it will turn out to be automatically satisfied. If we also ignore the constraint (22) then it is clear that it will be optimal for A_1 to set $\hat{u}_2(\bar{\theta}_2) = 0$, and replicate the DS solution (where $c(\hat{m}_1(\theta), z_\alpha(\hat{m}_2(\theta))) = c(\theta_1, h_2(\theta_2))$ and $\hat{q}_i(\theta) = q_i(\theta_1, h_2(\theta_2))$). Then with $\alpha < 1$ we will have $q_2^\alpha > \hat{q}_2$ almost everywhere, and (22) will be violated. Hence (22) must bind and $\hat{u}_2(\bar{\theta}_2) = E[(h_2(\theta_2) - \theta_2)(q_2^\alpha(\theta) - \hat{q}_2(\theta))]$.

This implies that the objective function can be written as

$$\begin{aligned} &E[X_\alpha^D(\hat{m}(\theta)) + J - \theta_1 \hat{q}_1(\theta) - h_2(\theta_2)\hat{q}_2(\theta) - (h_2(\theta_2) - \theta_2)(q_2^\alpha(\theta) - \hat{q}_2(\theta))] \\ &= E[X_\alpha^D(\hat{m}(\theta)) + J - \theta_1 \hat{q}_1(\theta) - \theta_2 \hat{q}_2(\theta) - (h_2(\theta_2) - \theta_2) q_2^\alpha(\theta)] \end{aligned} \tag{24}$$

which has to be maximized subject to the constraint (besides (23)):

$$E[(h_2(\theta_2) - \theta_2)(q_2^\alpha(\theta) - \hat{q}_2(\theta))] \geq 0. \tag{25}$$

Let μ denote the multiplier corresponding to (25). Then the objective is to pointwise maximize (subject to (23) alone) in state θ:

$$X_\alpha^D(\hat{m}(\theta)) + J - \theta_1 \hat{q}_1(\theta) - z_\mu(\theta_2)\hat{q}_2(\theta) - (1 - \mu)(h_2(\theta_2) - \theta_2) q_2^\alpha(\theta) \tag{26}$$

The 'effective' *ex ante* marginal delivery cost of the final output for A_1 is now $c(\theta_1, z_\mu(\theta_2))$. So (conditional on participation in the grand contract in state θ) A_1 will select messages $\hat{m}(\theta)$ so that the output scale is determined according to

$$c(\hat{m}_1(\theta), z_\alpha(\hat{m}_2(\theta))) = c(\theta_1, z_\mu(\theta_2)) \tag{27}$$

and then allocate production assignments within the coalition: $\hat{q}_i(\theta) = q_i(\theta_1, z_\mu(\theta_2))$ in state θ.

Since (25) requires that

$$E[(h_2(\theta_2) - \theta_2)(q_2(\theta_1, z_\alpha(\theta_2)) - q_2(\theta_1, z_\mu(\theta_2)))] \geq 0, \tag{28}$$

it follows that $\mu \geq \alpha$. We now claim that $\mu = \alpha$. Otherwise $\mu > \alpha \geq 0$, implying that (25) (and therefore also (28)) must bind owing to complementary slackness. But $\mu > \alpha$ also implies that (28) is a strict inequality, a contradiction.

Since $\mu = \alpha$, the optimal messages selected in the side contract (conditional on participation in the grand contract) in state θ are $\hat{m}_1(\theta_1) = \theta_1, \hat{m}_2(\theta_2) = \theta_2$. Since $J \geq 0$, it is optimal for the coalition to participate in the grand contract in every state. This replicates the outcome of the null side contract. So the null side contract is optimal within the class of revelation side contracts that satisfy truthtelling and nonnegative interim utility constraints for A_2, and therefore also within the set of revelation side contracts that replicate the outcome of any PBE of any other side contract. This establishes that the grand contract offered by P is collusion proof, establishing part (b) of the Claim.

■

Proof of Proposition 3: The first step is to describe the outcome of side contracting among M, A_1 and A_2, following any grand contract offered by P. Specifically we need to check that the delivery cost of this coalition equals first-best cost $c(\theta)$.

Fix any state θ, and let the grand contract be denoted by $X_M(m), X_i(m), q_i(m), i = 1, 2$. Let $(m^*(\theta), \{\tilde{q}_i(\theta)\}_{i=1,2})$ denote a maximizer of

$$J(m) \equiv X_M(m) + X_1(m) + X_2(m) - \theta_1\tilde{q}_1 - \theta_2\tilde{q}_2 \tag{29}$$

subject to the constraint $Q(q_1(m), q_2(m)) = Q(\tilde{q}_1, \tilde{q}_2)$. Then it is evident $m^*(\theta)$ also maximizes

$$X_M(m) + X_1(m) + X_2(m) - c(\theta)Q(q_1(m), q_2(m)) \tag{30}$$

and the two problems have the same maximum value (denoted $J^*(\theta)$). Let $M^*(\theta)$ denote the set of messages $m^*(\theta)$ with this property.

We claim that if $J^*(\theta) \geq 0$, M and the two agents will agree to participate in state θ, and submit a joint message $m^*(\theta) \in M^*(\theta)$. And if $J^*(\theta) < 0$ they will decide to not participate in state θ. This implies that the delivery cost for the coalition is the first-best cost $c(\theta)$ in state θ.

The first step in establishing the claim is that following the mechanism offered by P, M can design a side-contract in which it is optimal for both agents to participate in all states and all stages, and conditional on their participation has a Bayesian equilibrium outcome resulting in the following expected payoff for M:

$$\begin{aligned} & E[X_M(m^*(\theta)) + X_1(m^*(\theta)) + X_2(m^*(\theta)) - c(\theta)Q(q_1(m^*(\theta)), q_2(m^*(\theta)))] \\ - & \Sigma_i E[X_i(m^B(\theta)) - \theta_i q_i(m^B(\theta))] \end{aligned} \tag{31}$$

where $m^B(.)$ denotes the Bayesian equilibrium resulting from noncooperative play of P's mechanism, when the two agents have prior beliefs about each other's cost.

Let $R_i^B(\theta_i) \equiv E_{\theta_j}[X_i(m^B(\theta)) - \theta_i q_i(m^B(\theta))|\theta_i]$ denote the interim rent that A_i earns from this equilibrium.

The side-contract that M can offer to achieve this outcome is the following. Conditional on mutually agreeing to participate in the side-contract, the three players M, A_1, A_2 will play the following revelation game. Let $\theta^M = (\theta_1^M, \theta_2^M)$ denote M's report about θ; θ_1^1 denotes agent 1's report about θ_1, and θ_2^2 denotes agent 2's report about θ_2. Also use e to denote $(\theta^M, \theta_1^1, \theta_2^2)$. The side-contract specifies production assignments $q_i^*(e)$, reports $m(e)$, and transfers $(t_1, t_2)(e)$ from M to A_1, A_2 respectively as follows.

Let $\tilde{q}_i(\theta)$ denote the cost-minimizing demand of M for q_i in state θ, i.e., the solution to minimizing $\theta_1\tilde{q}_1+\theta_2\tilde{q}_2$ subject to the constraint $Q(\tilde{q}_1, \tilde{q}_2) = Q(q_1(m^*(\theta)), q_2(m^*(\theta)))$. Then the production assignment is

$$q_i^*(e) = \tilde{q}_i(\theta_1^1, \theta_2^2).$$

The reports are

$$m(e) \equiv m^*(\theta_1^1, \theta_2^2)$$

and the side payments are

$$\begin{aligned} & t_1(e) \equiv -X_1(m^*(\theta_1^1, \theta_2^2)) + \theta_1^1 q_1^*(\theta_1^1, \theta_2^2) \\ + & \int_{\theta_1^1}^{\bar{\theta}_1} q_1^*(s, \theta_2^2) ds - \int_{\theta_1^M}^{\bar{\theta}_1} q_1^*(s, \theta_2^2) ds \\ + & R_1^B(\theta_1^M) + \delta_1(\theta_2^M, \theta_2^2) \end{aligned}$$

where $\delta_1(\theta_2^M, \theta_2^2) = 0$ if $\theta_2^M = \theta_2^2$, and K otherwise, where K is a large positive number.

$$\begin{aligned} & t_2(e) \equiv -X_2(m^*(\theta_1^1, \theta_2^2)) + \theta_2^2 q_2^*(\theta_1^1, \theta_2^2) \\ + & \int_{\theta_2^2}^{\bar{\theta}_2} q_2^*(\theta_1^1, s) ds - \int_{\theta_2^M}^{\bar{\theta}_2} q_2^*(\theta_1^1, s) ds \\ + & R_2^B(\theta_2^M) + \delta_2(\theta_1^M, \theta_1^1) \end{aligned}$$

where $\delta_2(\theta_1^M, \theta_1^1) = 0$ if $\theta_1^M = \theta_1^1$, and equal to K otherwise.

Consider the play of the side contract, conditional on mutual participation. We claim that each A_i will have a dominant strategy of reporting truthfully: $\theta_i^i = \theta_i$. Consider A_1: the argument for A_2 will be analogous. Given his true cost θ_1, agent 1's *ex post* payoff as a function of the reports is

$$\begin{aligned} & X_1(m^*(\theta_1^1, \theta_2^2)) - \theta_1 q_1^*(\theta_1^1, \theta_2^2) + t_1(\theta_1^1, \theta^M, \theta_2^2) \\ = & (\theta_1^1 - \theta_1) q_1^*(\theta_1^1, \theta_2^2) \\ + & \int_{\theta_1^1}^{\bar{\theta}_1} q_1^*(s, \theta_2^2) ds - \int_{\theta_1^M}^{\bar{\theta}_1} q_1^*(s, \theta_2^2) ds \\ + & R_1^B(\theta_1^M) + \delta_1(\theta_2^M, \theta_2^2) \end{aligned}$$

It is clear that this payoff is maximized at $\theta_1^1 = \theta_1$, irrespective of the reports θ^M, θ^2 of others, since $q_1^*(., \theta_2^2)$ is a non-increasing function.

Given that the two agents report truthfully, it follows that it is a best response for M to also report truthfully, if K is sufficiently large. Hence the revelation mechanism is incentive compatible.

In a truthful Bayesian equilibrium, type θ_1 of A_1 will end up receiving an expected payoff $R_1^B(\theta_1)$, exactly equal to his interim payoff which he gets by not participating in the side-contract (given the assumption that off-equilibrium path beliefs are the same as the prior beliefs) in the stage that A_1 observes θ_1. The ex-ante payoff $E_{\theta_1}[R_1^B(\theta_1)]$ is also equal to that obtained in his choosing non-participation at the *ex ante* stage. So it is optimal for A_1 to participate in the side-contract in all stages. The same argument applies to A_2.

The second step in establishing the claim is that (following any given mechanism offered by P) it is always optimal for M to offer a side-contract. This follows from the first step: he can always assure himself a payoff of at least (31), which is by construction at least as great as his expected payoff $E_\theta[X_M(m^B(\theta))]$ from not offering a side-contract.

The third step is to show that M cannot design a side contract which generates higher ex-ante payoff than (31). Suppose otherwise. Then there exists a side contract $(m(e), \tilde{q}_i(e), t_i(e))$ in which both agents agree to participate ex-ante, resulting in equilibrium messages $\hat{m}(\theta)$, transfers $\hat{t}_i(\theta)$ and production assignments $\hat{q}_i(\theta)$, and *ex ante* payoff for M in the equilibrium

$$E_\theta[X_M(\hat{m}(\theta)) - \sum_{i=1}^{2} \hat{t}_i(\theta)] \tag{32}$$

which is strictly larger than (31). Since both agents participate ex-ante in this side contract, A_i's ex-ante payoff in the equilibrium:

$$E_\theta[X_i(\hat{m}(\theta)) + \hat{t}_i(\theta) - \theta_i \hat{q}_i(\theta)], \tag{33}$$

is at least as large as $E_{\theta_i}[R_i^B(\theta_i)]$ for each $i = 1, 2$. This implies that the ex-ante joint payoff in the equilibrium must be strictly larger than $E_\theta[J^*(\theta)]$. This contradicts the definition of $J^*(\theta)$, establishing the third step.

Finally, the second and the third steps imply that the side-contract described above is optimal for M. Moreover, if the selected message in state θ is not in the set $M^*(\theta)$, then applying the reasoning of the third step we would conclude that M obtains a payoff lower than (31). Hence M must offer a side-contract which results in a message from the state $M^*(\theta)$ in state θ, generates *ex ante* payoff equal to (31) for himself and $E_{\theta_i}[R_i^B(\theta_i)]$ for A_i, i.e., a joint payoff of $J^*(\theta)$. ■

Proof of Proposition 4: This result follows from Proposition 1, since it can easily be verified that the unit cost of A_1 $c(\theta_1, h_2(\theta_2))$ in DS has a continuous positive density throughout its support, and is first-order stochastically dominated by the unit cost $c(\theta_1, \theta_2)$ of M in DM. ■

Proof of Lemma 1. We first show that in case (1), the optimal production assignments in DM will be substitutes. If $c + G(c)/g(c)$ is non-decreasing in c, P's expected payoff is represented by $E_c[\pi(H(c))]$ where $\pi(H) \equiv \max_q[V(q) - Hq]$ and

$$H(c) \equiv c + \frac{G(c)}{g(c)}.$$

The input level of A_i is $q_i^{DM}(\theta_i, \theta_j) = q(H(c(\theta_i, \theta_j)))c_i(\theta_i, \theta_j)$ where $q(H) = \arg\max_q[V(q) - Hq]$.[19] The resulting assignments are substitutes if

$$\begin{aligned} & \partial q_i(\theta_i, \theta_j)/\partial\theta_j = \{\partial^2 c(\theta_i, \theta_j)/\partial\theta_i\partial\theta_j\}q(H(c(\theta_i, \theta_j))) \\ + \quad & \{\partial c(\theta_i, \theta_j)/\partial\theta_i\}\{\partial c(\theta_i, \theta_j)/\partial\theta_j\}q^{'}(H(c(\theta_i, \theta_j)))H^{'}(c(\theta_i, \theta_j)) > 0, \end{aligned}$$

for any (θ_1, θ_2). With a CES function $Q(q_1, q_2) = (q_1^\alpha + q_2^\alpha)^{1/\alpha}$, this condition is satisfied when the elasticity of substitution $\frac{1}{1-\alpha}$ is sufficiently large:

$$-[H(c)q^{'}(H(c))/q(H(c))][cH^{'}(c)/H(c)] \leq \frac{1}{1-\alpha}.$$

Clearly the substitute property is also satisfied with infinite elasticity of substitution, so the result holds in case (1a) as well.

Now turn to the complementarity property in C. Since the optimal assignments in C maximize

$$V(Q(q_1, q_2)) - \sum_{i=1}^{2} h(\theta_i)q_i$$

it follows that they will be complements if

$$\frac{\partial^2 V(Q(q_1, q_2))}{\partial q_1 \partial q_2} > 0.$$

■

Proof of Proposition 5:

Note first that given any nonincreasing function $\hat{q}(c)$ representing the scale of output at different unit cost levels, and corresponding cost-minimizing production assignments $\hat{q}_i(\theta) \equiv [\partial c(\theta_i, \theta_j)/\partial\theta_i]\hat{q}(c(\theta))$, these are implementable in DM and C (by some set of payment functions). Implementability in C is obvious because $\hat{q}_i$ is nonincreasing in θ_i. In DM it is implementable because production assignments are always selected in a cost-minimizing fashion by M given the scale of output, and the latter is implementable because $\hat{q}(c(\theta))$ is nonincreasing.

[19] It is difficult to verify whether or not $H(c)$ is a nondecreasing function in general, given only our assumption of the monotonicity of the hazard rate of θ_i. However, this property does hold in the case where the elasticity of substitution is infinite or zero. For the case of perfect complementarity, see Eaton (1987).

Next we shall show that any such production assignment can be implemented more cheaply in C (resp. DM) if the input demands are substitutes (resp. complements), i.e., $\hat{q}_2$ is increasing (resp. decreasing) in θ_1.

The minimum cost to P of implementing any such production assignments in DM equals

$$E[X_M(m(\theta))] = E[c(\theta)\hat{q}(c(\theta)) + \int_{c(\theta)}^{c(\bar{\theta})} \hat{q}(c)dc]$$

where $\bar{\theta}$ denotes $(\bar{\theta}_1, \bar{\theta}_2)$. In C, it equals

$$\Sigma_i E[\theta_i \hat{q}_i(\theta) + \int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(\theta, \theta_j)d\theta].$$

Since $\Sigma_i \theta_i \hat{q}_i(\theta) = c(\theta)\hat{q}(c(\theta))$, it follows that

$$E[X_M(m(\theta))] - \Sigma_i E[X_i(m(\theta))] = E[\int_{c(\theta)}^{c(\bar{\theta})} \hat{q}(c)dc - \Sigma_i \int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(\theta, \theta_j)d\theta]$$

Now note that

$$\int_{c(\theta)}^{c(\bar{\theta})} \hat{q}(c)dc = \int_{c(\theta_1,\theta_2)}^{c(\bar{\theta}_1,\theta_2)} \hat{q}(c)dc + \int_{c(\bar{\theta}_1,\theta_2)}^{c(\bar{\theta}_1,\bar{\theta}_2)} \hat{q}(c)dc \tag{34}$$

Take the first integral on the right-hand-side of (34); by a change of variable from c to θ_1:

$$\begin{aligned} \int_{c(\theta_1,\theta_2)}^{c(\bar{\theta}_1,\theta_2)} \hat{q}(c)dc &= \int_{\theta_1}^{\bar{\theta}_1} \hat{q}(c(\theta, \theta_2)) \frac{\partial c(\theta, \theta_2)}{\partial \theta} d\theta \\ &= \int_{\theta_1}^{\bar{\theta}_1} \hat{q}_1(\theta, \theta_2)d\theta, \end{aligned}$$

upon using the definition of cost-minimizing assignments. Using the same manipulation of the second integral on the right hand side of (34), the minimum rent that need to be paid in DM is

$$\int_{c(\theta)}^{c(\bar{\theta})} \hat{q}(c)dc = \int_{\theta_1}^{\bar{\theta}_1} \hat{q}_1(\theta, \theta_2)d\theta + \int_{\theta_2}^{\bar{\theta}_2} \hat{q}_2(\bar{\theta}_1, \theta)d\theta \tag{35}$$

which can be compared with the minimum rents in C:

$$\int_{\theta_1}^{\bar{\theta}_1} \hat{q}_1(\theta, \theta_2)d\theta + \int_{\theta_2}^{\bar{\theta}_2} \hat{q}_2(\theta_1, \theta)d\theta \tag{36}$$

Comparing (35) with (36), DM is more costly if and only if

$$\int_{\theta_2}^{\bar{\theta}_2} [\hat{q}_2(\bar{\theta}_1, \theta) - \hat{q}_2(\theta_1, \theta)]d\theta > 0 \tag{37}$$

which is true if the input demands are substitutes, while the reverse is true if they are complements. That C dominates DM in the substitute case now follows. In the complements case, and under the additional assumption that θ_1, θ_2 are identically and exponentially distributed, with a lower bound for the support equal to 0, the hazard rates $h_i(\theta_i)$ are linear and identical for both suppliers. This implies that relative virtual costs equal relative costs, so the optimal production assignments in C are cost-minimizing, and the scale of output is a function only of the unit cost $c(\theta)$. Hence the production assignments can be implemented in DM; the argument above establishes the cost of implementing them in DM would be lower. ■

Proof of Proposition 6: It suffices to show that CC dominates DM in the substitutes case. It is easy to show in this case that the expected profit of P in CC is at least as high as in DM, by adapting the proof of Proposition 2. Take $q(c)$ to be the optimal output function in DM (instead of DS), and select $\alpha = 0$. Then the argument of Proposition 2 ensures that P can implement the optimal output function in DM in CC at an expected cost of $\max\{X_0^D, X_0^C\}$. Since the two inputs are substitutes, we have $X_0^C \leq X_0^D$. So it can be implemented in CC at an expected cost of X_0^D. This is the same as the cost when P contracts for delivery of the output with a single agent who has a unit cost of $c(\theta_1, z_0(\theta_2)) = c(\theta)$ in state θ, which is exactly the same as in DM. Hence P can achieve the same profit in CC as in DM.

To establish that P can achieve a strictly higher profit in CC, however, this argument cannot be modified: we need an entirely different construction of a grand contract, and the assumption that agents have passive beliefs off the equilibrium path.

Consider the optimal output function in DM $q(c)$, and a continuously differentiable non-decreasing $\Lambda(\theta_2)$ defined over $[\underline{\theta}_2, \bar{\theta}_2]$ satisfying $0 \leq \Lambda(\theta_2) \leq 1$. Define

$$z(\theta_2) = \theta_2 + \frac{F_2(\theta_2) - \Lambda(\theta_2)}{f_2(\theta_2)}. \tag{38}$$

Extend $q(.)$ for c below $c(\underline{\theta}_1, \underline{\theta}_2)$ by setting it equal to $c(\underline{\theta}_1, \underline{\theta}_2)$ at all such values.

We shall show that there exists a nondecreasing $z(.)$ function of this form with the property that:

(a) $z(\theta_2) \leq \theta_2$ for all θ_2, with strict inequality holding over a set of positive measure, and

(b) P can implement $q(.)$ in CC at a cost $X_z^D \equiv E[X_z^D(\theta)]$ where

$$X_z^D(\theta) \equiv c(\theta_1, z(\theta_2))q(c(\theta_1, z(\theta_2))) + \int_{c(\theta_1, z(\theta_2))}^{c(\bar{\theta}_1, z(\bar{\theta}_2))} q(c)dc.$$

Since this is exactly the outcome where P contracts with a single agent with unit cost $c(\theta_1, z(\theta_2))$ and implements output function $q(c)$, it follows from Proposition 1 that P achieves a higher expected profit in CC compared with DM.

We first establish (b) for any nondecreasing $z(.)$ function satisfying (a), (38) and a supplemental condition specified below. Then we shall verify that such a $z(.)$ function can indeed be constructed in the case of substitutes.

Consider the following grand contract where A_2's message space is enlarged to include a decision which of two noncooperative games (N,C) to play, apart from a participation decision and type report: $(X_1(\tilde{m}), X_2(\tilde{m}), q_1^z(\tilde{m}), q_2^z(\tilde{m}))$, where $\tilde{m} = (\tilde{m}_1, \tilde{m}_2)$, $\tilde{m}_1 \in \tilde{M}_1 \equiv [\underline{\theta}_1, \bar{\theta}_1] \cup Exit$ and $\tilde{m}_2 \in \tilde{M}_2 \equiv \{(\theta_2, N) \mid \theta_2 \in [\underline{\theta}_2, \bar{\theta}_2]\} \cup \{(\theta_2, C) \mid \theta_2 \in [\underline{\theta}_2, \bar{\theta}_2]\} \cup Exit$. If either player decides to exit, $X_1 = X_2 = q_1^z = q_2^z = 0$. Otherwise the noncooperative game chosen by A_2 is played between the two agents. If A_2 selected C, the game is

$$X_1(\theta_1, (\theta_2, C)) = X_z^D(\theta)$$

$$X_2(\theta_1, (\theta_2, C)) = 0.$$

$$q_1^z(\theta_1, (\theta_2, C)) = q_1(\theta_1, z(\theta_2)) \equiv c_1(\theta_1, z(\theta_2))q(c(\theta_1, z(\theta_2)))$$

$$q_2^z(\theta_1, (\theta_2, C)) = q_2(\theta_1, z(\theta_2)) \equiv c_2(\theta_1, z(\theta_2))q(c(\theta_1, z(\theta_2)))$$

If instead A_2 selected N, the game is

$$X_1(\theta_1, (\theta_2, N)) \equiv X_1^N(\theta) = \theta_1 q_1(\theta_1, z(\theta_2)) + \int_{\theta_1}^{\bar{\theta}_1} q_1(\theta, z(\theta_2))d\theta$$

$$X_2(\theta_1, (\theta_2, N)) \equiv X_2^N(\theta) = \theta_2 q_2(\theta_1, z(\theta_2)) + \int_{\theta_2}^{\bar{\theta}_2} E_{\theta_1}[q_2(\theta_1, z(\theta))]d\theta.$$

$$q_1^z(\theta_1, (\theta_2, N)) = q_1(\theta_1, z(\theta_2))$$

$$q_2^z(\theta_1, (\theta_2, N)) = q_2(\theta_1, z(\theta_2))$$

So only the payment rules are modified between games N and C.

Let us first consider the outcomes of noncooperative play of the grand contract. Note first that selection of game C is a strictly dominated strategy for all types of A_2: it involves positive production assignments and no payments. So there cannot be any Bayesian equilibrium where A_2 selects the game C. Conditional on this it is a dominant strategy for every type of A_1 to agree to participate and report truthfully. And (given $z(.)$ is nondecreasing) conditional on A_1 agreeing to participate and reporting truthfully, and A_2 holding prior beliefs over A_1's type, it is a best response for A_2 to participate, select N and report truthfully. Hence if A_2 does not update her prior beliefs, there is a unique Bayesian equilibrium (which survives iterated elimination of dominated strategies) where A_2 obtains an interim payoff of[20]

$$u_2(\theta_2) \equiv \int_{\theta_2}^{\bar{\theta}_2} E_{\theta_1}[q_2(\theta_1, z(\theta))]d\theta.$$

[20]There is also a Bayesian equilibrium where both players decide never to participate, which does not survive iterated elimination of dominated strategies. A slight modification of the payment function can costlessly eliminate this equilibrium, e.g., where one player is given a small reward for being the only one electing to participate.

Moreover, note that even with a different set of beliefs, A_2 can always guarantee herself at least this interim payoff in any Bayesian equilibrium by selecting N and reporting truthfully. It follows therefore that irrespective of how the agents may update their beliefs upon observing interim exit/participation decisions from any side-contract offered by A_1, $u_2(\theta_2)$ is a lower bound to A_2's outside options from refusing to participate in the side-contract.

Claim. Assume that the agents have passive beliefs off the equilibrium path. Then if P offers a grand contract of the above form which satisfies the condition that

$$X_z^D(\theta) \geq X_1^N(\theta) + X_2^N(\theta)$$

for all θ, it is optimal for A_1 to offer the following side-contract: (in every state) the agents decide to participate in the grand contract, report costs truthfully, A_2 selects the game C, and A_1 pays a transfer $\hat{t}_2^C(\theta)$ to A_2 satisfying

$$E_{\theta_1}[\hat{t}_2^C(\theta)] = u_2(\theta_2) + E_{\theta_1}[\theta_2 q_2(\theta_1, z(\theta_2))].$$

This grand contract implements $q(c(\theta_1, z(\theta_2)))$ as a Perfect Bayesian Equilibrium (PBE) in CC at an expected cost of X_z^D.

The proof of this Claim is in two steps. We first establish an upper bound to A_1's expected payoff from any side-contract, and then show that the side contract described in the Claim achieves this upper bound under the assumption of passive beliefs.

Using Lemma 2, the outcome of an arbitrary side contract can be replicated by a revelation side contract which is accepted by both agents at the interim stage, and in which they report truthfully. Consider any such revelation side contract. Let $i(\theta) \in \{0, 1\}$ denote choice of game N, C, with $i = 0$ denoting choice of N. The outcome of any revelation side contract can be denoted as follows:

$$(i(\theta); (\hat{m}^C(\theta), \hat{q}_1^C(\theta), \hat{q}_2^C(\theta), \hat{t}_2^C(\theta)); (\hat{m}^N(\theta), \hat{q}_1^N(\theta), \hat{q}_2^N(\theta), \hat{t}_2^N(\theta)))$$

satisfying the restrictions $Q(\hat{q}_1^C(\theta), \hat{q}_2^C(\theta)) = Q(q^z(\hat{m}^C(\theta)))$ and $Q(\hat{q}_1^N(\theta), \hat{q}_2^N(\theta)) = Q(q^z(\hat{m}^N(\theta)))$. Conditional on $i(\theta) = 1$, $(\hat{m}^C(\theta), \hat{q}_1^C(\theta), \hat{q}_2^C(\theta), \hat{t}_2^C(\theta))$ respectively denotes their coalitional report to P, input reassignments and transfer from A_1 to A_2. On the other hand, $i(\theta) = 0$ means they choose game N in state θ, and then select $(\hat{m}^N(\theta), \hat{q}_1^N(\theta), \hat{q}_2^N(\theta), \hat{t}_2^N(\theta))$.

Let the resulting ex-post realizations of payoff of A_2, production of A_1 and A_2, and total payment received from P be denoted respectively by $\hat{u}_2(\theta)$, $\bar{q}_1(\theta)$, $\bar{q}_2(\theta)$ and $\bar{X}(\theta)$:

$$\begin{aligned} \hat{u}_2(\theta) \;\; &\equiv \;\; i(\theta)\{\hat{t}_2^C(\theta) - \theta_2 \hat{q}_2^C(\theta)\} \\ &+ \;\; (1 - i(\theta))\{\hat{t}_2^N(\theta) + X_2^N(\hat{m}^N(\theta)) - \theta_2 \hat{q}_2^N(\theta)\} \end{aligned}$$

$$\bar{q}_1(\theta) \equiv i(\theta)\hat{q}_1^C(\theta) + (1 - i(\theta))\hat{q}_1^N(\theta)$$

$$\bar{q}_2(\theta) \equiv i(\theta)\hat{q}_2^C(\theta) + (1 - i(\theta))\hat{q}_2^N(\theta)$$

$$\bar{X}(\theta) \equiv i(\theta) X_z^D(\hat{m}^C(\theta)) + (1 - i(\theta))\{X_1^N(\hat{m}^N(\theta)) + X_2^N(\hat{m}^N(\theta))\}$$

The resulting ex-post payoff $\hat{u}_1(\theta)$ of A_1 is denoted by

$$\hat{u}_1(\theta) \equiv \bar{X}(\theta) - \theta_1 \bar{q}_1(\theta) - \theta_2 \bar{q}_2(\theta) - \hat{u}_2(\theta).$$

Since both agents decide to participate in the side contract and report each others costs truthfully, the following relationships must hold:

$$E_{\theta_1}[\hat{u}_2(\theta_1, \theta_2)] = E_{\theta_1}[\hat{u}_2(\theta_1, \bar{\theta}_2)] + \int_{\theta_2}^{\bar{\theta}_2} E_{\theta_1}[\bar{q}_2(\theta_1, \theta)] d\theta$$

$$E_{\theta_1}[\bar{q}_2(\theta)] \quad \text{is non-increasing in } \theta_2$$

$$E_{\theta_1}[\hat{u}_2(\theta_1, \theta_2)] \geq u_2(\theta_2)$$

The third one follows from the fact that A_2 can guarantee herself an interim payoff of $u_2(\theta_2)$ in non-cooperative play of the grand contract.

A_1's ex-ante payoff from this side-contract is

$$E[\bar{X}(\theta) - \theta_1 \bar{q}_1(\theta) - h_2(\theta_2) \bar{q}_2(\theta) - \hat{u}_2(\theta_1, \bar{\theta}_2)]. \tag{39}$$

Now by definition of $z(.)$ we have $h_2(\theta_2) = z(\theta_2) + \frac{\Lambda(\theta_2)}{f_2(\theta_2)}$. Hence (39) can be written as

$$E[\bar{X}(\theta) - \theta_1 \bar{q}_1(\theta) - z(\theta_2) \bar{q}_2(\theta)] - \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[\bar{q}_2(\theta)] d\theta_2 - E[\hat{u}_2(\theta_1, \bar{\theta}_2)]. \tag{40}$$

Integration by parts and the fact that $u_2(\bar{\theta}_2) = 0$ implies that

$$\begin{aligned}
\int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[\bar{q}_2(\theta)] d\theta_2 &= \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2)[E_{\theta_1} \bar{q}_2(\theta) - E_{\theta_1} q_2(\theta_1, z(\theta_2))] d\theta_2 \\
&\quad + \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[q_2(\theta_1, z(\theta_2))] d\theta_2 \\
&= -\Lambda(\bar{\theta}_2) E_{\theta_1} \hat{u}_2(\theta_1, \bar{\theta}_2) + \Lambda(\underline{\theta}_2)[E_{\theta_1} \hat{u}_2(\theta_1, \underline{\theta}_2) - u_2(\underline{\theta}_2)] \\
&\quad + \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda'(\theta_2)[E_{\theta_1} \hat{u}_2(\theta) - u_2(\theta_2)] d\theta_2 \\
&\quad + \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[q_2(\theta_1, z(\theta_2))] d\theta_2
\end{aligned}$$

Hence (39) can be rewritten as

$$\begin{aligned}
&E[\bar{X}(\theta) - \theta_1 \bar{q}_1(\theta) - z(\theta_2) \bar{q}_2(\theta)] - (1 - \Lambda(\bar{\theta}_2)) E_{\theta_1} \hat{u}_2(\theta_1, \bar{\theta}_2) \\
&- \Lambda(\underline{\theta}_2)[E_{\theta_1} \hat{u}_2(\theta_1, \underline{\theta}_2) - u_2(\underline{\theta}_2)] \\
&- \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda'(\theta_2)[E_{\theta_1} \hat{u}_2(\theta) - u_2(\theta_2)] d\theta_2 - \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[q_2(\theta_1, z(\theta_2))] d\theta_2.
\end{aligned} \tag{41}$$

Since $u_2(\theta_2)$ is a lower bound to A_2's outside option payoff, (41) is maximized among all incentive compatible revelation side contracts satisfying the interim participation constraint when $E_{\theta_1}[\hat{u}_2(\theta_1,\theta_2)] = u_2(\theta_2)$, given the properties of $\Lambda(\theta_2)$. Moreover, given that $X_z^D(m) \geq X_1^N(m_1)+X_2^N(m_2)$ for all m_1, m_2 while production assignments are unaltered by selection of game N rather than C, it follows that (41) is maximized when $i(\theta)=1$ for all θ. It is evident from (41) that it is optimal for A_1 to set $\hat{m}^C(\theta)=\theta$ and $\hat{q}_i^C(\theta) = q_i(\theta_1, z(\theta_2))$. The resulting upper bound to A_1's ex ante profit is

$$\begin{aligned} & E[X_z^D(\theta) - c(\theta_1, z(\theta_2))q(c(\theta_1, z(\theta_2)))] \\ - & \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[q_2(\theta_1, z(\theta_2))] d\theta_2 \\ = & E[\int_{\theta_1}^{\bar{\theta}_1} q_1(y, z(\theta_2))dy + \int_{z(\theta_2)}^{z(\bar{\theta}_2)} q_2(\bar{\theta}_1, z)dz] \\ & - \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[q_2(\theta_1, z(\theta_2))] d\theta_2. \end{aligned}$$

Now note that the definition of $\Lambda(.)$ implies

$$\int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[q_2(\theta_1, z(\theta_2))] d\theta_2 = E[(h_2(\theta_2) - z(\theta_2)) q_2(\theta_1, z(\theta_2))].$$

Consider the following transfers that implement production assignments $q_i(\theta_1, z(\theta_2))$ in C:

$$X_z^C(\theta) \equiv \theta_1 q_1(\theta_1, z(\theta_2)) + \theta_2 q_2(\theta_1, z(\theta_2)) + \int_{\theta_1}^{\bar{\theta}_1} q_1(\theta, z(\theta_2)) d\theta + \int_{\theta_2}^{\bar{\theta}_2} q_2(\theta_1, z(\theta)) d\theta.$$

and let

$$X_z^C \equiv E[X_z^C(\theta)]$$

denote the associated expected cost of these transfers. Then it follows that

$$\begin{aligned} X_z^D - X_z^C &= E[(z(\theta_2) - \theta_2) q_2(\theta_1, z(\theta_2)) + \int_{z(\theta_2)}^{z(\bar{\theta}_2)} q_2(\bar{\theta}_1, z)dz - \int_{\theta_2}^{\bar{\theta}_2} q_2(\theta_1, z(y))dy] \\ &= E[\int_{z(\theta_2)}^{z(\bar{\theta}_2)} q_2(\bar{\theta}_1, z)dz] - \int_{\underline{\theta}_2}^{\bar{\theta}_2} \Lambda(\theta_2) E_{\theta_1}[q_2(\theta_1, z(\theta_2))] d\theta_2. \end{aligned}$$

Therefore the upper bound to A_1's ex ante profit from a side contract can be expressed as

$$E[\int_{\theta_1}^{\bar{\theta}_1} q_1(y, z(\theta_2))dy] + X_z^D - X_z^C. \quad (42)$$

Next we show that there is a side-contract of the kind described in the Claim which realizes this upper bound payoff in a PBE if beliefs are passive. Consider the following side-contract: $i(\theta) = 1$, $\hat{m}^C(\theta) = \theta$, $\hat{q}_i^C(\theta) = q_i(\theta_1, z(\theta_2))$ and

$$\begin{aligned} \hat{t}_2^C(\theta) &= \theta_2 q_2(\theta_1, z(\theta_2)) + \int_{\theta_2}^{\bar{\theta}_2} q_2(\theta_1, z(\theta)) d\theta \\ &+ [X_z^D(\theta) - X_z^C(\theta)] - E_{\theta_1}[X_z^D(\theta) - X_z^C(\theta)]. \end{aligned}$$

If the agents do not update their prior beliefs then conditional on playing the side contract it is easily checked that they report truthfully, since

$$E_{\theta_1}[\hat{t}_2^C(\theta)] = E_{\theta_1}[\theta_2 q_2(\theta_1, z(\theta_2)) + \int_{\theta_2}^{\bar{\theta}_2} q_2(\theta_1, z(\theta)) d\theta].$$

and

$$E_{\theta_2}[X_z^D(\theta) - \hat{t}_2^C(\theta)] = E_{\theta_2}[\theta_1 q_1(\theta_1, z(\theta_2)) + \int_{\theta_1}^{\bar{\theta}_1} q_1(y, z(\theta_2)) dy] + X_z^D - X_z^C.$$

On the other hand if they play the grand contract noncooperatively with their prior beliefs then A_2 would obtain exactly the same interim payoff. And A_1 would obtain an interim payoff of

$$E_{\theta_2}[\int_{\theta_1}^{\bar{\theta}_1} q_1(y, z(\theta_2)) dy]$$

which is less than the payoff from playing the side contract because

$$X_z^D - X_z^C = E[X_z^D(\theta) - \{X_1^N(\theta) + X_2^N(\theta)\}] \geq 0.$$

It follows that interim participation (and truthful reporting) in the side contract constitutes a PBE with passive beliefs, which concludes the proof of the Claim.

To complete the proof, we have to show that there exists a $z(.)$ function satisfying (a), (38) and the property that

$$X_z^D(\theta) \geq X_1^N(\theta) + X_2^N(\theta)$$

for all θ, if the two inputs are substitutes. Note first that the substitutes assumption implies

$$\begin{aligned} & X_z^D(\theta) - [X_1^N(\theta) + X_2^N(\theta)] \\ = \;& (z(\theta_2) - \theta_2) q_2(\theta_1, z(\theta_2)) + \int_{z(\theta_2)}^{z(\bar{\theta}_2)} q_2(\bar{\theta}_1, z) dz - \int_{\theta_2}^{\bar{\theta}_2} E_{\theta_1}[q_2(\theta_1, z(\theta))] d\theta \\ \geq \;& (z(\theta_2) - \theta_2) q_2(\bar{\theta}_1, z(\theta_2)) + \int_{z(\theta_2)}^{z(\bar{\theta}_2)} q_2(\bar{\theta}_1, z) dz - \int_{\theta_2}^{\bar{\theta}_2} E_{\theta_1}[q_2(\theta_1, z(\theta))] d\theta. \end{aligned}$$

Therefore it suffices to show that

$$(z(\theta_2) - \theta_2) q_2(\bar{\theta}_1, z(\theta_2)) + \int_{z(\theta_2)}^{z(\bar{\theta}_2)} q_2(\bar{\theta}_1, z) dz - \int_{\theta_2}^{\bar{\theta}_2} E_{\theta_1}[q_2(\theta_1, z(\theta))] d\theta \geq 0. \quad (43)$$

Define $\Phi(\theta_2)$ by

$$\Phi(\theta_2) \equiv \int_{\theta_2}^{\bar{\theta_2}} [q_2(\bar{\theta}_1, y) - E_{\theta_1} q_2(\theta_1, y)]dy.$$

Since the two inputs are substitutes, Φ is a nonnegative, nonincreasing continuous function with $\Phi(\bar{\theta}_2) = 0$ and $\Phi(\underline{\theta}_2) > 0$. So we can select $\hat{\theta}_2 \in (\bar{\theta}_2, \underline{\theta}_2)$ such that $\Phi(\hat{\theta}_2) > 0$.

Next select $\tilde{\theta}_2 \in [\underline{\theta}_2, \hat{\theta}_2)$ such that the density f_2 is everywhere positive on $I \equiv [\tilde{\theta}_2, \hat{\theta}_2]$. Let $L > 0$ denote the minimum value of f_2 on I. Take a sufficiently small $\gamma > 0$ so that $\tilde{\theta}_2 - \gamma/f_2(\tilde{\theta}_2) < \hat{\theta}_2 - \gamma/f_2(\hat{\theta}_2)$. Then since $\theta_2 - \gamma/f_2(\theta_2)$ is continuous, there exists $\tilde{I} \equiv [\theta_2^{'}, \theta_2^{''}] \subset I$ so that $\theta_2 - \gamma/f_2(\theta_2)$ is increasing in θ_2 on $\tilde{I}$.

Choose a continuously differentiable function $g : [\underline{\theta}_2, \hat{\theta}_2] \to [0, 1]$, with a uniformly bounded derivative and the following properties: $g(\theta_2) = g'(\theta_2) = 0$ for all $\theta_2 \in [\underline{\theta}_2, \theta_2^{'}] \cup [\theta_2^{''}, \hat{\theta}_2]$, while $g(\theta_2)$ is positive for every θ_2 in the interior of $\tilde{I}$. Let R denote the minimum value of $g'(\theta_2)$ over $\tilde{I}$. Since g must be somewhere strictly decreasing over $\tilde{I}$, R is a negative number.

Define $z(\theta_2) = \theta_2$ for all $\theta_2 \in (\hat{\theta}_2, \bar{\theta}_2]$, and $z(\theta_2) = \theta_2 - \epsilon \frac{g(\theta_2)}{f_2(\theta_2)}$ for $\theta_2 \leq \hat{\theta}_2$, where ϵ is a positive number satisfying the following properties (with $\bar{f}$ and $\bar{g}$ respectively denoting the maximum value of f_2 and g over $\tilde{I}$, and M denoting an upper bound to the rate of change of g)[21]:

(i) $\epsilon < \frac{L^2}{M\bar{f} + \bar{g}\bar{f}^2/\gamma}$

(ii) $\epsilon < \frac{L}{-R}$

(iii) $\epsilon Q + S(\epsilon) < \Phi(\hat{\theta}_2)$

where

$$Q \equiv \max_{\theta_2 \in \tilde{I}} [\frac{q_2(\bar{\theta}_1, z(\theta_2))}{f_2(\theta_2)}]$$

and

$$S(\epsilon) \equiv \int_{\theta_2^{'}}^{\theta_2^{''}} [q_2(\bar{\theta}_1, y - \frac{\epsilon}{f_2(y)}) - q_2(\bar{\theta}_1, y)]dy.$$

Since $S(\epsilon)$ tends to 0 as ϵ tends to 0, it is evident that such an ϵ can be chosen. Property (i) implies that z is nondecreasing. This follows from the following argument. We know that for $\theta_2, \theta_2 + h \in \tilde{I}$,

$$\begin{aligned} 0 &< [\theta_2 + h - \frac{\gamma}{f_2(\theta_2 + h)}] - [\theta_2 - \frac{\gamma}{f_2(\theta_2)}] \\ &= \frac{\gamma h}{f_2(\theta_2 + h) f_2(\theta_2)} [\frac{f_2(\theta_2 + h) f_2(\theta_2)}{\gamma} + \frac{f_2(\theta_2 + h) - f_2(\theta_2)}{h}] \\ &\leq \frac{\gamma h}{f_2(\theta_2 + h) f_2(\theta_2)} [\bar{f}^2/\gamma + \frac{f_2(\theta_2 + h) - f_2(\theta_2)}{h}], \end{aligned}$$

[21] In other words, $|g(\theta_2 + h) - g(\theta_2)| < Mh$ for all $\theta_2, \theta_2 + h \in \tilde{I}$.

implying that $f_2(\theta_2 + h) > f_2(\theta_2) - h\bar{f}^2/\gamma$, and by definition of M,

$$g(\theta_2 + h) < g(\theta_2) + Mh$$

which implies that

$$\begin{aligned} & [\theta_2 + h - \epsilon\frac{g(\theta_2 + h)}{f_2(\theta_2 + h)}] - [\theta_2 - \epsilon\frac{g(\theta_2)}{f_2(\theta_2)}] = h[1 - \epsilon\frac{g(\theta_2 + h)f_2(\theta_2) - g(\theta_2)f_2(\theta_2 + h)}{hf_2(\theta_2)f_2(\theta_2 + h)}] \\ > \quad & h[1 - \epsilon\frac{M\bar{f} + \bar{g}\bar{f}^2/\gamma}{L^2}], \end{aligned}$$

which is positive owing to property (i). The associated $\Lambda(\theta_2)$ equals $F_2(\theta_2)$ above $\hat{\theta}_2$ and $F_2(\theta_2) + \epsilon g(\theta_2)$ below $\hat{\theta}_2$. The continuous differentiability of g implies the same for Λ. Property (ii) implies that Λ is nondecreasing. Since $\Lambda(\bar{\theta}_2) \leq 1$, this implies that $\Lambda(\theta_2) \leq 1$ for all θ_2. So Λ satisfies all the required properties.

For this choice of $z(.)$, (43) is automatically satisfied for any $\theta_2 \geq \hat{\theta}_2$ by virtue of the substitutes assumption. For any $\theta_2 < \hat{\theta}_2$, it requires

$$-\epsilon\frac{g(\theta_2)q_2(\bar{\theta}_1, z(\theta_2))}{f_2(\theta_2)} + \int_{z(\theta_2)}^{\hat{\theta}_2} q_2(\bar{\theta}_1, z)dz - \int_{\theta_2}^{\hat{\theta}_2} E_{\theta_1}q_2(\theta_1, z(y))dy + \Phi(\hat{\theta}_2) > 0. \quad (44)$$

This is ensured by property (iii) and the substitutes assumption, since for any $\theta_2 < \hat{\theta}_2$:

$$\begin{aligned} & \int_{z(\theta_2)}^{\hat{\theta}_2} q_2(\bar{\theta}_1, z)dz - \int_{\theta_2}^{\hat{\theta}_2} E_{\theta_1}q_2(\theta_1, z(y))dy \\ = \quad & \int_{z(\theta_2)}^{\theta_2} q_2(\bar{\theta}_1, z)dz + \int_{\theta_2}^{\hat{\theta}_2} q_2(\bar{\theta}_1, y)dy - \int_{\theta_2}^{\hat{\theta}_2} E_{\theta_1}q_2(\theta_1, z(y))dy \\ \geq \quad & \int_{z(\theta_2)}^{\theta_2} q_2(\bar{\theta}_1, z)dz - S(\epsilon) + \int_{\theta_2}^{\hat{\theta}_2} [q_2(\bar{\theta}_1, z(y)) - E_{\theta_1}q_2(\theta_1, z(y))]dy \\ \geq \quad & -S(\epsilon). \end{aligned}$$

This concludes the proof. ■