Vol. 26, No. 4, Winter 1995
pp. 654–672

Hierarchical decentralization of incentive contracts

Nahum D. Melumad*

Dilip Mookherjee**

and

Stefan Reichelstein***

Agents in a hierarchy are commonly delegated authority to communicate and contract with agents at lower levels. While delegation reduces the burden of communication and information processing on the principal, it also introduces additional incentive problems. We find that with sufficient monitoring of the agents' contributions to joint production, and a particular sequence of contracting, the additional incentive problems inherent in delegation can be completely resolved. These conditions are generally also necessary for delegation to achieve second-best results.

1. Introduction

■ Delegation of authority is commonly observed within economic organizations. Owners of a firm usually delegate to top managers the responsibility for organizing production within the firm, employing subordinates, and contracting with external suppliers. These managers, in turn, hierarchically delegate responsibility to divisional managers, and so forth. In the context of procurement, purchasers of a final product often contract with a single supplier, who in turn is authorized to organize production and procure necessary inputs from subcontractors. One alternative to such patterns of delegation would be a centralized arrangement in which the owners (or purchasers) retain authority over all decisions and contract directly with all employees (or suppliers). The purpose of this article is to compare the performance of alternative delegation mechanisms with centralized ones.

* Columbia University.

** Boston University and Indian Statistical Institute, New Delhi.

*** University of California, Berkeley, and Universität Wien.

We received helpful comments and suggestions from Jacques Crémer, Joel Demski, Ron Dye, Rody Manuelli, Preston McAfee, Sherwin Rosen, David Sappington, Joseph Stiglitz, Timothy Van Zandt, two anonymous referees, and Michael Whinston (the Editor). We are also grateful to seminar participants at Berkeley, Cal Tech, Chicago, Columbia, the London School of Economics, Minnesota, Oxford, Stanford, and UCLA. Mookherjee's and Reichelstein's research was supported by the National Science Foundation (grant no. SES 9209455).

One alleged advantage of delegation is that it reduces the communication requirements between principals (owners, purchasers) and agents (managers, suppliers). Delegation can also be an instrument for distributing the burden of information processing more evenly among the members of the organization.[1] A possible disadvantage of delegation is that it may exacerbate incentive problems. Intermediate agents who have been given authority over certain decisions may pursue their own self-interest rather than that of the principal. Some authors have therefore suggested that hierarchical organizations are subject to a "loss of control," owing to the delegation of responsibility (Williamson, 1967; Calvo and Wellisz, 1978; and McAfee and McMillan, 1993).

We do not model the benefits of delegation explicitly. Instead, we take the perspective that the widespread practice of delegation makes it important to study the effectiveness of such arrangements. Specifically, we seek to delineate circumstances under which delegation becomes costly due to a loss of control by the principal.

Our model involves two agents who are privately informed about their respective costs of production. In a centralized arrangement, the principal contracts directly with both agents. Delegated contracting, on the other hand, involves the selection of a prime contractor to whom the principal delegates the authority to organize production and contract with the other agent. For the class of contracts we consider, the revelation principle implies that a centralized arrangement weakly dominates all other arrangements. The relevant issue then becomes the potential loss in performance as the principal replaces a two-tier hierarchy by one with three tiers.

A major result of the article is that, under certain conditions, the additional incentive problem inherent in delegated contracting can be alleviated completely. In particular, we exhibit a delegation mechanism that matches the expected profit for the principal under the optimal revelation mechanism. For this result to hold, it is essential that the principal can (i) monitor the contribution of the primary contractor to the joint product and (ii) design the sequence of contracts appropriately.

If the principal does not monitor the intermediate agent's contribution, that agent will be able to take advantage of his monopoly power. The resulting problem is similar in nature to the "double marginalization of rents" that frequently characterizes vertical relationships (see Tirole (1988)). In particular, the prime contractor will bias the assignment of production tasks in his favor, and assign the other agent production levels that are too low.[2] The severity of this problem depends on the degree to which the contributions of the two agents are substitutes for one another.

The issue of sequencing in the three-tier hierarchy has two distinct aspects. First, it is important that the intermediate agent accept the primary contract before he enters into the subcontract with the second agent. If this acceptance decision can be delayed, the intermediate agent obtains additional private information before contracting with the principal. This tends to increase the informational rent captured by the intermediate agent. A second aspect of the sequencing of communication and contracting is that the intermediate agent should report his own cost before entering the subcontract.[3] Such

[1] The importance of limited information-processing capabilities for hierarchical organization has been emphasized in Williamson (1975, 1985). For models of mechanism design with computation costs, see Radner (1993) and Radner and Van Zandt (1992). Marschak and Reichelstein (1994) analyze the structure of efficient networks and hierarchies when communication is costly.

[2] This prediction is consistent with similar claims in the government contracting literature. For instance, Rogerson (1992) argues that existing allocation rules for overhead costs will bias prime contractors against awarding subcontracts for components.

[3] In a model involving many agents and multiple branches of a hierarchy, it would be essential for the prime contractor(s) to report to the principal again after they have received information from their subordinates. Otherwise the organization would not be able to solve the problem of coordinating the activities of different branches; see Mookherjee and Reichelstein (1995).

sequencing is essential to screen the agent's private information and thereby to align his objective with that of the principal.

A number of recent articles have studied contractual hierarchies in the context of asymmetric information. In McAfee and McMillan (1993), the intermediate agent has no private information. The corresponding three-tier hierarchy nonetheless entails a loss relative to centralized contracting, since the intermediate agent's incentive contract is subject to bankruptcy constraints. These constraints are also present in our model when the sequence of contracting is such that the intermediate agent can refuse the primary contract after communicating with the subcontractor.

The models of Baron and Besanko (1992, 1994) and Gilbert and Riordan (1995) focus on settings where the two agents supply strictly complementary inputs. As a consequence, they find that centralized and delegated contracting can be performance equivalent for scenarios in which our model predicts the inferiority of the latter.[4] In addition, these authors study another organizational arrangement—labelled informational consolidation—where the two agents are "merged" into a single entity that knows both agents' production costs. We make further comparisons between their models and ours in Section 4.

In Melumad, Mookherjee, and Reichelstein (1992), we consider delegated contracting with a different kind of monitoring. The principal is assumed to be able to monitor the payment that the intermediate agent makes to the subcontractor. If that information becomes available for the prime contractor's reward scheme, the principal can alleviate the bias inherent in the intermediate agent's preference over alternative subcontracts.[5] While monitoring financial information is common within firms, such incentive mechanisms appear to be particularly vulnerable to collusion, i.e., the agents can do better for themselves by signing an (unobservable) side contract.

The remainder of the article is organized as follows. Section 2 introduces the model, and Section 3 presents the characterization of the optimal centralized mechanism. In Section 4 we develop the main results about the relative performance of centralized and delegated contracting. The article concludes in Section 5. All proofs are in the Appendix.

2. The model

■ We consider contracting problems involving a principal and two agents. Agent i can take some productive action a_i from a set A_i. The set of jointly feasible actions is the product $A = A_1 \times A_2$. For every action profile $a \equiv (a_1, a_2)$, the organization experiences costs and benefits. All benefits accrue directly to the principal according to a benefit function $B(a)$. If agent i takes action a_i, he bears the cost $C_i(a_i, \theta_i)$, where $\theta_i \in \Theta_i \equiv [\underline{\theta}_i, \overline{\theta}_i]$ is a parameter representing the agent's private information (his type). As a consequence, the cost $C_i(a_i, \theta_i)$ is not observable to anyone except agent i. The cost parameters, θ_i, are drawn independently from commonly known prior distributions $F_i(\theta_i)$ with positive densities $f_i(\theta_i)$.

Agents are compensated for the costs they bear by transfer payments x_i made either by the principal or by another agent. All parties are assumed to be risk neutral. Hence,

---

[4] In Baron and Besanko (1992), it is assumed that the principal observes the cost report that the subcontractor makes to the prime contractor. That assumption is relaxed in Baron and Besanko (1994). In both models, the prime contractor has less discretion than in our framework, since the principal specifies both agents' production levels, contingent on a report by the intermediate agent. This approach seems more appropriate for a setting with strictly complementary inputs, where the issue of task allocation does not arise.

[5] This result requires, however, that the prime contractor's cost function satisfy a certain separability condition. Melumad, Mookherjee, and Reichelstein (1992) also model the idea that agents cannot fully communicate their private information for contracting purposes. Delegated contracting may then outperform a centralized arrangement, as the prime contractor can act on better information about his own environment.

each agent's utility is the (expected) difference between the transfer payment he receives and his cost. Similarly, the principal seeks to maximize the difference between the benefit $B(a)$ and the sum of the (expected) payments made to the agents.[6] Our analysis invokes the following technical assumptions:

Assumption 1. $A_i = \mathbf{R}_+$.

Assumption 2. $B(\cdot, \cdot)$ is increasing and continuously differentiable in both variables.

Assumption 3. $F_i(\theta_i)/f_i(\theta_i)$ is increasing in θ_i.

Assumption 4. $C_i(a_i, \theta_i)$ is thrice differentiable, increasing in both arguments, and it satisfies

$$\frac{\partial^2}{\partial a_i \partial \theta_i} C_i(a_i, \theta_i) > 0, \quad \text{and} \quad \frac{\partial}{\partial \theta_i}\left[\frac{\partial^2}{\partial a_i \partial \theta_i} C_i(a_i, \theta_i)\right] \geq 0 \qquad \text{for all } a_i > 0,\ \theta_i \in \Theta_i.$$

Assumption 2 rules out the possibility of strictly complementary inputs as considered in Baron and Besanko (1992, 1994) and Gilbert and Riordan (1995). The second derivative in Assumption 4 represents the familiar "single-crossing property." The remaining parts of Assumptions 3 and 4 are also standard assumptions in adverse selection models (see, for example, Guesnerie and Laffont (1984)). These conditions ensure that a solution derived by recognizing only the "local" incentive constraints will also be globally incentive compatible.

3. Direct revelation mechanisms

■ The revelation principle implies that for any principal-agent problem, the principal can maximize her expected payoff by engaging the agents in a comprehensive "grand-contract." Such a contract corresponds to a two-tier hierarchical arrangement in which the agents simultaneously report their entire private information to the principal. The actions are then determined by a decision rule (to which the principal has committed herself) specifying actions for all possible configurations of the agents' reports. The corresponding mechanism design problem is represented by the following optimization program:[7]

$$\mathbf{P_0}: \qquad \max_{a(\theta), x(\theta)} E_\theta\left[B\big(a(\theta)\big) - \sum_{i=1}^{2} x_i(\theta)\right]$$

$$\text{subject to:} \qquad \text{for all } \theta_i \in \Theta_i,\ i \in \{1, 2\},$$

$$\theta_i \in \underset{\tilde{\theta}_i}{\operatorname{argmax}}\ E_{\theta_j}\Big[x_i(\tilde{\theta}_i, \theta_j) - C_i\big(a_i(\tilde{\theta}_i, \theta_j), \theta_i\big)\Big] \tag{i}$$

$$E_{\theta_j}\Big[x_i(\theta_i, \theta_j) - C_i\big(a_i(\theta_i, \theta_j), \theta_i\big)\Big] \geq 0. \tag{ii}$$

The first constraint in $\mathbf{P_0}$ is the standard incentive-compatibility condition, requiring

[6] Our model applies to a variety of team production problems in which the actions of the two agents need to be coordinated (see, for example, McAfee and McMillan (1991)). One particular variant is a procurement setting where the principal can buy a divisible good from two competing suppliers. In the simplest case, we have $A_i = \mathbf{R}_+$, $C_i(a_i, \theta_i) = \theta_i \cdot a_i$, and $B(a) \equiv \hat{B}(a_1 + a_2)$, i.e., the agents produce perfect substitutes at constant marginal cost.

[7] For convenience, we shall use the vector notation $\theta \equiv (\theta_1, \theta_2)$, $x \equiv (x_1, x_2)$, and $a \equiv (a_1, a_2)$. The notation $E_\theta[\cdot]$ represents the expected value with respect to (θ_1, θ_2), i.e., $E_\theta[\cdot] = \int_{\underline{\theta}_1}^{\bar{\theta}_1} \int_{\underline{\theta}_2}^{\bar{\theta}_2} [\cdot]\, dF_2(\theta_2)\, dF_1(\theta_1)$. Similarly, $E_{\theta_j}[\cdot] = \int_{\underline{\theta}_j}^{\bar{\theta}_j} [\cdot]\, dF_j(\theta_j)$.

that truth telling be a Bayesian-Nash equilibrium. Constraint (ii) represents a participation constraint reflecting that each agent knows his cost prior to contracting.[8]

The solution to $\mathbf{P}_0$ is well known from the literature on adverse selection. For any given production assignments, $a(\theta) \equiv (a_1(\theta_1, \theta_2), a_2(\theta_1, \theta_2))$, the payments to the agents are uniquely determined by the incentive and participation constraints. Furthermore, the agents will earn informational rents because of their private information, that is, their expected payoffs will generally exceed the lower bound given by the participation constraints. The principal needs to trade off production efficiency against the informational rents earned by the agents. The optimal balance between these conflicting objectives is summarized by the following result.

Lemma 1. Given Assumptions 1, 3, and 4, the optimal production assignments, $a^*(\theta)$, in $\mathbf{P}_0$ satisfy

$$\left(a_1^*(\theta_1, \theta_2), a_2^*(\theta_1, \theta_2)\right) \in \underset{(a_1,a_2)}{\operatorname{argmax}} \left\{B(a_1, a_2) - h_1(a_1, \theta_1) - h_2(a_2, \theta_2)\right\}, \quad (1)$$

where
$$h_i(a_i, \theta_i) \equiv C_i(a_i, \theta_i) + \frac{F_i(\theta_i)}{f_i(\theta_i)} \frac{\partial}{\partial \theta_i} C_i(a_i, \theta_i).$$

Following Myerson (1981), we refer to $h_i(a_i, \theta_i)$ as agent i's "virtual cost." The objective function in (1) reflects that $h_i(\cdot)$ is the relevant cost for the principal in choosing the production assignments rather than the underlying cost function $C_i(a_i, \theta_i)$. In expectation, the virtual cost exceeds the production cost $C_i(\cdot)$ by an amount equal to the agent's informational rent.

An alternative formulation of the incentive problem in $\mathbf{P}_0$ might impose the constraint that truthful reporting be a dominant strategy for each agent. Constraint (i) would then be replaced by

$$\theta_i \in \underset{\tilde{\theta}_i}{\operatorname{argmax}} \left\{x_i(\tilde{\theta}_i, \theta_j) - C_i\left(a_i(\tilde{\theta}_i, \theta_j), \theta_i\right)\right\} \qquad \text{for all } \theta_j \in \Theta_j. \qquad (\text{i}')$$

Further, the individual rationality constraint (ii) could be replaced by the stronger *ex post* constraint

$$x_i(\theta_i, \theta_j) - C_i\left(a_i(\theta_i, \theta_j), \theta_i\right) \geq 0 \qquad \text{for all } \theta_j \in \Theta_j. \qquad (\text{ii}')$$

Mookherjee and Reichelstein (1992) show that if constraints (i) and (ii) are replaced by the stronger constraints (i′) and (ii′), respectively, the principal's expected payoff in $\mathbf{P}_0$ remains unchanged. Thus, dominant strategies and *ex post* individual rationality can be obtained "for free" in this context.

A revelation mechanism may be viewed as a two-tier hierarchy in which both agents are subordinates of the principal. We subsequently refer to this organizational structure as H_0. The maximum payoff attainable by the principal under H_0 is the value $E_\theta[B(a^*(\theta)) - h_1(a_1^*(\theta), \theta_1) - h_2(a_2^*(\theta), \theta_2)]$. Since it represents the largest payoff attainable by any incentive-feasible mechanism, we refer to it as the second-best performance level. It serves as a natural benchmark for evaluating delegation mechanisms.

For future reference we also note the following revenue equivalence theorem, as observed by Myerson (1981) in the context of optimal auction design. Let $\{a(\theta), x(\theta)\}_{\theta \in \Theta}$

[8] Alternatively, the agents may find out their costs after contracting but they cannot be required to pay damages if they quit the contract.

be any Bayesian incentive-compatible revelation mechanism, i.e., one that satisfies constraint (i) in $\mathbf{P}_0$. The principal's expected payoff then equals

$$E_\theta\Big[B\big(a(\theta)\big), - h_1\big(a_1(\theta), \theta_1\big) - h_2\big(a_2(\theta), \theta_2\big)\Big] - \Pi_1(\overline{\theta}_1) - \Pi_2(\overline{\theta}_2),$$

where $\Pi_i(\theta_i) \equiv E_{\theta_j}[x_i(\theta_i, \theta_j) - C_i(a_i(\theta_i, \theta_j), \theta_i)]$ denotes agent i's expected payoff when his cost is θ_i. In particular, optimality requires that the production assignments are chosen according to (1), and that the participation constraints bind for the highest-cost types, i.e., $\Pi_1(\overline{\theta}_1) = \Pi_2(\overline{\theta}_2) = 0$.

To conclude this section, we present an example that will also illustrate several of our results below.

Example. Consider a procurement problem with *ex ante* identical agents. Let $\Theta_i = [0, 1]$, $C_i(a_i, \theta_i) = \theta_i{\cdot}a_i$ and $F_1(\cdot) = F_2(\cdot) = F(\cdot)$. Agent i's virtual cost then becomes $\alpha(\theta_i){\cdot}a_i$, where $\alpha(\theta_i) \equiv \theta_i + F(\theta_i)/f(\theta_i)$. The agents produce perfect substitutes, i.e.,

$$B(a_1, a_2) = \hat{B}(a_1 + a_2).$$

There is some exogenously given output level $\overline{B}$ that the principal wants to procure for all (θ_1, θ_2). The corresponding cost-minimization problem then becomes

$$\min_{a_1(\cdot), a_2(\cdot)} \int_0^1 \int_0^1 \big[\alpha(\theta_1){\cdot}a_1(\theta_1, \theta_2) + \alpha(\theta_2){\cdot}a_2(\theta_1, \theta_2)\big]\, dF(\theta_1)\, dF(\theta_2),$$

subject to $\hat{B}(a_1(\theta_1, \theta_2) + a_2(\theta_1, \theta_2)) \geq \overline{B}$. It follows immediately that only one agent should produce, that is, $a_i^*(\theta_1, \theta_2) \equiv \hat{B}^{-1}(\overline{B})$ and $a_j^*(\theta_1, \theta_2) = 0$ whenever $\alpha(\theta_i) < \alpha(\theta_j)$. Since $F(\theta_i)/f(\theta_i)$ is increasing in θ_i, we find that $\alpha(\theta_i) < \alpha(\theta_j)$ if and only if $\theta_i < \theta_j$. As observed by Myerson (1981), a standard second-price auction is optimal in this setting. Agent i is paid the amount θ_j for producing $\hat{B}^{-1}(\overline{B})$ if $\theta_i < \theta_j$, and zero otherwise. It can be verified that $E_{\theta_j}[\theta_j{\cdot}a_i^*(\theta_i, \theta_j)] = E_{\theta_j}[\alpha(\theta_i){\cdot}a_i^*(\theta_i, \theta_j)]$, i.e., an agent's expected payment is indeed equal to his expected virtual cost.

4. Delegated contracting

■ Our model of delegated contracting corresponds to a three-tier hierarchy in which the principal contracts only with agent 1, who is given the authority to contract with agent 2. Agent 1 can decide the production levels (a_1, a_2), though the choice of a_2 is determined by the subcontract with agent 2.[9] We consider several variants of the contractual three-tier hierarchy. These variants differ in the extent of monitoring by the principal and in the sequence of contracting and information reporting between the principal and the agents.

Monitoring of individual production contributions by the principal will generally be costly. Although we do not attempt to measure these costs explicitly, our results illustrate the value of monitoring, i.e., the extent to which monitoring the agents' individual contributions allows the principal to achieve better control in a three-tier hierarchy. In some situations the sequence of contracts and reports in a three-tier hierarchy can be chosen by the principal. In other situations she may have only partial control over the sequence of events. For instance, the principal may be unable to elicit

[9] In contrast to our analysis, Melumad and Suehiro (1995) make the choice of prime contractor an endogenous component of the model.

a report from agent 1 *before* that agent has an opportunity to communicate and contract with agent 2.

□ **The base scenario.** In our first variant of delegated contracting, the principal monitors the individual contributions a_1 and a_2 for the purpose of contracting. The first stage consists of two moves: the principal offers agent 1 a contract of the form $x_1(a, \tilde{\theta}_1)$. Agent 1 then responds by his acceptance of the contract and by a report $\tilde{\theta}_1$ regarding his own cost θ_1.[10] In effect, agent 1 chooses one contract from a menu of alternatives. The second stage begins with agent 1 offering a (sub)contract to agent 2. That agent then responds by agreeing to produce some a_2 as part of the subcontract. The third and final move of the second stage lets agent 1 choose an action a_1 in response to the a_2 chosen by the second agent.

We assume that the second-stage interaction between the two agents is not observed by the prinicipal, i.e., she neither monitors the incentive scheme agent 1 offers to agent 2 nor observes the subsequent report by agent 2. On the other hand, it is of no importance to our analysis whether agent 2 has knowledge of the prime contract, which the principal offers to agent 1, and the report subsequently made by the latter. To see this, note that without loss of generality agent 1 offers agent 2 a revelation scheme $\{a_2(\theta_2), x_2(\theta_2)\}_{\theta_2 \in \Theta_2}$ at the second stage. Agent 1's preferences over such schemes depend on his type θ_1, his earlier report $\tilde{\theta}_1$ to the principal, and the contract $x_1(a, \tilde{\theta}_1)$. Since both agents' utility functions are linear in x_2, the informed-principal problem (Myerson, 1983 and Maskin and Tirole, 1990) does not arise in this context: Agent 1 cannot gain by delaying the revelation of his private information to agent 2 until after the latter has responded to the subcontract.[11]

For a given contract $x_1(a, \tilde{\theta}_1)$, we shall denote the second-stage revelation mechanism for agent 2 by $\{a_2(\theta_2 | \theta_1, \tilde{\theta}_1), x_2(\theta_2 | \theta_1, \tilde{\theta}_1)\}_{\theta_2 \in \Theta_2}$ to reflect its dependence on θ_1 and $\tilde{\theta}_1$. Following agent 2's (truthful) report, θ_2, agent 1 will subsequently choose a_1 so as to maximize $x_1(a_1, a_2(\theta_2 | \theta_1, \tilde{\theta}_1), \tilde{\theta}_1) - C_1(a_1, \theta_1)$. The resulting sequential optimization problem can be represented as a simultaneous choice of the functions $\{a_1(\cdot | \theta_1, \tilde{\theta}_1), a_2(\cdot | \theta_1, \tilde{\theta}_1), x_2(\cdot | \theta_1, \tilde{\theta}_1)\}$ in order to solve the following subcontracting program[12]

$$\mathbf{SP}(\theta_1 | x_1(a, \tilde{\theta}_1)): \qquad \max_{a(\cdot), x_2(\cdot)} E_{\theta_2}\left[x_1(a(\theta_2 | \theta_1, \tilde{\theta}_1), \tilde{\theta}_1) - C_1(a_1(\theta_2 | \theta_1, \tilde{\theta}_1), \theta_1) - x_2(\theta_2 | \theta_1, \tilde{\theta}_1)\right]$$

subject to: for all $\theta_2 \in \Theta_2$, and all $\theta_1, \tilde{\theta}_1 \in \Theta_1$,

$$\theta_2 \in \operatorname*{argmax}_{\tilde{\theta}_2} \left\{x_2(\tilde{\theta}_2 | \theta_1, \tilde{\theta}_1) - C_2(a_2(\tilde{\theta}_2 | \theta_1, \tilde{\theta}_1), \theta_2)\right\} \tag{i}$$

$$x_2(\theta_2 | \theta_1, \tilde{\theta}_1) - C_2(a_2(\theta_2 | \theta_1, \tilde{\theta}_1), \theta_2) \geq 0. \tag{ii}$$

Constraints (i) and (ii) respectively represent the incentive and participation constraints for agent 2. By $\Gamma_1(\theta_1 | x_1(a, \tilde{\theta}_1))$ we denote the value of the optimization program $\mathbf{SP}(\theta_1 | x_1(a, \tilde{\theta}_1))$. The function $\Gamma_1(\cdot | \cdot)$ represents the "reduced-form" utility of agent 1

[10] Without loss of generality, we can assume that all types of agent 1 accept the contract and therefore there is no need to model the acceptance decision explicitly.

[11] This observation follows in particular from Proposition 11 in Maskin and Tirole (1990).

[12] For notational ease, we write $\{a_1(\cdot | \theta_1, \tilde{\theta}_1), a_2(\cdot | \theta_1, \tilde{\theta}_1), x_2(\cdot | \theta_1, \tilde{\theta}_1)\}$ instead of

$$\{a_1(\theta_2 | \theta_1, \tilde{\theta}_1), a_2(\theta_2 | \theta_1, \tilde{\theta}_1), x_2(\theta_2 | \theta_1, \tilde{\theta}_1)\}_{\theta_2 \in \Theta_2}.$$

when he reports to the principal at the first stage. Therefore, the principal's problem at the first stage reduces to

$$\mathbf{P_1}: \qquad \max_{x_1(a,\theta_1),a(\theta)} E_\theta\big[B(a(\theta)) - x_1(a(\theta),\, \theta_1)\big]$$

subject to: for all $\theta_1 \in \Theta_1$,

$$\theta_1 \in \operatorname*{argmax}_{\tilde{\theta}_1} \Gamma_1(\theta_1 \mid x_1(a,\, \tilde{\theta}_1)) \tag{i}$$

$$\Gamma_1(\theta_1 \mid x_1(a,\, \theta_1)) \geq 0 \tag{ii}$$

$$\text{there exist } \{a(\cdot \mid \theta_1,\, \theta_1),\, x_2(\cdot \mid \theta_1,\, \theta_1)\} \in \operatorname{argmax} \mathbf{SP}(\theta_1 \mid x_1(a,\, \theta_1)),$$
$$\text{such that } a(\theta) \equiv a(\theta_2 \mid \theta_1,\, \theta_1), \text{ for all } \theta_2. \tag{iii}$$

Constraints (i) and (ii) represent the incentive and participation constraints for agent 1, based on his reduced-form utility function Γ_1. The third constraint reflects that the desired production assignments must coincide with those that agent 1 will choose when he subcontracts with agent 2. We refer to the delegated contracting arrangement just described as H_1. For this organizational arrangement, the maximum payoff attainable by the principal is the value of $\mathbf{P_1}$. As noted above, the revelation principle implies that the three-tier hierarchy H_1 cannot do better than H_0. The relevant question then is whether H_1 can attain the benchmark of second-best performance. If so, the multitier hierarchy may be preferable, since the principal only communicates and contracts with one agent.

Theorem 1. Given Assumptions 1, 3, and 4, the three-tier hierarchy H_1 achieves second-best performance.

Delegation is potentially prone to a "control loss" problem, since agent 1 is given monopoly power in his subcontract with agent 2. As explained below, the prime contractor has a tendency to bias the production assignments in his favor. To counteract this tendency, the principal may shift agent 1's preferences by adopting the following incentive scheme in H_1:

$$x_1(a,\, \tilde{\theta}_1) = B(a_1,\, a_2) - \big[h_1(a_1,\, \tilde{\theta}_1) - C_1(a_1,\, \tilde{\theta}_1)\big] + \beta(\tilde{\theta}_1). \tag{2}$$

The last term on the right-hand side of (2) is a lump-sum amount (depending only on the report $\tilde{\theta}_1$) and is designed to elicit truthful reporting from the agent. The second term amounts to a tax on agent 1's production contribution, depending on the observed a_1 and the reported $\tilde{\theta}_1$. This tax induces agent 1 to mark up his own cost and, provided $\tilde{\theta}_1 = \theta_1$, agent 1 will impute the virtual cost $h_1(a_1,\, \theta_1)$ for his own contribution. Since under any incentive-compatible scheme the expected payment to agent 2 must equal his expected virtual cost, the objective function of agent 1 reduces to

$$E_{\theta_2}\big[B(a_1,\, a_2) - h_1(a_1,\, \theta_1) - h_2(a_2,\, \theta_2)\big] + \beta(\theta_1).$$

Thus, agent 1 internalizes the principal's objective as given in (1). The proof of Theorem 1 shows that it is possible to construct the function $\beta(\theta_1)$ so that agent 1 will indeed report truthfully at the first stage and his expected profit will be equal to zero at $\theta_1 = \overline{\theta}_1$. The claim then follows from the revenue equivalence theorem.

We note that the incentive scheme in (2) does not amount to "selling the business" to agent 1, in the sense that the principal receives a constant payoff. For the above incentive scheme, the principal's payoff varies with both θ_1 and θ_2, since $a_1^*(\cdot)$ depends on these variables. In order to make agent 1 the full residual claimant, the selling price would have to include an informational rent which exceeds that under the optimal revelation mechanism.[13]

Theorem 1 would readily extend to situations where agent 1 contracts with an arbitrary number of subcontractors at the third tier of the hierarchy. The principal can implement optimal production choices by monitoring aggregate output, $B(a)$, and a_1. In particular, there would be no need to observe the individual contributions of the remaining agents, irrespective of the size of the team.

If one takes the perspective that monitoring the contributions of individual team members is costly, it is natural to ask whether Theorem 1 remains true if the principal can observe agent 1's contribution imperfectly. A one-dimensional random variable s is said to be an unbiased signal for a_1, if $E[s|a_1] = a_1$, i.e., the expected value of s, conditional on a_1, is always equal to a_1.

Corollary to Theorem 1. Suppose agent 1's cost function takes the form

$$C_1(a_1, \theta_1) = v_1(\theta_1)\cdot a_1.$$

If the principal observes only the aggregate output $B(a_1, a_2)$ and an unbiased signal for a_1, the three-tier hierarchy H_1 still achieves second-best performance.

When agent 1's cost function is multiplicatively separable, i.e., $C_1(a_1, \theta_1) = v_1(\theta_1)\cdot a_1$, the virtual cost becomes $h_1(a_1, \theta_1) = (v_1(\theta_1) + v_1'(\theta_1)\cdot F_1(\theta_1)/f_1(\theta_1))\cdot a_1$.[14] In the three-tier hierarchy H_1, the principal can then adopt a variant of the incentive scheme in (2):

$$x_1(B, s, \theta_1) = B - v_1'(\theta_1)\cdot\frac{F_1(\theta_1)}{f_1(\theta_1)}\cdot s + \beta(\theta_1).$$

Since s is an unbiased signal of a_1, the incentives for agent 1 and the expected payoffs are unchanged from the setting where a_1 was observed perfectly by the principal. It should be noted that the assumed risk neutrality of agent 1 is crucial in this context: the result holds irrespective of how noisy the signal is, and agent 1 bears the entire risk associated with the realization of the signal.

□ **Lack of monitoring.** We now turn to a scenario in which the principal continues to observe the joint output, B, but does not obtain any additional signal for agent 1's contribution. We refer to the corresponding hierarchy as H_2. It involves the same sequence of moves as H_1, except that the principal's contract with agent 1 has to take the form $x_1(B, \tilde{\theta}_1)$. To examine how H_2 compares with H_0, note that the multiagent problem is degenerate if there exists a second-best policy, $\{a_1^*(\cdot), a_2^*(\cdot)\}$, such that for one of the agents, say agent i, $a_i^*(\theta_1, \theta_2) \equiv 0$ for all θ_1 and θ_2. In that case we call agent i dispensable. It is readily verified that the three-tier hierarchy H_2 can attain the same performance as H_0 when either agent is dispensable. We therefore exclude this case in the following result.

[13] To see this, note that if agent 1 were the residual claimant, he would choose a production level a_1 that is optimal relative to $C_1(\cdot)$ rather than $h_1(\cdot)$. Any such mechanism must involve higher rents for agent 1 than those resulting from second-best production assignments.

[14] If the cost function is multiplicatively separable, i.e., $C_1(a_1, \theta_1) = v_1(\theta_1)\cdot w_1(a_1)$, we can always change the units of agent 1's contribution so that $w_1(a_1) \equiv a_1$.

Theorem 2. Suppose Assumptions 1–4 hold, and neither agent is dispensable. Then the three-tier hierarchy H_2 does not achieve second-best performance.

The hierarchy H_2 is subject to a loss of control and a double marginalization of rents. Since the prime contractor's incentive can be based only on aggregate output, he will trade off his own production cost, $C_1(a_1, \theta_1)$, against the virtual cost, $h_2(a_2, \theta_2)$, of agent 2. But the marginal cost $\partial C_1/\partial a_1$ is less than the marginal virtual cost $\partial h_1/\partial a_1$, and as a consequence, agent 1 will assign himself more production than is desired by the principal. In addition, agent 1's informational rent will increase beyond the second-best level. To illustrate the bias in agent 1's objective further, we consider again the example introduced in Section 3.

Example (continued). The principal offers agent 1 a fixed payment $\bar{x}_1$ for delivery of $\bar{B}$. Since the payment $\bar{x}_1$ is sunk for agent 1 at the second stage, he will make a take-it-or-leave-it offer to agent 2 for the entire quantity $a_2 = \hat{B}^{-1}(\bar{B})$. The unit price, p, at which agent 1 offers production to agent 2 is chosen so as to minimize

$$\left[p \cdot F(p) + \theta_1 \cdot \left(1 - F(p)\right)\right] \cdot \hat{B}^{-1}(\bar{B}). \tag{3}$$

Since $F(\theta)/f(\theta)$ is increasing, this is a convex minimization problem. The first-order condition shows that the optimal p is given by $p = \alpha^{-1}(\theta_1)$, where $\alpha(\theta) \equiv \theta + F(\theta)/f(\theta)$. Hence, agent 1 produces the entire quantity if and only if $\theta_2 > \alpha^{-1}(\theta_1)$. Production will then be distorted for cost environments where θ_2 is between $\alpha^{-1}(\theta_1)$ and θ_1. In those environments, agent 1 will be the supplier in H_2, even though the second-best solution requires agent 2 to produce.[15]

The result of Theorem 2 is only seemingly at odds with those obtained by Baron and Besanko (1992, 1994) and Gilbert and Riordan (1995). They find that a three-tier hierarchy can replicate the optimal revelation mechanism even though the principal observes only the aggregate output. The difference arises because their models consider strict complementarities between the two suppliers.[16] The aggregate output then fully reveals each agent's contribution. Formally, we note that when $B(a_1, a_2) = \min\{a_1, a_2\}$, the principal can resort again to the incentive scheme in (2), that is, she may set $x_1(B, \theta_1) = B - [h_1(B, \theta_1) - C_1(B, \theta_1)] + \beta(\theta_1)$. The hierarchy H_2 then achieves second-best performance.

Theorems 1 and 2 indicate that the three-tier hierarchy H_2 is inferior because it entails both delegation of contracting and lack of monitoring. A result by McAfee and McMillan (1991) indicates that lack of monitoring by itself may have no adverse consequences. Under centralized contracting, they find that second-best performance remains attainable if the principal is confined to observing aggregate output rather than individual contributions. To implement the second-best production assignments, the principal constructs a two-stage mechanism, wherein agents first report their types and then choose their a_i's independently.

□ **Alternative sequencing.** We now explore the role of the sequencing of contracts in the hierarchical mechanism H_1, where the principal monitors agent 1's contribution to joint output. It is conceivable that the principal may not be able to elicit a report

[15] Bulow and Roberts (1989) link this distortion to a standard monopoly problem, in which a seller faces a buyer with unknown reservation value θ_2, distributed according to the function $F(\theta_2)$. If the seller's reservation value is θ_1, and he charges a price p, the expected profit will be $(p - \theta_1)F(p)$. Interpreting $F(\theta_2)$ as a demand curve, agent 1's problem is precisely that of a monopolist having marginal cost θ_1.

[16] As noted above, the models of Baron and Besanko (1992, 1994) differ from ours also in several other respects.

from agent 1 before that agent signs a contract with agent 2. Specifically, we study the following variant of the three-tier hierarchy H_1, which we label H_1'. Assume that the principal monitors a_1 and a_2, and the sequence of events is as follows. At the first stage, the principal offers agent 1 a contract of the form $x_1(a_1, a_2, m)$, where $m \in M$ denotes a report (message) that agent 1 can send to the principal after communicating with agent 2. The first stage ends without agent 1 responding to the contract offered to him.

At the second stage, the first two moves are the same as in H_1. That is, agent 1 offers a (revelation) contract to agent 2, who then makes a report and thereby commits to producing a particular level of a_2. The third and final move of the second stage calls for agent 1 to choose a_1 and to report some m to the principal. The participation constraint now requires that agent 1 can refuse the principal's contract at the end of the second stage, i.e., agent 1 will be paid (at least) zero for delivering $a_1 = a_2 = 0$.[17] Such an *ex post* participation constraint is also imposed in Baron and Besanko (1992, 1994).[18] We note that requiring the individual-rationality constraints to hold *ex post* does not immediately make H_1' inferior to H_1 (and H_0). As argued in Section 3, the principal's expected payoff remains unchanged if one imposes that the participation constraints in $\mathbf{P}_0$ must hold *ex post.*

A standard argument shows that the principal can without loss set $M = \Theta_1 \times \Theta_2$ and induce agent 1 to report (θ_1, θ_2) truthfully at the second stage of the mechanism. As a consequence, the principal faces a two-dimensional adverse-selection problem with agent 1. Alternatively, the principal can also resort to a pure delegation scheme, wherein agent 1 does not make any report but simply delivers a_1 and a_2.[19] This observation is useful for comparing the performance of H_1 and H_1'. To state the formal result, we extend the earlier definition of a dispensable agent in the following way: Agent 2 is dispensable (in H_0) at $\theta_1 = \overline{\theta}_1$, if for some second-best policy, $a^*(\theta)$, it is true that $a_2^*(\overline{\theta}_1, \theta_2) = 0$ for all θ_2.

Theorem 3. The three-tier hierarchy H_1 weakly dominates H_1'. Given Assumptions 1–4, H_1 strictly dominates H_1' if agent 2 is not dispensable at $\theta_1 = \overline{\theta}_1$.

The condition that agent 2 not be dispensable at $\theta_1 = \overline{\theta}_1$ is generally stronger than the condition that agent 2 not be dispensable (as invoked in Theorem 2). We note, however, that if the agents' contributions to aggregate output are weak substitutes, then agent 2's contribution will be largest when $\theta_1 = \overline{\theta}_1$. In particular, $a_2^*(\theta_1, \theta_2)$ is increasing in θ_1, provided $\partial^2 B(a_1, a_2)/\partial a_1 \partial a_2 \leq 0$ for all a_1 and a_2.

Corollary to Theorem 3. Given Assumptions 1–4, suppose that $\partial^2 B(a_1, a_2)/\partial a_1 \partial a_2 \leq 0$ and that agent 2 is not dispensable. Then H_1 strictly dominates H_1'.

As noted in connection with Theorem 2, the nondispensability of agent 2 is also a necessary condition for strict dominance: H_1' and H_1 are performance equivalent whenever agent 2 is dispensable.[20] Thus we find that, under the assumptions stated in

[17] We are assuming implicitly that $C_i(0, \theta_i) = 0$ for all θ_i. We also assume that even though agent 1 may refuse the principal's contract at the end, he can sign a binding contract with agent 2 at the beginning of the second stage. Finally, agent 1 could always withhold agent 2's contribution and deliver zero output to the principal.

[18] Below we also consider a hybrid form of delegated contracting, where agent 1 can commit to accepting the contract at the first stage, but delays reporting of his information until the second stage.

[19] Given any scheme $x_1(a_1, a_2, m)$, the principal can create an equivalent scheme:

$$\hat{x}_2(a_1, a_2) \equiv \sup_{m \in M}\{x_1(a_1, a_2, m)\}.$$

[20] The optimal revelation mechanism takes the form $\{a_1^*(\theta_1), x_1(\theta_1)\}_{\theta_1 \in \Theta_1}$ when agent 2 is dispensable. Such a mechanism can equivalently be implemented by a delegation scheme $x_1(a_1)$, such that $x_1(a_1) \geq 0$ for all a_1.

the above corollary, the nondispensability of agent 2 is necessary and sufficient for H_1 to strictly dominate H_1'.

The proof of Theorem 3 centers on the fact that in H_1' the informational rent of agent 1 must exceed that under the optimal revelation mechanism. This feature results from two interacting requirements in H_1': Agent 1 can refuse the principal's contract after "talking" to agent 2, and agent 1 needs to be given incentives for adopting the "proper" subcontract. If the highest-cost type, $\overline{\theta}_1$, were to break even on average, he would earn positive rents for some (low) values of θ_2 and earn less than zero for high values of θ_2. However, since agent 1 can obtain a payoff of zero by not delivering any output, he cannot be compelled to enter into an unprofitable contract with the high types of agent 2.

Can agent 1 be induced to implement the second-best production assignments in H_1', irrespective of the informational rents such a policy may require? We recall that in H_1 the principal can shift agent 1's preferences over alternative subcontracts by imposing a tax on his contribution. This tax depends on agent 1's report at the time of accepting the principal's contract. Since H_1' does not permit such reporting, agent 1 can no longer be given a corresponding incentive scheme before finding out the other agent's cost. We note that for a single-agent problem, it would not matter whether the principal can screen the agent's type, since a pure delegation scheme can always replicate the outcome of an incentive-compatible revelation mechanism. In the three-tier hierarchy H_1', however, communication may be essential because of the presence of agent 2. The following example illustrates this point.

Example (continued). Consider again the procurement setting introduced in Section 3. The principal needs to procure a given $\overline{B}$; the agents produce perfect substitutes ($B(a_1, a_2) = \hat{B}(a_1 + a_2)$) and their costs are *ex ante* identical and linear in a_i. The second-best assignments are given by $a_1^*(\theta_1, \theta_2) = \hat{B}^{-1}(\overline{B})$ if $\theta_1 \leq \theta_2$ and $a_1^*(\theta_1, \theta_2) = 0$ otherwise. Suppose there exists an incentive scheme $\hat{x}_1\,(a_1, a_2)$ that induces agent 1 to implement these assignments in H_1'. In equilibrium, it would then be true that either $a_i = \hat{B}^{-1}(\overline{B})$ or $a_i = 0$ for all θ_1 and θ_2. Thus, without loss of generality,

$$\hat{x}_1(a_1, a_2) \equiv u_1(a_1, a_2) \cdot \hat{B}^{-1}(\overline{B}),$$

where

$$u_1(a_1, a_2) = \begin{cases} u_1^1 & \text{if } a_1 = \hat{B}^{-1}(\overline{B}) \\ u_1^2 & \text{if } a_2 = \hat{B}^{-1}(\overline{B}) \\ -k & \text{otherwise,} \end{cases}$$

where k is a large number. Given this payment function, type θ_1 of agent 1 can do no better than to offer agent 2 the right to produce $\hat{B}^{-1}(\overline{B})$ for a payment of $p \cdot \hat{B}^{-1}(\overline{B})$, or not produce at all. The unit price p is calculated so as to maximize

$$(u_1^2 - p) \cdot F(p) + (u_1^1 - \theta_1) \cdot (1 - F(p)),$$

or equivalently $p = \alpha^{-1}(\theta_1 + u_1^2 - u_1^1)$, for p interior in (0, 1).

Agent 1 will implement the second-best assignments only if $\theta_1 = \alpha^{-1}(\theta_1 + u_1^2 - u_1^1)$, for all θ_1. However, that is impossible because $\alpha(\theta) \equiv \theta + F(\theta)/f(\theta)$. In contrast, in H_1 the principal has the option of setting $u_1^1(\theta_1) = u_1^2 - F(\theta_1)/f(\theta_1)$ and thus aligning agent 1's incentives with her own.

To conclude this section, we mention two possible hybrid versions of the three-tier hierarchies H_1 and H_1'. Consider first a setting where the principal can elicit a report $\tilde{\theta}_1$ from agent 1 at the first stage (as in H_1), but agent 1 can still refuse the principal's contract at the end of the second stage, i.e., agent 1 can always claim a payment of zero for not delivering any output. We refer to this delegated contracting arrangement as H_1''. It is easily checked that the proof of Theorem 3 would apply without change, if agent 1 were to make a report at the first stage. Therefore, Theorem 3 and its corollary remain true if one replaces H_1' by H_1'' in the statements.

Finally, consider the three-tier hierarchy that results when agent 1 can make a commitment to accept the principal's contract at the first stage, but agent 1 can delay any reporting of information until the end of the second stage (for instance, agent 1 may need to do some research before he finds out his own cost). As in H_1, the participation constraints for agent 1 have to hold only in an interim sense. Such an arrangement will again be weakly dominated by H_1, and the dominance will be strict whenever agent 1 cannot be induced to implement the second-best production assignments. While this issue needs to be analyzed in more detail, we can conclude that the sequencing in the three-tier hierarchy H_1 is the most desirable from the principal's perspective.

5. Concluding remarks

■ In a two-agent model we have demonstrated that the performance of an optimal revelation mechanism can be replicated by a three-tier hierarchy, wherein the principal contracts with only one agent and delegates to that agent the authority to contract with the other agent. The three-tier hierarchy is prone to a control loss because the intermediate agent will have a tendency to bias the allocation of production in his favor. To avoid such distortions, the principal must monitor the contribution of the intermediate agent and ensure that the subcontract is not executed before the prime contract.

Our model has not accounted explicitly for the benefits associated with delegation mechanisms. One possible benefit frequently mentioned in the management literature is that delegation entails greater flexibility in a world where agents cannot communicate their entire private information to others. Such constraints may arise from expertise possessed by some agents and not shared by others, or from the presence of communication and information-processing costs. When there are limitations on the amount of information that agents can communicate for contracting purposes, the potential control loss associated with delegation mechanisms may resurface. Any such loss needs to be traded off against the advantage in delegating decision making to a better informed agent.[21]

Future research will show to what extent the results obtained in the present model also apply to larger organizations. In our model, the intermediate agent's report to the principal is needed only to avoid a loss of control, but this report serves no coordination or planning purpose. In a hierarchy with multiple branches, information reporting will be required for both coordination and control. In particular, the report by an intermediate agent to his superior will effectively have to convey the private information of all subordinates of the intermediate agent. In general, larger hierarchies therefore require a more complex information-reporting structure than the one considered here.[22]

[21] In a setting where the principal can monitor payments between the two agents, our earlier work (Melumad, Mookherjee, and Reichelstein, 1992) has shown that with limited communication, delegation schemes always outperform centralized ones. If payments between the agents cannot be monitored, examples show that the comparison between the two organizational modes can go either way.

[22] Contractual hierarchies with multiple branches have been considered by Crémer and Riordan (1987) and Mookherjee and Reichelstein (1995). In these models, intermediate agents have typically less discretion than in our framework.

The ability of intermediate agents to make commitments and to monitor select variables is likely to play a key role in making delegated contracting an effective organizational mode.

Our analysis of centralized mechanisms has assumed noncooperative behavior by the agents. The literature on incentives has recognized for some time that such mechanisms are generally vulnerable to collusion, i.e., to unobservable side contracting between the agents (see Green and Laffont (1979), Tirole (1992), Felli (1993), and Crémer (1994)). Three-tier hierarchies, on the other hand, make explicit allowance for side contracting, giving the intermediate agent the entire bargaining power. A natural question arising from our analysis is whether the three-tier hierarchy H_1' (in which the intermediate agent enters into the subcontract before he accepts the principal's contract) is in fact inferior to centralized contracting, once the possibility of side agreements is recognized explicitly in the formulation of centralized contracting.[23]

Appendix

■ Proofs of Lemma 1 and Theorems 1–3 follow.

Proof of Lemma 1. Let $\Pi_i(\theta_i)$ denote agent i's expected payoff when his type is θ_i. As Mirrlees (1986) has shown, the requirement of local incentive compatibility implies that

$$\Pi_i'(\theta_i) = E_{\theta_j}\left[-\frac{\partial}{\partial\theta_i}C_i\big(a_i(\theta_i, \theta_j), \theta_i\big)\right]. \tag{A1}$$

Therefore

$$\Pi_i(\theta_i) - \Pi_i(\overline{\theta}_i) = \int_{\theta_i}^{\overline{\theta}_i} E_{\theta_j}\left[\frac{\partial}{\partial\theta_i}C_i\big(a_i(t, \theta_j), t\big)\right] dt. \tag{A2}$$

Since the agent's individual-rationality constraint is binding only for the highest-cost type, it follows that $\Pi_i(\overline{\theta}_i) = 0$. By definition, $E_{\theta_j}[x_i(\theta_i, \theta_j)] = E_{\theta_j}[C_i(a_i(\theta_i, \theta_j), \theta_i)] + \Pi_i(\theta_i)$. To substitute for the payments, $x_i(\cdot)$, in the objective function of $\mathbf{P_0}$, we integrate $x_i(\cdot)$ with respect to θ_i. Integration by parts shows that

$$\int_{\Theta_i} \Pi_i(\theta_i)\, dF_i(\theta_i) = E_\theta\left[\frac{\partial}{\partial\theta_i}C_i\big(a_i(\theta_i, \theta_j), \theta_i\big)\cdot\frac{F_i(\theta_i)}{f_i(\theta_i)}\right]$$

and therefore, $E_\theta[x_i(\theta_i, \theta_j)] = E_\theta[h_i(a_i(\theta_i, \theta_j), \theta_i)]$, where $h_i(a_i, \theta_i) \equiv C_i(a_i, \theta_i) + (F_i(\theta_i))/(f_i(\theta_i))\cdot\partial C_i(a_i, \theta_i)/\partial\theta_i$ denotes the virtual cost as defined in the statement of the lemma. It remains to check that a solution $(a_1^*(\theta_1, \theta_2), a_2^*(\theta_1, \theta_2))$ to the maximization problem in (1) is indeed globally incentive compatible, that is, $\Pi_i(\theta_i) \geq E_{\theta_j}[x_i(\tilde{\theta}_i, \theta_j) - C_i(a_i^*(\tilde{\theta}_i, \theta_j), \theta_i)]$ for all θ_i, $\tilde{\theta}_i$. This inequality can be rewritten as

$$E_{\theta_j}\left[\int_{\theta_i}^{\tilde{\theta}_i} \frac{\partial}{\partial\theta_i}C_i\big(a_i^*(t, \theta_j), t\big)\, dt\right] \geq E_{\theta_j}\Big[C_i\big(a_i^*(\tilde{\theta}_i, \theta_j), \tilde{\theta}_i\big) - C_i\big(a_i^*(\tilde{\theta}_i, \theta_j), \theta_i\big)\Big].$$

Since $\partial C_i(a_i, \theta_i)/\partial\theta_i$ is increasing in a_i by Assumption 4, the above inequality will hold as long as $a_i^*(\theta_i, \theta_j)$ is decreasing in θ_i for any given θ_j. That, in turn, follows from the assumption that $\partial h_i(a_i, \theta_i)/\partial a_i$ is increasing in θ_i. *Q.E.D.*

Proof of Theorem 1. We show that for a proper choice of the function $\beta(\tilde{\theta}_1)$, the compensation scheme

$$x_1(a_1, a_2, \tilde{\theta}_1) = B(a_1, a_2) - \big[h_1(a_1, \tilde{\theta}_1) - C_1(a_1, \tilde{\theta}_1)\big] + \beta(\tilde{\theta}_1) \tag{A3}$$

will induce agent 1 to report truthfully and implement action choices that are optimal in $\mathbf{P_0}$, i.e., maximize $B(a_1, a_2) - h_1(a_1, \theta_1) - h_2(a_2, \theta_2)$. Following a report of $\tilde{\theta}_1$, agent 1's expected payoff equals

[23] Laffont and Martimort (1995a, 1995b) find that the possibility of collusion does not lower the performance of centralized contracting below the second-best level. As these authors note, however, their result is likely to be driven by the special modelling assumption that an agent's private information can take only one of two possible values.

$$\Gamma_1(\theta_1 | x_1(a, \tilde{\theta}_1)) = \max_{a_1, a_2} E_{\theta_2}\Big[B(a_1, a_2) - \big[h_1(a_1, \tilde{\theta}_1) - C_1(a_1, \tilde{\theta}_1)\big] + \beta(\tilde{\theta}_1) - C_1(a_1, \theta_1) - h_2(a_2, \theta_2)\Big]. \quad \text{(A4)}$$

For brevity, we shall from here on simply write $\Gamma_1(\theta_1, \tilde{\theta}_1)$ instead of $\Gamma_1(\theta_1 | x_1(a, \tilde{\theta}_1))$. Clearly, if agent 1 reports truthfully, he will subsequently choose $\{a_1(\cdot), a_2(\cdot)\}$ so as to maximize pointwise

$$B(a_1, a_2) - h_1(a_1, \theta_1) - h_2(a_2, \theta_2).$$

To establish incentive compatibility, we invoke Lemmas 6.1 and 6.3 of Mirrlees (1986). It suffices to show that for all θ_1,

$$\partial\Gamma_1(\theta_1, \tilde{\theta}_1)/\partial\theta_1 \text{ is (weakly) increasing in } \tilde{\theta}_1. \quad \text{(A5)}$$

Let $a(\theta_1, \theta_2, \tilde{\theta}_1) \equiv (a_1(\theta_1, \theta_2, \tilde{\theta}_1), a_2(\theta_1, \theta_2, \tilde{\theta}_1))$ denote a maximizer of the right-hand side of (A4). By the envelope theorem,

$$\frac{\partial}{\partial\theta_1}\Gamma_1(\theta_1, \tilde{\theta}_1) = -E_{\theta_2}\left[\frac{\partial}{\partial\theta_1}C_1\big(a_1(\theta_1, \theta_2, \tilde{\theta}_1), \theta_1\big)\right].$$

To verify that $\partial\Gamma_1(\theta_1, \tilde{\theta}_1)/\partial\theta_1$ is (weakly) increasing in $\tilde{\theta}_1$, it suffices to show that $a_1(\theta_1, \theta_2, \tilde{\theta}_1)$ is (weakly) decreasing in $\tilde{\theta}_1$, since the single-crossing property is assumed to hold, i.e., $\partial^2 C_1(\cdot, \cdot)/\partial\theta_1\partial a_1 > 0$. By definition, $a(\theta_1, \theta_2, \tilde{\theta}_1)$ is a pointwise maximizer of

$$B(a) - \frac{F_1(\tilde{\theta}_1)}{f_1(\tilde{\theta}_1)}\cdot\frac{\partial}{\partial\theta_1}C_1(a_1, \tilde{\theta}_1) - C_1(a_1, \theta_1) - h_2(a_2, \theta_2).$$

By virtue of Assumptions 3 and 4

$$\frac{\partial}{\partial a_1}\left[\frac{F_1(\tilde{\theta}_1)}{f_1(\tilde{\theta}_1)}\cdot\frac{\partial}{\partial\theta_1}C_1(a_1, \tilde{\theta}_1) + C_1(a_1, \theta_1)\right]$$

is increasing in $\tilde{\theta}_1$. A standard revealed-preference argument then shows that $a_1(\theta_1, \theta_2, \tilde{\theta}_1)$ is (weakly) decreasing in $\tilde{\theta}_1$. That establishes (A5).

To complete the proof, we define

$$N(\theta_1) \equiv E_{\theta_2}\big[B(a^*(\theta)) - h_1(a_1^*(\theta), \theta_1) - h_2(a_2^*(\theta), \theta_2)\big]$$

and set

$$\beta(\theta_1) = -N(\theta_1) + \int_{\theta_1}^{\bar{\theta}_1} E_{\theta_2}\left[\frac{\partial}{\partial\theta_1}C_1\big(a_1(t, \theta_2, t), t\big)\right] dt. \quad \text{(A6)}$$

We find that $\Gamma_1(\theta_1, \theta_1)$ is decreasing in θ_1 and that $\Gamma_1(\bar{\theta}_1, \bar{\theta}_1) = 0$. Thus the participation constraints are met and the claim follows from the revenue equivalence theorem. *Q.E.D.*

Proof of Theorem 2. It suffices to show that the second-best production assignment will not be implemented in H_2. We distinguish two cases and in both cases obtain a proof by contradiction.

Case 1. There exists $(\tilde{\theta}_1, \tilde{\theta}_2)$ such that for all (θ_1, θ_2) in a neighborhood of $(\tilde{\theta}_1, \tilde{\theta}_2)$ it is true that $a_1^*(\theta_1, \theta_2) > 0$ and $a_2^*(\theta_1, \theta_2) > 0$, where $\{a_1^*(\cdot), a_2^*(\cdot)\}$ again denote some second-best production assignments, i.e., they maximize

$$B(a) - h_1(a_1, \theta_1) - h_2(a_2, \theta_2). \quad \text{(A7)}$$

If $\{a_1^*(\cdot), a_2^*(\cdot)\}$ are also optimal in H_2, they must be maximizers in agent 1's subproblem $\mathbf{SP}(\theta_1 | x_1(a, \theta_1))$. In particular, it must be true that

$$\begin{aligned} \big(a_1^*(\theta_1, \theta_2), a_2^*(\theta_1, \theta_2)\big) &\in \underset{(a_1, a_2)}{\operatorname{argmin}}\big\{C_1(a_1, \theta_1) + h_2(a_2, \theta_2)\big\} \\ &\text{subject to } B(a_1, a_2) \geq B\big(a_1^*(\theta_1, \theta_2), a_2^*(\theta_1, \theta_2)\big). \end{aligned} \quad \text{(A8)}$$

The corresponding first-order conditions are

$$\frac{\partial}{\partial a_1} C_1(a_1^*, \theta_1) - \mu \frac{\partial}{\partial a_1} B(a_1^*, a_2^*) = 0$$

$$\frac{\partial}{\partial a_2} h_2(a_2^*, \theta_2) - \mu \frac{\partial}{\partial a_2} B(a_1^*, a_2^*) = 0,$$

where μ is the Lagrange multiplier. It follows from (A7) that $\mu = 1$ and, therefore,

$$\frac{\partial}{\partial a_1} h_1(a_1^*, \theta_1) = \frac{\partial}{\partial a_1} C_1(a_1^*, \theta_1),$$

implying that

$$\frac{F_1(\theta_1)}{f_1(\theta_1)} \cdot \frac{\partial^2}{\partial a_1 \partial \theta_1} C_1(a_1^*, \theta_1) = 0.$$

That would contradict Assumption 4.

Case 2. For almost all (θ_1, θ_2), either $a_1^*(\theta_1, \theta_2) = 0$ or $a_2^*(\theta_1, \theta_2) = 0$. We partition the space of environments into

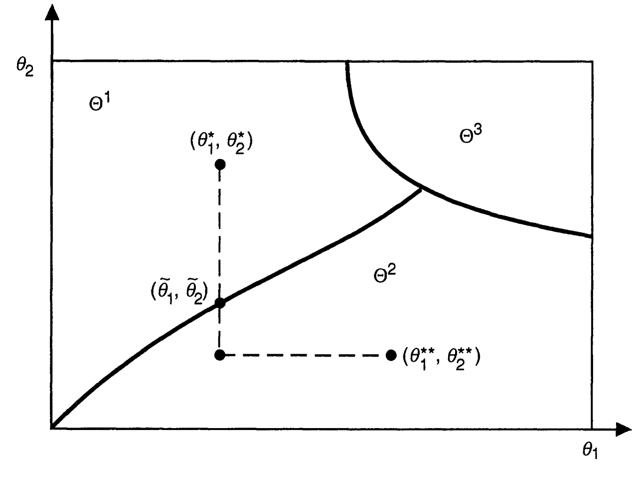
$$\Theta^1 = \left\{(\theta_1, \theta_2) \mid a_1^*(\theta_1, \theta_2) > 0\right\}$$
$$\Theta^2 = \left\{(\theta_1, \theta_2) \mid a_2^*(\theta_1, \theta_2) > 0\right\}$$
$$\Theta^3 = \left\{(\theta_1, \theta_2) \mid a_1^*(\theta_1, \theta_2) = 0 \quad \text{and} \quad a_2^*(\theta_1, \theta_2) = 0\right\}.$$

Since neither agent is dispensable, both Θ^1 and Θ^2 have nonempty interiors. Let (θ_1^*, θ_2^*) and $(\theta_1^{**}, \theta_2^{**})$ be two environments belonging to the interior of Θ^1 and Θ^2, respectively. Clearly, if $(\theta_1, \theta_2) \in \Theta^1$, then $(\theta_1 - \Delta\theta_1, \theta_2 + \Delta\theta_2) \in \Theta^1$ for any $\Delta\theta_1, \Delta\theta_2 \geq 0$ and, conversely, $(\theta_1 + \Delta\theta_1, \theta_2 - \Delta\theta_2) \in \Theta^2$ if $(\theta_1, \theta_2) \in \Theta^2$. It follows that there exists a point $(\tilde{\theta}_1, \tilde{\theta}_2)$ that is on the boundary of the sets Θ^1 and Θ^2 and either lies on the segment $(\theta_1^*, \theta_2^{**}) - (\theta_1^*, \theta_2^*)$ or on the segment $(\theta_1^*, \theta_2^{**}) - (\theta_1^{**}, \theta_2^{**})$; see Figure A1.

Suppose $(\tilde{\theta}_1, \tilde{\theta}_2)$ lies on the segment $(\theta_1^*, \theta^{**}{}_2) - (\theta_1^*, \theta_2^*)$, so that $\tilde{\theta}_1 = \theta_1^*$ (the other case is dealt with by a parallel argument). By hypothesis, the actions $(a_1^*(\theta_1^*, \theta_2^*), 0) \equiv (a_1^*, 0)$ solve (A7). At the same time, these action choices solve the minimization problem in (A8) for any environment $(\tilde{\theta}_1, \tilde{\theta}_2 + \Delta\tilde{\theta}_2)$ with $\Delta\tilde{\theta}_2 > 0$. To establish the contradiction, it suffices to show that

$$C_1(a_1^*, \tilde{\theta}_1) + h_2(0, \tilde{\theta}_2 - \Delta\tilde{\theta}_2) < C_1(0, \tilde{\theta}_1) + h_2\left(a_2^*(\tilde{\theta}_1, \tilde{\theta}_2 - \Delta\tilde{\theta}_2), \tilde{\theta}_2 - \Delta\tilde{\theta}_2\right) \qquad \text{(A9)}$$

FIGURE A1



for $\Delta\tilde{\theta}_2$ sufficiently small, where $(0, a_2^*(\tilde{\theta}_1, \tilde{\theta}_2) - \Delta\tilde{\theta}_2))$ is the solution to (A7) at $(\tilde{\theta}_1, \tilde{\theta}_2 - \Delta\tilde{\theta}_2)$. Since $(a_1^*, 0)$ solves (A8) for $(\tilde{\theta}_1, \tilde{\theta}_2 + \Delta\tilde{\theta}_2)$, we know that

$$C_1(a_1^*, \tilde{\theta}_1) + h_2(0, \tilde{\theta}_2 + \Delta\tilde{\theta}_2) \leq C_1(0, \tilde{\theta}_1) + h_2\big(a_2^*(\tilde{\theta}_1, \tilde{\theta}_2 - \Delta\tilde{\theta}_2), \tilde{\theta}_2 + \Delta\tilde{\theta}_2\big)$$

$$- \frac{F_1(\tilde{\theta}_1)}{f_1(\tilde{\theta}_1)} \cdot \frac{\partial}{\partial \theta_1}\big[C_1(a_1^*, \tilde{\theta}_1) - C_1(0, \tilde{\theta}_1)\big].$$

Since the last term on the right-hand side is a positive constant, the desired contradiction, i.e., the inequality in (iii) follows from the fact that for any given ϵ,

$$\Big|h_2\big(a_2^*(\tilde{\theta}_2 - \Delta\tilde{\theta}_2), \tilde{\theta}_2 - \Delta\tilde{\theta}_2\big) - h_2\big(a_2^*(\tilde{\theta}_2 - \Delta\tilde{\theta}_2), \tilde{\theta}_2 + \Delta\tilde{\theta}_2\big)\Big| < \epsilon$$

for $\Delta\tilde{\theta}_2$ sufficiently small. *Q.E.D.*

Proof of Theorem 3. Let $x_1(a, m)$ be an incentive scheme for H_1', and define $\hat{x}_1(a) \equiv \sup_{m\in M} \{x_1(a, m)\}$. The *ex post* participation constraint in H_1' requires that $\hat{x}_1(0, 0) \geq 0$. Let $a(\theta)$ denote the production assignments that agent 1 will implement when given the incentive scheme $\hat{x}_1(a)$ (or, equivalently the scheme $x_1(a_1, m)$). In particular, type θ_1 of agent 1 offers a contract $\{a_2(\theta_1, \cdot), x_2(\theta_1, \cdot)\}$ to agent 2, and the payments, $x_2(\cdot, \cdot)$, satisfy $E_{\theta_2}[x_2(\theta_1, \theta_2)] = E_{\theta_2}[h_2(a_2(\theta_1, \theta_2), \theta_2)]$. It must then be true that

$$E_\theta\big[\hat{x}_1\big(a(\theta)\big) - C_1\big(a_1(\theta), \theta_1\big) - h_2\big(a_2(\theta), \theta_2\big)\big] \geq E_\theta\big[\hat{x}_1(0, 0) - C_1(0, \theta_1) - h_2(0, \theta_2)\big] = E_\theta\big[\hat{x}_1(0, 0)\big] \geq 0.$$

Thus the contract $\hat{x}_1\,(a_1, a_2)$ is feasible in H_1: it is trivially incentive compatible and satisfies the interim participation constraint imposed in $\mathbf{P_1}$. Hence H_1 weakly dominates H_1'.

To establish the strict dominance of H_1 over H_1', suppose now that agent 2 is not dispensable in H_0 at $\theta_1 = \overline{\theta}_1$, and let

$$\hat{\theta}_2 = \sup\{\theta_2 \,|\, a_2^*(\overline{\theta}_1, \theta_2) > 0\},$$

where $\{a_1^*(\theta), a_2^*(\theta)\}_{\theta\in\Theta}$ denotes some second-best policy. We define

$$\hat{\Pi}_1(\theta_1, \theta_2) = \hat{x}_1\big(a^*(\theta_1, \theta_2)\big) - C_1\big(a_1^*(\theta), \theta_1\big) - h_2\big(a_2^*(\theta), \theta_2\big).$$

Clearly, $E_{\theta_2}[\hat{\Pi}_1(\theta_1, \theta_2)] = \Pi_1(\theta_1)$ for all θ_1. We note that $\hat{\Pi}_1(\cdot, \cdot)$ is continuous in both θ_1 and θ_2. For any given θ_1, $\hat{\Pi}_1(\theta_1, \cdot)$ is weakly decreasing in θ_2. Furthermore, $\hat{\Pi}_1(\theta_1, \theta_2) > \hat{\Pi}_1(\theta_1, \theta_2 + \Delta\theta_2)$, for $\Delta\theta_2 > 0$, whenever $a_2^*(\theta_1, \theta_2) > 0$.

Suppose first that $\hat{\Pi}_1(\overline{\theta}_1, \hat{\theta}_2) < 0$. We argue that type $\overline{\theta}_1$ would then implement a policy other than $\{a_1^*(\overline{\theta}_1, \theta_2), a_2^*(\overline{\theta}_1, \theta_2)\}_{\theta_2\in\Theta_2}$. Let $\tilde{\theta}_2$ be the unique value for which $\hat{\Pi}_1(\overline{\theta}_1, \tilde{\theta}_2) = 0$. Type $\overline{\theta}_1$ of agent 1 could choose the following assignments:

$$\tilde{a}_1(\overline{\theta}_1, \theta_2) = \begin{cases} a_1^*(\overline{\theta}_1, \theta_2) & \text{if } \theta_2 \leq \tilde{\theta}_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\tilde{a}_2(\overline{\theta}_1, \theta_2) = \begin{cases} a_2^*(\overline{\theta}_1, \theta_2) & \text{if } \theta_2 \leq \tilde{\theta}_2 \\ 0 & \text{otherwise.} \end{cases}$$

For this alternative policy agent 1 obtains an expected profit of

$$\int_{\underline{\theta}_2}^{\tilde{\theta}_2} \hat{\Pi}_1(\overline{\theta}_1, \theta_2)\, dF_2(\theta_2). \qquad \text{(A10)}$$

Because $\hat{\Pi}_1(\overline{\theta}_1, \theta_2) < 0$ for $\tilde{\theta}_2 < \theta_2 < \hat{\theta}_2$, it follows that the expression in (A10) is larger than $E_{\theta_2}[\hat{\Pi}_1(\overline{\theta}_1, \theta_2)]$, and therefore type $\overline{\theta}_1$ of agent 1 would not implement the second-best policy $\{a_1^*(\overline{\theta}_1, \cdot), a_2^*(\overline{\theta}_1, \cdot)\}$.

The argument just given can be extended to all types θ_1 close to $\overline{\theta}_1$. Continuity of $\hat{\Pi}_1(\cdot, \hat{\theta}_2)$ in θ_1 ensures that $\hat{\Pi}_1(\theta_1, \hat{\theta}_2) < 0$ for θ_1 sufficiently close to $\overline{\theta}_1$. It follows that for all θ_1 close to $\overline{\theta}_1$, we have $\hat{\Pi}_1(\theta_1, \hat{\theta}_2(\theta_1)) < 0$, where $\hat{\theta}_2(\theta_1) \equiv \sup\{\theta_2 \,|\, a_2^*(\theta_1, \theta_2) > 0\}$. From here on the argument is identical to that given above; if θ_1 is sufficiently close to $\overline{\theta}_1$, agent 1 will not implement the second-best policy. Thus we conclude that $\hat{\Pi}_1(\overline{\theta}_1, \theta_2) \geq 0$.

We note next that $\hat{\Pi}_1(\overline{\theta}_1, \theta_2) = \hat{\Pi}_1(\overline{\theta}_1, \hat{\theta}_2) \geq 0$ for all $\theta_2 > \hat{\theta}_2$, because $a_2^*(\overline{\theta}_1, \theta_2) = 0$ for all $\theta_2 > \hat{\theta}_2$. Since the function $\hat{\Pi}_1(\overline{\theta}_1, \cdot)$ is strictly decreasing in θ_2 on the range $[\underline{\theta}_2, \hat{\theta}_2]$, we find that

$$\Pi_1(\overline{\theta}_1) = E_{\theta_2}\left[\hat{\Pi}_1(\overline{\theta}_1, \theta_2)\right] > 0.$$

The revenue equivalence theorem then implies that principal's payoff in H_1' is less than the second-best. By Theorem 1, H_1 therefore strictly dominates H_1'. *Q.E.D.*

References

Baron, D.P. and Besanko, D. "Information, Control, and Organizational Structure." *Journal of Economics and Management Strategy,* Vol. 1 (1992), pp. 237–275.

——— and ———. "Informational Hierarchies, Self-Remedying Hidden Gaming, and Organizational Neutrality." Mimeo, Stanford University, 1994.

Bulow, J. and Roberts, J. "The Simple Economics of Optimal Auctions." *Journal of Political Economy,* Vol. 97 (1989), pp. 1060–1090.

Calvo, G.A. and Wellisz, S. "Supervision, Loss of Control, and the Optimum Size of the Firm." *Journal of Political Economy,* Vol. 86 (1978), pp. 943–952.

Crémer, J. "Manipulation by Coalitions under Asymmetric Information: The Case of Groves Mechanisms." Mimeo, Virginia Polytechnic Institute, 1994.

——— and Riordan, M.H. "On Governing Multilateral Transactions with Bilateral Contracts." *RAND Journal of Economics,* Vol. 18 (1987), pp. 436–451.

Felli, L. "Collusion in Organizations: Does Delegation Help?" Mimeo, London School of Economics, 1993.

Gilbert, R. and Riordan, M.H. "Regulating Complementary Products: A Comparative Institutional Analysis." *RAND Journal of Economics,* Vol. 26 (1995), pp. 243–256.

Green, J. and Laffont, J.-J. "On Coalition Incentive Compatibility." *Review of Economic Studies,* Vol. 44 (1979), pp. 243–254.

Guesnerie, R. and Laffont, J.-J. "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm." *Journal of Public Economics,* Vol. 25 (1984), pp. 329–369.

Laffont, J.-J. and Martimort, D. "Collusion under Asymmetric Information." Mimeo, IDEI, Université de Toulouse, 1995a.

——— and ———. "Collusion and Delegation." Mimeo, IDEI, Université de Toulouse, 1995b.

Marschak, T. and Reichelstein, S. "Network Mechanisms, Informational Efficiency, and Hierarchies." Mimeo, University of California, Berkeley, 1994.

Maskin, E. and Tirole, J. "The Principal-Agent Relationship With an Informed Principal: The Case of Private Values." *Econometrica,* Vol. 58 (1990), pp. 379–409.

McAfee, R.P. and McMillan, J. "Optimal Contracts for Teams." *International Economic Review,* Vol. 32 (1991), pp. 561–577.

——— and ———. "Organizational Diseconomies of Scale." Mimeo, University of California, San Diego, 1993.

Melumad, N., Mookherjee, D., and Reichelstein, S. "A Theory of Responsibility Centers." *Journal of Accounting and Economics,* Vol. 15 (1992), pp. 445–484.

——— and Suehiro, H. "Optimal Supervision in Hierarchies." Mimeo, Graduate School of Business, Columbia University, 1995.

Mirrlees, J. "Optimal Taxation." In K.J. Arrow and M. Intriligator, eds., *Handbook of Mathematical Economics.* New York: North-Holland, 1986.

Mookherjee, D. and Reichelstein, S. "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules." *Journal of Economic Theory,* Vol. 56 (1992), pp. 378–399.

——— and ———. "Budgeting and Hierarchical Control." Mimeo, University of Vienna, 1995.

Myerson, R.B. "Optimal Auction Design." *Mathematics of Operations Research,* Vol. 6 (1981), pp. 58–73.

———. "Mechanism Design by an Informed Principal." *Econometrica,* Vol. 51 (1983), pp. 1767–1797.

Radner, R. "The Organization of Decentralized Information Processing," *Econometrica,* Vol. 61 (1993), pp. 1109–1146.

——— and Van Zandt, T. "Information Processing and Returns to Scale." *Annales d'Economie et de Statistique,* No. 25–26 (1992), pp. 265–298.

Rogerson, W.P. "Overhead Allocation and Incentives for Cost Minimization in Defense Procurement." *Accounting Review,* Vol. 67 (1992), pp. 671–690.

Tirole, J. "*The Theory of Industrial Organization.* Cambridge, Mass.: MIT Press, 1988.

———. "Collusion and the Theory of Organizations." In J.-J. Laffont, ed., *Advances in Economic Theory, Sixth World Congress of the Econometric Society.* Cambridge: Cambridge University Press, 1992.

WILLIAMSON, O.E. "Hierarchical Control and Optimum Firm Size." *Journal of Political Economy,* Vol. 75 (1967), pp. 123–138.
———. *Markets and Hierarchies.* New York: The Free Press, 1975.
———. *The Economic Institutions of Capitalism.* New York: The Free Press, 1985.