A theory of responsibility centers*

Nahum Melumad
Stanford University, Stanford, CA 94305-5015, USA

Dilip Mookherjee
Indian Statistical Institute, New Delhi, India

Stefan Reichelstein
University of California, Berkeley, CA 94720, USA

We consider a principal–agent model to examine the effectiveness of responsibility centers, in particular cost or profit centers. We show that rather than contracting with each agent directly, the principal can create equally powerful incentives by setting up a responsibility center structure. The principal contracts with only the 'manager' of the center and delegates contracting with other agents and coordinating their activities. The principal then must monitor some measure of financial performance such as the center's cost or profit. We also find that responsibility centers dominate direct contracting with the agents when communication is limited.

1. Introduction

Responsibility centers are a common feature of large organizations.[1] Familiar types of responsibility centers are profit centers, cost centers, and investment centers. Common to the different types of responsibility centers is that certain decisions are delegated to the center's management with instructions to optimize

Correspondence to: Nahum Melumad, Graduate School of Business, Stanford University, Stanford, CA 94305-5015, USA.

*We thank the referee, Stanley Baiman, and the editor, Ray Ball, for their many comments and suggestions. We are also grateful to the seminar participants at Boston University, University of California–Berkeley, Carnegie Mellon University, London School of Economics, University of Minnesota, New York University, University of Pennsylvania, Northwestern University, Stanford University, University of Texas–Austin, and Yale University for their comments.

[1] Horngren and Foster (1991) define responsibility centers as 'parts, segments, or subunits of an organization whose managers are accountable for specific sets of activities'. See also Higgins (1952), Prince (1975), and Kaplan and Atkinson (1989) for a detailed discussion of the various types of responsibility centers observed in large firms.

Published in: *Journal of Accounting and Economics*

some financial performance variable like cost, profit, or revenue. For example, a cost center is often given an exogenous production (or service provision) target. The center's manager has discretion over 'internal' decisions, e.g., production assignments within the center, input sourcing, and make-or-buy decisions. The general objective for the manager of a cost center might be to minimize a given measure of cost. Often, the center's performance is evaluated on the basis of cost achieved, and managerial rewards such as cash bonuses, promotions, and dismissals are tied to this performance measure.

We analyze a formal model of responsibility centers and compare them to other organizational arrangements. We consider a team production setting including a principal and two agents. Each agent's production cost is known only to that agent. As a consequence, the principal incurs certain agency costs. We show that responsibility centers can be an effective means of economizing on these agency costs.

We model a responsibility center as an organizational unit that comprises the two agents.[2] The principal contracts with only one of the agents, say agent 1, who may be viewed as the 'manager' of the center. Agent 1, in turn, contracts with agent 2 and thereby determines the allocation of production within the center. This arrangement amounts to a three-tier hierarchy involving delegation of contracting. We model a cost center as an arrangement wherein the principal verifies the delivery of some aggregate output without observing the agents' individual contributions. In addition, the principal monitors the cost incurred by the center and uses this variable to create incentives for agent 1. We find that a cost center can provide optimal incentives for the agents if the organization's aggregate output level is fixed exogenously. The key to this result is that by monitoring the cost incurred by the center, the principal can affect agent 1's incentives so that agent 1 will internalize the principal's objective in determining production assignments.

Our analysis also shows that the incentive scheme for agent 1 can be linear in the observed cost. Effectively, agent 1 will bear a fraction of all cost overruns relative to some cost target. We establish an explicit formula for the optimal cost-share parameter in terms of the underlying agency characteristics. Finally, we find that under certain conditions there is no need to disaggregate the costs incurred by the center. The principal is no worse off by conditioning agent 1's performance evaluation on aggregate cost rather than on the individual line items.

It follows from the Revelation Principle that a responsibility center (more generally, any three-tier hierarchy) can at best replicate the performance of an optimal revelation mechanism.[3] The latter mechanism can be viewed as two-tier

[2]The agents themselves may be individuals or organizational subunits. In our analysis, they are taken to be primitive building blocks of the organization design problem.

[3]A general version of the Revelation Principle is given in Myerson (1982). Consider any mechanism in which agents report selected pieces of information, possibly to other agents (rather than the

organizations in which all agents simultaneously report their entire information to the principal. All decisions are then determined according to a comprehensive contract (a 'grand contract') to which the principal is committed. One implication of the Revelation Principle is that as long as one accepts its assumptions a responsibility center arrangement can never emerge as superior to the centralized two-tier arrangement corresponding to an optimal revelation mechanism.[4]

To generate a demand for responsibility centers we remove the assumption of unlimited, costless communication. Management textbooks generally assert that decision making is delegated within firms because communication of relevant information is costly.[5] We incorporate the idea of costly communication by requiring that agents communicate their information by selecting messages from a message set that is not large enough to permit full revelation of all private information. This notion appears to be fairly descriptive. Frequently, decisions and incentive schemes are based on internal accounting numbers such as product cost reports or divisional income statements that condense large amounts of information into a few numbers. We do not, however, explicitly model the nature of underlying communication, contracting, or information processing costs. Instead, we explore the implications of an exogenous limitation on the size of the message sets.

When communication is limited, a three-tier hierarchy will have a flexibility advantage over the centralized two-tier arrangement. Agent 1 can use the exact information about his own environment when allocating the production tasks between agent 2 and himself. In contrast, in a centralized mechanism, the principal has to base those decisions on 'coarse' reports of agents 1 and 2. At the same time, it will no longer be possible for the principal to align agent 1's incentives with his own preferences. Even when the principal monitors the cost (or profit) of the center, there will be a 'control loss' when agent 1 assigns production to agent 2. This control loss emerges because of limited communication between the principal and agent 1. Under certain conditions, we find that the flexibility gain outweighs the control loss, generating a demand for

---

principal) and possibly in a sequential rather than simultaneous fashion. Furthermore, suppose that the organization's decisions are delegated to different agents. Corresponding to any equilibrium of this mechanism there exists a revelation mechanism in which the principal commits himself to 'mimic' the equilibrium behavior of the agents in their original mechanism. As a consequence, agents will find it in their interest to reveal their information truthfully and the decisions made by the principal will be the same as in the equilibrium of the original mechanism.

[4] The Revelation Principle has also made it difficult to explain the emergence of cost allocation schemes [see, for instance, Baiman and Noel (1985) and Rajan (1988)] or transfer pricing mechanisms [see, for example, Amershi and Cheng (1990), Christensen and Demski (1990), Ronen and Balachandran (1988), and Vaysman (1992)].

[5] To cite just one such viewpoint, Ralph Cordiner (1956), a former president of GE, states: 'Unless we could put the responsibility and authority for decision making close to the scene of the problem, where *complete understanding* and prompt actions are possible, the company would not be able to compete . . .'

a responsibility center vis-a-vis a centralized two-tier hierarchy. Hence, responsibility centers become a means of economizing on agency costs when there are constraints on communication for contracting purposes.

The problem of delegation has been addressed in previous research on management control; see, for instance, Demski, Patell, and Wolfson (1984), Demski and Sappington (1987), Melumad and Reichelstein (1987), Penno (1988), and Ramakrishnan (1990).[6] These papers study the effectiveness of alternative delegation arrangements for organizations involving a single agent. However, in single-agent settings it is impossible to explore the issue of delegated contracting that seems essential to the understanding of responsibility centers.

The present paper is also related to the work of Alchian and Demsetz (1972). In the context of team production, these authors explore the value of making one agent the 'boss'. As in our model, Alchian and Demsetz assume that the principal can observe only the aggregate output. Unlike the present model, however, Alchian and Demsetz assume that each agent observes the other agent's production contribution in addition to his own. Furthermore, the setting considered by Alchian and Demsetz (1972) is one of complete information and therefore issues of communication do not arise.

Another relevant strand of literature is the economics of hierarchies. Cremer (1979), Geanakaplos and Milgrom (1985), Marschak and Reichelstein (1992), and Radner (1990) study the question of how to arrange agents in a hierarchy or, more generally, in a network mechanism. These papers emphasize the issues of communication and information processing in organizations. How should the agents interact to solve a given problem in the shortest possible time and/or with minimal communication requirements? All of these studies ignore motivational issues. Agents are assumed to carry out their instructions faithfully and hence there is no need for contracting and incentives.[7]

The main precursor of the present paper is Melumad, Mookherjee, and Reichelstein (1991) (hereafter referred to as MMR).[8] The paper studies contractual hierarchies and exhibits a performance loss associated with three-tier hierarchies relative to two-tier hierarchies. The contractual settings in MMR are similar to the ones in the present paper, except that the principal does not monitor the payments that agent 1 makes to agent 2. Agent 1 may then be viewed as a 'prime contractor', whose reward is the difference between the amount he is paid by the principal and what he pays his 'subcontractors'. The

[6] See Baiman (1990) for a recent survey of this literature.

[7] Calvo and Wellisz (1978) consider hierarchical organizations in which agents need to supervise their subordinates in order to alleviate moral hazard problems. In their model, however, there is no coordination issue. All information is publicly known, and the agents' production processes are independent of one another.

[8] Another related paper is McAfee and McMillan (1992), who examine the performance loss associated with multi-tier hierarchies.

analysis in MMR does show that the principal can mitigate the control loss associated with a three-tier hierarchy by monitoring agents' individual production contributions, as determined by agent 1. Monitoring payments rather than physical quantities has the general advantage that the former can be aggregated across different layers of the organization. For our present model, though, it remains an open question as to what form of monitoring is more effective.

The remainder of this paper is organized as follows. Section 2 presents the model. In section 3 we consider a setting where the total output is determined exogenously and the only issue is how to distribute production between the two agents. First, in subsection 3.1 we analyze the two-tier arrangement corresponding to an optimal revelation mechanism. In Theorem 1 we show that the performance of the optimal revelation mechanism can also be attained by a cost center. Limited communication is introduced in subsection 3.2. The main result (Theorem 2) is that with limited communication a cost center dominates a centralized, two-tier arrangement. In section 4 we study settings where the total output is determined endogenously. For those settings we establish results similar to those of Theorems 1 and 2 if one expands the notion of responsibility center from cost to profit center. We conclude in section 5 suggesting directions for future work.

2. The model

The basic scenario considered in this paper involves a principal (headquarters) and two agents (departmental managers). Each agent i produces an output a_i that belongs to a set A_i of possible output choices. For simplicity, we shall assume that A_i is an interval of real numbers, though the reader may check that several of our results will continue to hold if A_i were discrete or multi-dimensional. For any combination of (a_1, a_2), the principal receives a monetary benefit $B(a_1, a_2)$. Production of $a_i \in A_i$ involves a known and observable cost $e_i(a_i)$, which we think of as a payment to external suppliers. In addition, production of $a_i \in A_i$ requires an unobservable cost $D_i(a_i, \theta_i)$ borne by agent i. The magnitude of this cost depends on the state of agent i's production environment. This state is represented by the random variable θ_i lying in an interval $\Theta_i = [\underline{\theta}_i, \bar{\theta}_i]$; the realization of θ_i is known only to agent i.

Throughout our analysis, we assume that the variables θ_i are drawn independently from prior probability distributions $F_i(\theta_i)$ with density functions $f_i(\theta_i)$. Agents can be compensated for the costs they incur by a transfer payment x_i. Depending on the organizational structure, this payment can be made by the principal or by another agent. Finally, all parties are assumed to be risk-neutral. Agent i maximizes the expected value of $x_i - D_i(a_i, \theta_i)$, while the principal maximizes the (expected) difference between his gross benefit $B(a_1, a_2)$ and the payments made to the agents and the external suppliers.[9]

Though the above model amounts to a standard adverse selection problem, it may be useful to sketch a number of alternative interpretations of the model. One natural setting is that of team production, where the agents produce components for a system. The variable a_i could then represent quantity or quality characteristics of agent i's contribution. The two agents may produce their components simultaneously or sequentially. For instance, a_2 may refer to an intermediate product that agent 2 transfers to agent 1. Agent 1 then converts this intermediate good into the final output $B(a_1, a_2)$. Another interpretation of our model is that the agents' physical production contributions are fixed, but that a_i measures the time it takes agent i to complete his job (since a_i is a 'good', it measures the reduction in delivery time relative to some benchmark).[10] In this case, if the agents carry out their production sequentially, the benefit function takes the special form $B(a_1, a_2) = B(a_1 + a_2)$, i.e., the benefit to the principal depends on the aggregate time taken by the two agents to deliver the product. Essential to the different interpretations of our model is that for a given level of final output the agents' contributions are substitutes.

The cost $D_i(a_i, \theta_i)$ can be interpreted as disutility associated with the managerial effort needed to produce a_i in state θ_i.[11] The payment x_i then represents a bonus to compensate the agent for his disutility. Alternatively, one can interpret $x_i - D_i(\theta_i, a_i)$ as 'slack'. To cover certain expenses, an agent is given a budget x_i. The amount $D_i(\theta_i, a_i)$ represents the actually needed expenses, while the residual (i.e., the slack) can be spent on items from which the agent derives personal utility.[12]

Our model also applies to contracting situations outside the firm. Consider, for example, a situation in which a government hires one firm, labeled agent 1, to supply a system. For simplicity, suppose the system is acquired under a fixed-price contract except for a major component which is procured under a cost-based contract.[13] This component could either be supplied by agent 1 or by

[9] We note that the external suppliers play no particular role in the contracting process. They have been introduced only to illustrate certain aggregation results (see the corollary to Theorem 1).

[10] Harrison, Holloway, and Patell (1989) discuss performance evaluation schemes based on delivery time in the semiconductor industry.

[11] Though the model presented here is one of adverse selection only, it also applies to settings involving asymmetric information and moral hazard. Suppose agent i produces a_i according to a production function $a_i = \phi(w_i, \theta_i)$, where w_i denotes 'effort'. To exert effort level w_i, agent i incurs an unobservable cost $V_i(w_i)$. Provided that $\phi(\cdot,\cdot)$ is monotone increasing in w_i, we can eliminate the effort variable by defining $D_i(a_i, \theta_i) = V_i(\tau(a_i, \theta_i))$ where $\phi(\tau(a_i, \theta_i), \theta_i) \equiv a_i$.

[12] This interpretation underlies the work of Antle and Eppen (1985). See also McAfee and McMillan (1992) and the references contained therein.

[13] Generally, the government uses cost-based contracts if there are substantial uncertainties regarding project costs and the contractor's *ex ante* cost calculation is not considered reliable.

agent 2, who would become a subcontractor. Let T be a target cost for this component, and let $T - a_i$ be the actual cost as verified by the government auditors. Hence, a_i represents the cost reduction that firm i achieves, provided it is the supplier. In this context, $D_i(a_i, \theta_i)$ measures the value of foregone project 'perks', such as expanded purchases of equipment or excessive tests and experimentation. This interpretation is similar to the one in Laffont and Tirole (1986).

In the derivation of our results below, we invoke some of the following assumptions:

A.1′: $D_i(a_i, \theta_i)$ is increasing in both variables; the function $\partial^2 D_i(a_i, \theta_i)/\partial a_i \partial \theta_i$ is nonnegative and (weakly) increasing in θ_i.[14]

A.1: $D_i(a_i, \theta_i) = L_i(\theta_i) \cdot K_i(a_i)$, with $L_i(\theta_i)$ increasing and convex. We may then choose the units for a_i so that $K_i(\cdot)$ becomes the identity function, i.e., $D_i(a_i, \theta_i) = L_i(\theta_i) \cdot a_i$.

A.2′: $F_i(\theta_i)/f_i(\theta_i)$ is increasing in θ_i.

A.2: $(L_i'(\theta_i)/L_i(\theta_i)) \cdot (F_i(\theta_i)/f_i(\theta_i))$ is increasing in θ_i.

Assumptions A.1′ and A.2′ are standard in adverse selection models [see, for example, Guesnerie and Laffont (1984)]. The 'single-crossing' condition $\partial^2 D_i(a_i, \theta_i)/\partial a_i \partial \theta_i \geq 0$ implies that higher types will be asked to produce less output a_i, since the marginal cost of producing a_i increases everywhere as θ_i increases. Assumption A.2′ says that the 'inverse hazard rate' of the distribution $F_i(\cdot)$ has to be increasing in θ_i. It is well known that this condition is satisfied for many of the commonly considered probability distributions. The significance of this condition and the condition that $\partial^2 D_i(a_i, \theta_i)/\partial a_i \partial \theta_i$ be increasing in θ_i is that one can solve for the optimal incentive mechanism by recognizing only the local incentive compatibility constraints. It turns out that the solution obtained will then also be globally incentive compatible.

The multiplicative separability assumption in A.1 will simplify parts of our analysis since it implies that the cost of producing a_i increases uniformly for higher types. Finally, the significance of assumption A.2 is that linear compensation schemes will be optimal in our model. A sufficient condition for A.2 to be satisfied is that the inverse hazard rate of $F_i(\cdot)$ is increasing (i.e., A.2′ holds) and that the function $L_i(\cdot)$ is log-convex [i.e., the function $(\ln \circ L_i)(\theta_i)$ is a convex function]. Obviously, though, it is possible for A.2 to hold even though one of the two ratios is not monotone increasing.

[14] From here on, we shall refer to a function as 'increasing' whenever it is weakly increasing.

In the subsequent analysis, we invoke assumptions A.1′ and A.2′ to characterize the solution to the centralized contracting problem with unlimited communication. In order to obtain our results regarding responsibility centers, however, we rely on the stronger assumptions A.1 and A.2.

3. Centralized contracting versus cost center

3.1. Unlimited communication

We first consider the special setting where the level of desired output is fixed exogenously at some level $\bar{B}$. The only issue then is the distribution of production assignments between the two agents. This setting is natural for the study of cost centers, where the cost target is given and the task for the center is to produce this output at minimum cost.

The Revelation Principle implies that the performance of an optimal revelation mechanism provides an upper bound for the performance attainable under any organizational arrangement. A direct revelation mechanism specifies the two agents' production assignments and compensation payments for all conceivable reports made by the agents concerning the realization of their cost variables θ_i. Subsequently, the principal verifies that each agent delivers the production assignment specified by the mechanism. Formally, the principal chooses $\{a_i(\theta_1, \theta_2), x_i(\theta_1, \theta_2)\}$ so as to solve the following centralized contracting problem:[15]

$$\text{P.1} \qquad \min_{\substack{a_i(\theta_1, \theta_2) \\ x_i(\theta_1, \theta_2)}} \sum_{i=1}^{2} \mathrm{E}_\theta[e_i(a_i(\theta_1, \theta_2)) + x_i(\theta_1, \theta_2)],$$

subject to: for all $\theta_i \in \Theta_i$, $1 \leq i \leq 2$,

$$\text{(i)} \quad \theta_i \in \operatorname*{argmax}_{\tilde{\theta}_i} \mathrm{E}_{\theta_j}[x_i(\theta_j, \tilde{\theta}_i) - D_i(a_i(\theta_j, \tilde{\theta}_i), \theta_i)],$$

$$\text{(ii)} \quad \mathrm{E}_{\theta_j}[x_i(\theta_j, \theta_i) - D_i(a_i(\theta_j, \theta_i), \theta_i)] \geq 0,$$

$$\text{(iii)} \quad B(a_1(\theta_1, \theta_2), a_2(\theta_1, \theta_2)) \geq \bar{B}.$$

[15] We adopt the notation $\mathrm{E}_\theta[\cdot] = \int_{\underline{\theta}_1}^{\bar{\theta}_1} \int_{\underline{\theta}_2}^{\bar{\theta}_2} [\cdot] \, \mathrm{d}F_2(\theta_2) \mathrm{d}F_1(\theta_1)$ and $\mathrm{E}_{\theta_j}[\cdot] = \int_{\underline{\theta}_j}^{\bar{\theta}_j} [\cdot] \, \mathrm{d}F_j(\theta_j)$.

The first constraint in P.1 is the incentive constraint requiring that truthful reporting by both agents constitutes a Bayesian Nash equilibrium.[16] Constraint (ii) represents a participation constraint, where each agent's outside opportunity payoff is normalized to zero. The interpretation of this constraint is that each agent has to decide whether or not to enter into a contract *after* learning the realization of his cost variable θ_i. Alternatively, the constraint may reflect a situation where agents receive their information after contracting, but cannot be prevented from quitting *ex post.*[17] The third constraint says that the production assignments must attain the fixed benefit $\bar{B}$.

Previous research has obtained the following characterization for the solution to P.1. For any production policy $\{a_1(\theta_1, \theta_2), a_2(\theta_1, \theta_2)\}$ that the principal may choose, the 'local' incentive compatibility and participation constraints fully determine the compensation payments for each agent. To see this, define the expected utility payoff of type θ_i of agent i to be

$$\Gamma_i(\theta_i) = \mathrm{E}_{\theta_j}[x_i(\theta_j, \theta_i) - D_i(a_i(\theta_j, \theta_i), \theta_i)]. \tag{1}$$

In order for θ_i to have an incentive to report truthfully, $\Gamma_i(\theta_i)$ must be at least as large as what θ_i could obtain by claiming to be type $\theta_i + \Delta\theta_i$,

$$\Gamma_i(\theta_i) \geq \Gamma_i(\theta_i + \Delta\theta_i) + \mathrm{E}_{\theta_j}[D_i(a_i(\theta_j, \theta_i + \Delta\theta_i), \theta_i + \Delta\theta_i) - D_i(a_i(\theta_j, \theta_i + \Delta\theta_i), \theta_i)].$$

Letting $\Delta\theta_i \to 0$, we obtain

$$\Gamma_i'(\theta_i) \geq -\mathrm{E}_{\theta_j}\left[\frac{\partial}{\partial\theta_i} D_i(a_i(\theta_j, \theta_i), \theta_i)\right].$$

The reverse inequality follows from the fact that type $\theta_i + \Delta\theta_i$ must be discouraged from reporting to be type θ_i, and hence

$$\Gamma_i'(\theta_i) = -\mathrm{E}_{\theta_j}\left[\frac{\partial}{\partial\theta_i} D_i(a_i(\theta_j, \theta_i), \theta_i)\right], \tag{2}$$

or equivalently,

$$\Gamma_i(\theta_i) - \Gamma_i(\bar{\theta}_i) = \mathrm{E}_{\theta_j}\left[\int_{\theta_i}^{\bar{\theta}_i} \frac{\partial}{\partial\theta_i} D_i(a_i(\theta_j, t), t)\,\mathrm{d}t\right]. \tag{3}$$

[16] An alternative formulation is to require that truthful reporting be a dominant strategy for each agent. As shown in Mookherjee and Reichelstein (1992), the dominant strategy and Bayesian formulations yield the same expected payoff to the principal in the present model.

[17] See Melumad (1989) for a discussion of agency models involving quitting.

Eq. (3) shows that type θ_i will earn an informational rent, i.e., his utility payoff will exceed the minimum amount given by the participation constraint (zero in our model). Higher-cost types will earn lower rents and for an optimal mechanism the participation constraint will be binding for the worst type, that is, $\Gamma_i(\bar{\theta}_i) = 0$.[18] Combining eqs. (1) and (3), we obtain

$$\mathrm{E}_{\theta_j}[x_i(\theta_j, \theta_i)] = \mathrm{E}_{\theta_j}\left[D_i(a_i(\theta_j, \theta_i), \theta_i) + \int_{\theta_i}^{\bar{\theta}_i} \frac{\partial}{\partial \theta_i} D_i(a_i(\theta_j, t), t)\,\mathrm{d}t \right]. \quad (4)$$

Eq. (4) shows that, given a production schedule $\{a_1(\theta_1, \theta_2), a_2(\theta_1, \theta_2)\}$, the payments to the agents are fully determined by the local incentive constraints ($\Delta\theta_i$ was taken to be small in the above argument) and the participation constraints. This result is sometimes referred to as the *revenue equivalence theorem.*[19] We note that the magnitude of the agents' rents depends on the production assignments chosen. The second term on the right-hand side of (4) can be integrated by parts, resulting in

$$\mathrm{E}_{\theta}[x_i(\theta_j, \theta_i)] = \mathrm{E}_{\theta}[D_i(a_i(\theta_j, \theta_i), \theta_i) + \frac{F_i(\theta_i)}{f_i(\theta_i)} \cdot \frac{\partial}{\partial \theta_i} D_i(a_i(\theta_j, \theta_i), \theta_i)].$$

Let

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

This function is usually referred to as agent i's 'virtual cost': the sum of the agent's production cost and an informational cost which is required to prevent misrepresentation of private information. Ignoring the global incentive constraints, the optimization problem in P.1 can then be restated as

$$\min_{a_1(\cdot),\, a_2(\cdot)} \sum_{i=1}^{2} \mathrm{E}_{\theta}\,[e_i(a_i(\theta_1, \theta_2)) + h_i(a_i, \theta_i)], \quad (5)$$

[18] If the function $a_i(\theta_j, t)$ were constant and equal to $\bar{a}$ on the interval $[\theta_i, \bar{\theta}_i]$, the rent expression in (3) would amount to $\Gamma_i(\theta_i) = D_i(\bar{a}, \theta_i) - D_i(\bar{a}, \theta_i)$ (i.e., type θ_i would have to be paid the cost of the highest-type θ_i in order to tell the truth). In general, though, an agent's rent will be smaller since the principal can induce the agent to self-select by procuring less output from higher-cost types.

[19] This term has been coined in the auction literature. Myerson (1981) and others have shown that any two auction mechanisms, which in equilibrium involve the same allocation rules for selling the object to a particular bidder, must yield the same expected revenue for the principal.

subject to

$$B(a_1(\theta_1, \theta_2), a_2(\theta_1, \theta_2)) \geq \bar{B}.$$

The problem in (5) can be solved pointwise. For each environment (θ_1, θ_2), the optimal production assignments satisfy

$$\min_{a_1, a_2} \sum_{i=1}^{2} [e_i(a_i) + h_i(a_i, \theta_i)], \tag{6}$$

subject to

$$B(a_1, a_2) \geq \bar{B}.$$

Assumptions A.1′ and A.2′ ensure that the virtual cost $h_i(a_i, \theta_i)$ is increasing in both arguments. As a consequence, the functions $a_i(\theta_j, \cdot)$, which solve (6), will be monotone decreasing in θ_i. It can be shown [see, for instance, Guesnerie and Laffont (1984)] that the resulting mechanism is also globally incentive-compatible and hence we have characterized the solution to P.1. We shall refer to the production assignments solving P.1 [or the minimization problem in (6)] as the second-best production assignments.

Before we turn to our study of cost centers, it will be useful to consider the following hierarchical contracting arrangement. The principal first communicates and contracts with agent 1 who, in turn, communicates and contracts with agent 2. If in the subcontract agent 2 agrees to produce a_2, agent 1 has to produce the residual quantity a_1, so that $B(a_1, a_2) = \bar{B}$.[20] The principal only verifies delivery of the (exogenously fixed) output $\bar{B}$. He does not monitor the terms of the contract between agent 1 and agent 2, nor does he observe the production contributions a_i made by the two agents. Such an arrangement is consistent with the general notion that higher-level contracts in the organization are expressed in terms of aggregates without specifying contingencies for all lower-level decisions. Upon delivery of the output $\bar{B}$, the principal pays agent 1, who then pays agent 2 and the external suppliers 'out of his own pocket'. Thus agent 1 becomes a residual claimant, while the principal pays a fixed price. Depending on his type θ_1, agent 1 will choose an incentive scheme

[20] This setup is related to that of Demski and Sappington (1989) who study a three-tier hierarchy where a supervisor can generate, at a personal cost, information about his subordinate's environment. Under certain conditions, the principal wants to condition the supervisor's compensation on the worker's performance measure.

for agent 2 $\{a_2(\theta_2 | \theta_1), x_2(\theta_2 | \theta_1)\}$ which solves

$$\min_{x_2(\cdot),\, a(\cdot)} \mathrm{E}_{\theta_2}[D_1(a_1(\theta_2 | \theta_1), \theta_1) + x_2(\theta_2 | \theta_1)$$

$$+ e_1(a_1(\theta_2 | \theta_1)) + e_2(a_2(\theta_2 | \theta_1))],$$

subject to the incentive and participation constraints for agent 2 and the requirement that $B(a_1(\theta_2 | \theta_1), a_2(\theta_2 | \theta_1)) \geq \bar{B}$. Using again the fact that in any incentive compatible mechanism agent 2's expected payment must equal his expected virtual cost, agent 1's problem amounts to the pointwise minimization problem:

$$\min_{(a_1, a_2)} \left\{ \sum_{i=1}^{2} e_i(a_i) + D_1(a_1, \theta_1) + h_2(a_2, \theta_2) \right\}, \tag{7}$$

subject to

$$B(a_1, a_2) \geq \bar{B}.$$

Comparison of (6) and (7) shows that the three-tier hierarchy involves a distortion of the production assignments: Agent 1 imputes too low a cost for himself, since he takes into account $D_1(\cdot\,,\cdot)$ rather than $h_1(\cdot\,,\cdot)$. As a consequence, agent 1's production contribution will be too large relative to the second-best production assignments. This reflects the monopoly power commanded by agent 1 in the delegated contracting arrangement. We subsequently refer to this distortion as the 'control loss'. The principal's expected cost will increase since production assignments chosen according to (7) cannot at the same time minimize the objective function in (6), which represents the principal's expected payments to the agents and the external suppliers.[21]

To further illustrate the control loss associated with the above three-tier hierarchy, we consider the following example, which will also be useful at later stages of the analysis.

Example: Consider a situation where the agents' production contributions are perfect substitutes, and $B(a) = a_1 + a_2$, $\bar{B} = 1$. Assume the two agents are *ex ante* symmetric, i.e., $F_1(\cdot) = F_2(\cdot) = F(\cdot)$ and $D_i(a_i, \theta_i) = \theta_i \cdot a_i$. Virtual cost

[21] The proof of this claim is a direct consequence of theorem 3.1 in MMR (1991). The proof relies on the assumption that the functions $e_i(\cdot)$ and $B(\cdot\,,\cdot)$ are differentiable.

can then be expressed as

$$h_i(a_i, \theta_i) \equiv a_i \cdot \alpha(\theta_i) \quad \text{where} \quad \alpha(\theta_i) \equiv \theta_i + \frac{F(\theta_i)}{f(\theta_i)}.$$

Suppose, in addition, that $e_i(a_i) = p \cdot a_i$, i.e., for each unit of a_i the external suppliers are paid a constant price of p. As a consequence, the payments of the external suppliers play no role in the allocation of production assignments.

For this example, the minimization program in (7) is linear in a_1 and a_2. Hence, agent i will be assigned to produce the entire amount if and only if $\alpha(\theta_i) \leq \alpha(\theta_j)$. Equivalently,

$$a_1(\theta_1, \theta_2) = 1 \quad \text{if and only if} \quad \theta_2 \geq \theta_1,$$

i.e., the agent with the lower cost realization will be assigned the entire production responsibility.

In the three-tier hierarchy, on the other hand, agent 1 chooses a_1 and a_2 in order to minimize $a_1 \cdot \theta_1 + a_2 \cdot \alpha(\theta_2)$. An elementary way of verifying this claim is to note that, in the subcontract, agent 1 makes agent 2 a take-it-or-leave-it offer to produce $a_2 = 1$ for a fixed price P. This price is chosen to maximize agent 1's expected payoff: $P \cdot F(P) + \theta_1 \cdot (1 - F(P))$. The first-order condition then requires that P satisfies $\theta_1 = P + F(P)/f(P)$ or $P = \alpha^{-1}(\theta_1)$. Hence, the allocation rule under this three-tier hierarchy involves

$$a_1(\theta_1, \theta_2) = 1 \quad \text{if and only if} \quad \theta_2 \geq \alpha^{-1}(\theta_1).$$

The resulting production assignments are inefficient whenever $\theta_1 \geq \theta_2 \geq \alpha^{-1}(\theta_1)$ (the hatched region in fig. 1). Within this region, agent 1 refuses to let agent 2 be the producer, despite agent 2 being the lower cost supplier.

Our notion of a *cost center* is similar to the three-tier hierarchy just described, except for additional monitoring by the principal. Again, the principal first contracts with agent 1 and delegates to him the task of determining the production contributions a_1 and a_2 via the subcontract with agent 2. In contrast to the previous three-tier arrangement, however, agent 1 no longer pays agent 2 and the external suppliers out of his own pocket. Instead, agent 1 is given authority to make these payments from the principal's account (i.e., he is given 'check writing' authority). Naturally, the principal can monitor the account balance,

$$c \equiv e_1(a_1) + e_2(a_2) + x_2,$$

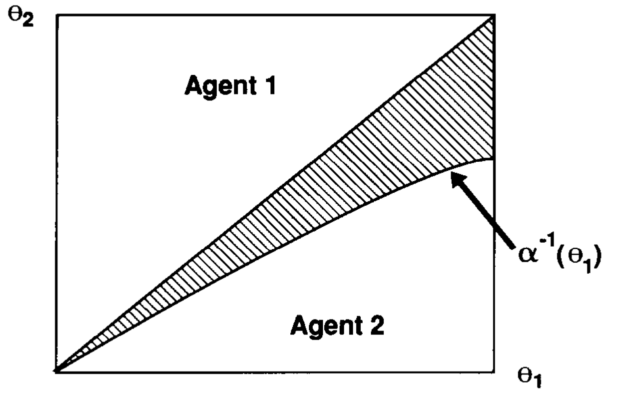

Fig. 1. Production assignments in two- and three-tier hierarchies. θ_i is agent i's production cost and $\alpha^{-1}(\cdot)$ is the inverse virtual cost.

which is the sum of all observable costs that agent 1 has control over. The cost c can be used to evaluate agent 1's performance. Formally, the principal offers agent 1 a menu of contracts taking the form $\{x_1(\theta_1, c)\}$. Thus, if agent 1 reports θ_1 in the first stage and subsequently incurs cost c, he will be paid $x_1(\theta_1, c)$ (provided he delivers the target output $\bar{B}$). For any report $\tilde{\theta}_1$ to the principal, type θ_1 of agent 1 designs an incentive contract $\{a_2(\theta_2 | \theta_1, \tilde{\theta}_1), x_2(\theta_2 | \theta_1, \tilde{\theta}_1)\}$ for agent 2, solving:

$$\text{P.2}_{\text{sub}} \quad \max_{\substack{a_1(\cdot), a_2(\cdot) \\ x_2(\cdot)}} E_{\theta_2}[x_1(\tilde{\theta}_1, c(\theta_2 | \theta_1, \tilde{\theta}_1)) - D_1(a_1(\theta_2 | \theta_1, \tilde{\theta}_1), \theta_1)],$$

subject to: for all θ_2,

$$\text{(i)} \quad \theta_2 \in \operatorname*{argmax}_{\tilde{\theta}_2} \{x_2(\theta_2 | \theta_1, \tilde{\theta}_1) - D_2(a_2(\theta_2 | \theta_1, \tilde{\theta}_1), \theta_2)\},$$

$$\text{(ii)} \quad x_2(\theta_2 | \theta_1, \tilde{\theta}_1) - D_2(a_2(\theta_2 | \theta_1, \tilde{\theta}_1), \theta_2) \geq 0,$$

$$\text{(iii)} \quad B(a_1(\theta_2 | \theta_1, \tilde{\theta}_1), a_2(\theta_2 | \theta_1, \tilde{\theta}_1)) \geq \bar{B},$$

$$\text{(iv)} \quad c(\theta_2 | \theta_1, \tilde{\theta}_1) \equiv e_1(a_1(\theta_2 | \theta_1, \tilde{\theta}_1)) + e_2(a_2(\theta_2 | \theta_1, \tilde{\theta}_1)) + x_2(\theta_2 | \theta_1, \tilde{\theta}_1).$$

Let $c(\theta_1, \theta_2) \equiv c(\theta_2 | \theta_1, \theta_1)$ and let $\Gamma(\theta_1 | x_1(\tilde{\theta}_1, c))$ denote the value of the optimization program P.2$_{\text{sub}}$ [induced by the scheme $x_1(\theta_1, c)$]. The principal's problem in the cost center arrangement then becomes:

$$\text{P.2} \qquad \min_{x_1(\cdot,\cdot)} \; E_\theta[c(\theta_1, \theta_2) + x_1(\theta_1, c(\theta_1, \theta_2))],$$

subject to: for all θ_1,

$$\text{(i)} \quad \theta_1 \in \operatorname*{argmax}_{\tilde{\theta}_1} \Gamma(\theta_1 | x_1(\tilde{\theta}_1, c)),$$

$$\text{(ii)} \quad \Gamma(\theta_1 | x_1(\theta_1, c)) \geq 0.$$

We are now in a position to state our first result which shows that a cost center can provide the same incentives as the optimal revelation mechanism.

Theorem 1. Given assumptions A.1 and A.2, a cost center arrangement can replicate the performance of the optimal revelation mechanism in P.1.

Proof. See appendix.

The proof of Theorem 1 amounts to showing that one can construct a compensation scheme $x_1(\theta_1, c)$, which induces agent 1 (the manager of the cost center) to implement the second-best production assignments. Consider incentive schemes linear in the observed aggregate cost c,

$$x_1(\theta_1, c) = b(\theta_1) - \beta(\theta_1)\cdot(c - c^*(\theta_1)). \tag{8}$$

The contracts in (8) are sometimes referred to as budget-based schemes.[22] Through his message θ_1, agent 1 selects a cost target $c^*(\theta_1)$, a cost share parameter $\beta(\theta_1)$, and a lump-sum payment $b(\theta_1)$. The function $c^*(\theta_1)$ is constructed such that, if agent 1 reports his information θ_1 truthfully, the expected cost equals $c^*(\theta_1)$ and therefore the expected budget variance is zero. Hence, $b(\theta_1)$ becomes the expected payment to agent 1. We note that agent 1's incentive to 'control' the center's cost hinges entirely on the choice of $\beta(\theta_1)$. For $\beta = 0$, agent 1 would delegate all production to agent 2, while $\beta = 1$ would make agent 1 the residual claimant and, as argued above, induce agent 1 to distort the production assignments. Consider, however, the following cost-share parameter

[22] See, for example, Kirby, Reichelstein, Sen, and Paik (1991).

with a value intermediate between zero and one:

$$\beta(\theta_1) = \frac{1}{\mu(\theta_1)} \quad \text{and} \quad \mu(\theta_1) = 1 + \frac{L_1'(\theta_1)}{L_1(\theta_1)} \cdot \frac{F_1(\theta_1)}{f_1(\theta_1)}. \tag{9}$$

We note that in light of A.2, the function $\beta(\theta_1)$ is decreasing in θ_1. This is sufficient for a menu of linear contracts to be incentive compatible. Given an incentive scheme of the form shown in (8), suppose agent 1 reports truthfully [for an appropriate choice of $b(\cdot)$ and $c^*(\cdot)$, this will turn out to be his best response]. Subsequently, agent 1 seeks to solve

$$\max_{a_1(\cdot),\, a_2(\cdot)} \quad \mathrm{E}_{\theta_2}[b(\theta_1) - \beta(\theta_1) \cdot (c(\theta_1, \theta_2) - c^*(\theta_1)) - L_1(\theta_1) \cdot a_1(\theta_1, \theta_2)],$$

subject to the four constraints listed in $\mathrm{P.2_{sub}}$. Once agent 1 has reported his private information, $b(\theta_1)$ and $c^*(\theta_1)$ are 'sunk' and his objective function effectively reduces to

$$\min_{a_1(\cdot),\, a_2(\cdot)} \quad \mathrm{E}_{\theta_2}[\beta(\theta_1) \cdot c(\theta_1, \theta_2) + L_1(\theta_1) \cdot a_1(\theta_1, \theta_2)].$$

Dividing this objective function by the constant $\beta(\theta_1)$ shows that agent 1 will be induced to choose production assignments that minimize the expected value of $c + [L_1(\theta_1)/\beta(\theta_1)] \cdot a_1$. Given the formula for $\beta(\theta_1)$ in (9), we find that $[L_1(\theta_1)/\beta(\theta_1)] \cdot a_1 = h_1(\theta_1, a_1)$, which equals agent 1's virtual cost. Since this virtual cost is multiplicatively separable, we shall write (with a slight abuse of notation) $h_i(a_i, \theta_i) \equiv h_i(\theta_1) \cdot a_i$, where $h_i(\theta_i) = L_i(\theta_i) + [F_i(\theta_i)/f_i(\theta_i)] L_i'(\theta_i)$. Any incentive compatible contract $\{a_2(\theta_2), x_2(\theta_2)\}$ for agent 2 has to satisfy $\mathrm{E}_{\theta_2}[x_2(\theta_2)] = \mathrm{E}_{\theta_2}[h_2(\theta_2) \cdot a_2(\theta_2)]$. Since $c = e_1(a_1) + e_2(a_2) + x_2$, it follows that type θ_1 of agent 1 will choose the production assignments $a_2(\theta_2 | \theta_1)$, $a_1(\theta_2 | \theta_1)$ so as to minimize

$$\mathrm{E}_{\theta_2}\left[\left\{\sum_{i=1}^{2} e_i(a_i(\theta_2 | \theta_1)) + h_2(\theta_2) \cdot a_2(\theta_2 | \theta_1)\right\} + h_1(\theta_1) \cdot a_1(\theta_2 | \theta_1)\right], \tag{10}$$

subject to

$$B(a_1(\theta_2 | \theta_1), a_2(\theta_2 | \theta_1)) \geq \bar{B}.$$

Again, this problem can be solved pointwise. For any given (θ_1, θ_2), agent 1 will choose a_1 and a_2 so as to minimize $\sum_{i=1}^{2} [e_i(a_i) + h_i(\theta_i) \cdot a_i]$ subject to

$B(a_1, a_2) \geq \bar{B}$. Comparison with (6) shows that the resulting production assignments are exactly the ones that the principal would have chosen in the grand-contracting problem P.1. Thus, agent 1 will internalize the principal's objective function precisely. To check that the principal's expected cost is indeed minimized by the cost center arrangement, recall that the principal's expected cost is given by

$$E_\theta[x_1(\theta_1, c(\theta_2 | \theta_1)) + c(\theta_2 | \theta_1)],$$

where

$$c(\theta_2 | \theta_1) = \sum_{i=1}^{2} e_i(a_i(\theta_2 | \theta_1)) + h_2(\theta_2) \cdot a_2(\theta_2 | \theta_1),$$

and $\{a_1(\theta_2 | \theta_1), a_2(\theta_2 | \theta_1)\}$ denotes a solution to (10) [or equivalently to (6)]. Hence the principal's expected cost will be the same as in P.1 if

$$E_\theta[x_1(\theta_1, c(\theta_2 | \theta_1))] = E_\theta[h_1(\theta_1) \cdot a_1(\theta_2 | \theta_1)].$$

In the proof of Theorem 1, we construct suitable functions $b(\theta_1)$ and $c^*(\theta_1)$ for the compensation scheme in (8), such that agent 1 will have an incentive to report his information truthfully and, furthermore, $E_\theta[x_1(\theta_1, c(\theta_2 | \theta_1))] = E_{\theta_1}[b(\theta_1)] = E_\theta[h_1(\theta_1) \cdot a_1(\theta_2 | \theta_1)]$.

The essential feature of the cost-share parameter $\beta(\theta_1)$ in (8) is that agent 1 'marks down' all other costs in c relative to his own unobservable cost $L_1(\theta_1) \cdot a_1$. Alternatively, agent 1 marks up his cost $L_1(\theta_1) \cdot a_1$ and for suitably chosen $\beta(\theta_1)$ the marked-up cost is exactly equal to the virtual cost $h_1(\theta_1) \cdot a_1$.[23] The principal effectively subsidizes contributions from agent 2 by choosing the cost-share parameter $\beta(\theta_1)$ less than 1. This alleviates the monopoly distortion described earlier, when the principal only monitors delivery of the output level $\bar{B}$. An interesting consequence of this argument is that the principal need not monitor the (individual) line items of the cost c [i.e., the values of $e_1(a_1)$, $e_2(a_2)$, and x_2]. Hence, line item reporting is of no value in this context.

Corollary to Theorem 1. Given assumptions A.1 and A.2, aggregation of all observable cost items is optimal.

[23] At this point the multiplicative separability assumption in A.1 is essential. Without multiplicative separability it would be impossible to ensure equality of $D_1(a_1, \theta_1)/\beta(\theta_1)$ and $h_1(a_1, \theta_1)$ for all values of a_1 and θ_1. However, for suitably chosen cost-share parameters, the production choices of the cost center may still provide a 'reasonable' approximation to those of the optimal revelation mechanism.

It follows immediately from the above reasoning that Theorem 1 and its corollary extend to an arbitrary number of agents in the cost center. The principal would still find an evaluation scheme of the form in (8) to be optimal for controlling agent 1. In fact, the formula for the cost-share parameter would remain unchanged. In order to align agent 1's interests with his own, the principal chooses the cost-share parameter based on characteristics pertaining to agent 1 only, rather than those of other agents in the center.

Previous literature on principal–agent models has exhibited conditions under which menus of linear contracts are optimal. It is well known that linear schemes have certain 'robustness' properties, insofar as the behavior induced by them is less sensitive to an exact specification of the underlying model parameters. In our context of hierarchical contracting, we find that the above analysis is robust to the introduction of noise in the observed cost c. For instance, the cost associated with the external suppliers might be random, though all parties know their expected values [e.g., $e_i(a_i) = e_i^*(a_i) + \varepsilon_i$ where ε_i is a random variable with mean zero]. Such variability in the observed cost will not matter with linear incentive schemes and risk-neutral agents.

3.2. Limited communication

In this subsection we compare cost centers with centralized contracting when there are limitations on the amount of information that agents can communicate to the principal or to the other agents in the organization. As stated in the introduction, we do not explicitly model the cost of communication and processing information. Instead, we simply impose the restriction that agents cannot reveal their private information in full detail since the space of messages available to them is smaller than the space of possible environments. Though this modeling restriction is somewhat *ad hoc*, we believe that it captures an essential feature of decision-making and contracting within firms.[24] A substantial portion of the formal communication within large organizations is channeled through the internal accounting reports, e.g., budgets, divisional P&L statements, or product cost reports. These statements generally provide a 'coarse' summary of the information relevant to a department or a division.

When communication is limited, there is a potential gain for the principal in delegating to agent 1 the authority to decide the production assignments and to contract with agent 2. Agent 1 is now in a position to make these decisions on the basis of better information about his own environment than the principal would have been able to obtain through a report by agent 1. We refer to this effect as the *flexibility gain* associated with delegation in a limited communica-

[24]Recent papers by Christensen and Demski (1990), Jordan (1989), and Vaysman (1992) have adopted a similar approach.

tion environment. In spite of this flexibility gain, delegation is not necessarily desirable because of the *control loss* associated with delegated contracting. The analysis in Theorem 1 shows that the principal needs full communication with agent 1 in order to alleviate the control loss entirely.[25] With limited communication, there will be a residual control loss that needs to be traded off against the flexibility gain.

We incorporate the notion of limited communication by requiring that agent i select messages from a finite message set $M_i \equiv \{m_i^1, \ldots, m_i^{k_i}\}$. In analogy to P.1 above, the contracting arrangement then specifies production assignments $a_i(m_1, m_2)$ and transfer payments $x_i(m_1, m_2)$ for all possible message combinations $(m_1, m_2) \in M_1 \times M_2$. In addition, the mechanism needs to specify message sending rules, which specify the message that an agent is supposed to send given his environment θ_i.[26]

It follows from the Revelation Principle that there is no loss of generality in having agents report their information simultaneously, provided there are no limitations on communication. With limited communication, it will generally be preferable to let agents send their messages sequentially. The principal has more flexibility if agent 2 sends his message in response to the information communicated by agent 1. Formally, the message sending rule for agent 1 then becomes a function $\lambda_1\colon \Theta_1 \to M_1$ and for agent 2 a function $\lambda_2\colon \Theta_2 \times M_1 \to M_2$. In our earlier paper we show that, under assumptions A.1 and A.2′, the principal may confine attention to interval partitions such that $\lambda_1(\theta_1) = m_1^u$ if and only if $\theta_1 \in (\theta_2^{u-1}, \theta_1^u]$, where $1 \leq u \leq k_1$ and $\theta_1^0 \equiv \underline{\theta}_1$. Similarly, $\lambda_2(\theta_2, m_1^u) = m_2^u$ if and only if $\theta_2 \in (\theta_1^{u,v-1}, \theta_2^{u,v}]$, for $1 \leq u \leq k_1$ and $1 \leq v \leq k_2$. This amounts to requiring agents to report which of the component intervals their environment happens to be in. The principal, nevertheless, has to choose how to partition an agent's set of possible environments into such intervals, i.e., how to partition $\Theta_1 \times \Theta_2$ into $k_1 \times k_2$ rectangles. The following result is proven in MMR.

Lemma. Suppose communication is limited with $|M_i| = k_i$. *Given assumptions A.1 and A.2′, the centralized contracting problem is equivalent to*:

$$\min_{\substack{\{\theta_1^u\},\{\theta_2^{uv}\}\\ \{a_i^{uv}\}}} \sum_{u=1}^{k_1} \sum_{v=1}^{k_2} \int_{\theta_1^{u-1}}^{\theta_1^u} \int_{\theta_2^{u,v-1}}^{\theta_2^{uv}} [h_1(\theta_1)\cdot a_1^{uv} + h_2(\theta_2)\cdot a_2^{uv}$$

$$+ e_1(a_1^{uv}) + e_2(a_2^{uv})]\, \mathrm{d}F_2(\theta_2)\, \mathrm{d}F_1(\theta_1), \tag{11}$$

[25] That is, the cost share parameter $\beta(\theta_1)$ in (8) typically varies with θ_1.

[26] With limited communication we could have considered more complex contracting arrangements in the two-tier hierarchy. For instance, the principal could delegate the choice of a_i to agent i. Agents would be asked to report their information sequentially to both the principal and the other agent, and subsequently, each agent would make his decision. The agents' compensation would be

subject to

$$B(a_1^{uv}, a_2^{uv}) \geq \bar{B}.$$

The characterization given in (11) is analogous to the one in (5). The limited communication requirement implies that the production assignments $a_i(\theta_1, \theta_2)$ have to be constant on each cell of the partition, i.e., $a_i(\theta_1, \theta_2) = a_1^{uv}$ if $\theta_1 \in [\theta_1^{u-1}, \theta_1^u]$ and $\theta_2 \in [\theta_2^{u,v-1}, \theta_2^{u,v}]$. As argued in connection with (5), incentive compatibility requires that, for *any* production schedule (in particular those that are step functions), the expected payment to an agent has to equal his expected virtual cost. However, if the production assignments $a_1(\theta_1, \theta_2)$ and $a_2(\theta_1, \theta_2)$ are constant on each rectangle, $[\theta_1^{u-1}, \theta_1^u] \times [\theta_2^{u,v-1}, \theta_2^{u,v}]$, then

$$\mathrm{E}_\theta\left[\sum_{i=1}^{2} h_i(\theta_i)\cdot a_i(\theta_1, \theta_2)\right]$$

$$= \sum_{u=1}^{k_1}\sum_{v=1}^{k_2}\int_{\theta_1^{u-1}}^{\theta_1^u}\int_{\theta_2^{u,v-1}}^{\theta_2^{uv}} [h_1(\theta_1)\cdot a_1^{uv} + h_2(\theta_2)\cdot a_2^{uv}]\,\mathrm{d}F_2(\theta_2)\,\mathrm{d}F_1(\theta_1).$$

Example (*continued*): Consider again the setting of the above example, and assume that $k_i = 2$, i.e., both agents can send one of two possible messages. Agent 1 is asked to report m_1^1 – to be interpreted as 'low cost' – if $\theta_1 < \theta_1^1$. The reporting instruction for agent 2 is conditional on agent 1's message. If agent 1 reports m_1^1, agent 2 is asked to report low cost (m_1^2) if and only if $\theta_2 \leq \theta_2^1$ for some value θ_2^1. However, if agent 1 reports m_1^2, agent 2's cut-off point changes to some value θ_2^2 (which can be shown to be larger than θ_2^1).

The optimal production assignments for each of the four cells are shown in fig. 2. Note that for either of the two messages that agent 1 may send, agent 2 will be asked to be the producer in the cell corresponding to lower values of θ_2. This reflects the principal's objective of selecting the agent with lower expected virtual cost, where the expectation is taken over all environments in the particular cell. Even if agent 1 reports low cost, agent 2's expected virtual cost will be lower in the southwest cell of fig. 2 provided $\theta_2^1 \leq \theta_1^1$.[27] Since the second-best (full communication) production assignments call for agent 1 to be the producer if and only if the environment (θ_1, θ_2) is above the diagonal, the hatched areas

a function of their messages and the decisions made. Because of its sequential nature, such an arrangement may be superior to the one considered here. The simpler setting we consider, however, is the natural representation of a centralized structure in which the principal retains control over decisions and compensation payments.

[27] It can be shown that the optimal cut-off point θ_1^1 satisfies $\theta_2^1 = \alpha^{-1}(\theta_1^1)$.

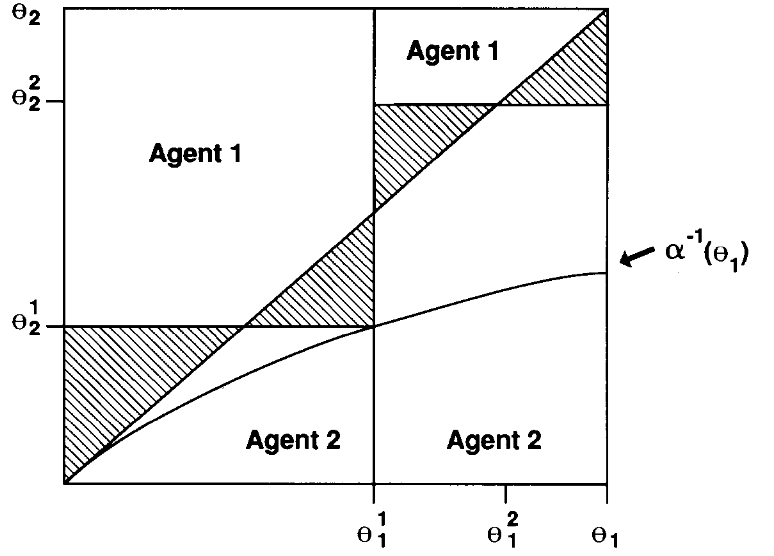

Fig. 2. Production assignments in a two-tier hierarchy with limited communication. θ_i is agent i's production cost, θ_i^j is the cut-off level for agent i's messages, and $\alpha^{-1}(\cdot)$ is the inverse virtual cost.

in fig. 2 indicate the distortions resulting from limited communication. For those environments the agent with higher cost will be asked to produce.

In the context of this example, consider now the three-tier hierarchy in which agent 1 is the residual claimant and the principal only observes delivery of the output $\bar{B}$. The communication constraint which limits each agent to a binary message set does not bind in this organizational setting. The principal pays agent 1 a fixed price, and agent 1 with cost θ_1 makes agent 2 a take-it-or-leave-it offer at a price of $\alpha^{-1}(\theta_1)$ [recall that $\alpha(\theta) = \theta + F(\theta)/f(\theta)$]. Hence, the solution characterized in fig. 1 remains valid in this limited communication example. We note the three-tier hierarchy indeed exhibits a flexibility advantage insofar as the contract that agent 1 proposes to agent 2 is responsive to agent 1's information θ_1 (in fact this contract fully reveals agent 1's information). However, this flexibility turns out to be detrimental to the principal because of the accompanying control loss. To see this, note that in the three-tier hierarchy the principal's fixed price equals the expected cost borne by the highest-cost type of agent 1 ($\bar{\theta}_1 = 1$),

$$\alpha^{-1}(1)\cdot F(\alpha^{-1}(1)) + 1\cdot(1 - F(\alpha^{-1}(1)).$$

Yet, the principal has the option of using the following contract in centralized two-tier contracting: Agent 2 sends one of two possible messages and agent 1 does not communicate information at all. Agent 2 is asked to produce if he reports that $\theta_2 \leq \alpha^{-1}(1)$, and is paid $\alpha^{-1}(1)$ in that case. If agent 2 reports $\theta_2 > \alpha^{-1}(1)$, agent 1 produces $a_1 = 1$ and is paid $\bar{\theta}_1 = 1$. The expected cost of this mechanism matches that of the fixed price calculated above, and we therefore conclude that the potential flexibility gain of the three-tier hierarchy (without monitoring of cost) has no value for the principal.[28]

We now turn to the study of cost centers with limited communication. As argued in connection with Theorem 1, the principal can overcome the control loss associated with delegated contracting by letting agent 1 bear a share of the overall cost. However, the cost-share parameter needs to be calibrated exactly to the agent's type. With limited communication this becomes impossible and one therefore expects to be left with a residual control loss. To illustrate that, we return again to our basic example.

Example (*continued*): As before, suppose each agent is constrained to a binary message set, i.e., $|M_i| = 2$. In the cost center arrangement, the principal may then offer agent 1 a menu of contracts with two 'entrées'. Similar to the construction in the full communication case, consider a menu of linear contracts

$$x_1^u(c) = b^u - \beta^u \cdot (c - c^u), \qquad 1 \leq u \leq 2. \tag{12}$$

The parameters b^u, β^u, and c^u are chosen such that agent 1 will find it in his interest to report 'low' cost, i.e., report m_1^1, if and only if $\theta_1 \leq \theta_1^1$, where θ_1^1 is the same cut-off point as in the optimal centralized arrangement (see fig. 2). Given the cost-share parameter β^u, agent 1 seeks to minimize

$$\beta^u \cdot [p \cdot a_1 + p \cdot a_2 + \alpha(\theta_2) \cdot a_2] + \theta_1 \cdot a_1 .$$

Following the same argument as in the earlier part of the example in section 3.1, agent 1 will make agent 2 a take-it-or-leave-it offer to produce $a_2 = 1$ for a payment of $\alpha^{-1}(\mu^u \cdot \theta_1)$, where $\mu^u = 1/\beta^u$. The question then becomes whether the cost-share parameters β^1 and β^2 can be chosen so as to improve upon the production assignments resulting from the optimal centralized contract. Fig. 3 presents the geometry of the arguments involved. Like fig. 2, fig. 3 shows the optimal production assignment resulting from the optimal centralized contract when each agent can send either of two possible messages. Again, the hatched areas indicate those environments for which the production assignments are

[28] The above example suggests a more general result which is established in theorem 4.2 of MMR (1991): for arbitrary message sets M_1 and M_2, the three-tier hierarchy without monitoring is dominated by the centralized contracting arrangement represented in (11). Hence the control loss always outweighs the flexibility gain if the principal does not monitor agent 1's interaction with agent 2.

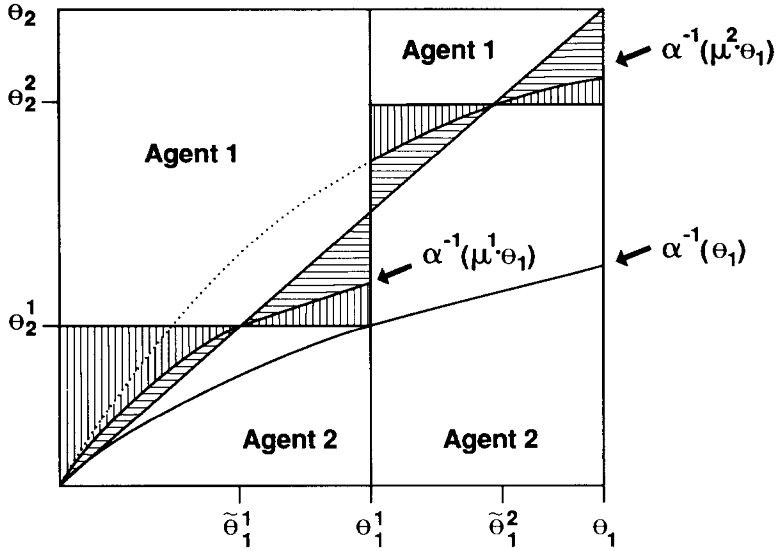

Fig. 3. Production assignments in a cost center with limited communication. θ_i is agent i's production cost, θ_i^j is the cut-off level for agent i's messages, $\alpha^{-1}(\cdot)$ is the inverse virtual cost, and $\tilde{\theta}_1^u$ is the average cost of agent 1 for the range $[\theta_1^{u-1}, \theta_1^u]$.

distorted, i.e., the agent with higher cost is the producer. To establish the superiority of a cost center over centralized contracting, it suffices to show that for an appropriate choice of the cost-share parameters β^1 and β^2 there will be fewer distortions. Specifically, we show that for appropriate β^1 and β^2, a cost center arrangement will avoid the distortions in the *vertically hatched* areas, without introducing any new distortions.[29]

To construct the cost-share parameters, let $\tilde{\theta}_1^1 \in [\underline{\theta}_1, \theta_1^1]$ and $\tilde{\theta}_1^2 \in [\theta_1^1, \bar{\theta}_1]$ be two types of agent 1 determined in the following way: $\tilde{\theta}_1^1$ is that type whose virtual cost [i.e., $\alpha(\theta_1^1)$] is equal to the average virtual cost of all types in the interval $[\underline{\theta}_1, \theta_1^1]$. Note that the principal must be indifferent as to who the producer is if $\theta_1 = \tilde{\theta}_1^1$ and θ_2 is equal to the cut-off point θ_2^1 (see fig. 3). Otherwise the principal could achieve lower expected cost in (11) by changing the value of θ_2^1. Type $\tilde{\theta}_1^2$ is determined in the same fashion on the interval $[\theta_1^1, \bar{\theta}_1]$. The cost-share parameters β^1 and β^2 are now chosen so that the compensation scheme in (12) induces the two types $\tilde{\theta}_1^u$, $1 \leq u \leq 2$, of agent 1 to implement

[29]The horizontally shaded areas represent environments for which the cost center remains inefficient relative to the optimal revelation mechanism (represented by the diagonal). With limited communication, however, the performance of revelation mechanisms is no longer the relevant comparison.

exactly the second-best production assignments in the cost center arrangement. As argued in connection with Theorem 1, in particular eq. (9), this requires that

$$1/\beta^u = \mu^u = \alpha(\tilde{\theta}_1^u)/\tilde{\theta}_1^u.$$

As a consequence, any type $\theta_1 \leq \theta_1^1$ will offer agent 2 to be the producer at a price of $\alpha^{-1}(\mu^1 \cdot \theta_1)$, while types $\theta_1 \geq \theta_1^1$ propose a price of $\alpha^{-1}(\mu^2 \cdot \theta_1)$. Fig. 3 shows that the induced production assignments are uniformly more efficient than those resulting from centralized contracting.[30] For environments in the vertically hatched areas, the agent with lower cost is now the producer, thereby eliminating the distortions of centralized contracting. The improved production efficiency results in lower virtual costs and therefore also in lower expected payments to the agents.

We now show that the analysis in the preceding example extends to a general result. Consider the general contracting problem associated with a cost center. If his information is θ_1, agent 1 is asked to report $\lambda_1(\theta_1) = m_1^u$ and thereby to select the contract $x_1(m_1^u, c)$ from the menu.[31] Subsequently, agent 1 designs a menu of contracts $\{x_2(m_2^v), a_2(m_2^v)\}$, $1 \leq v \leq k_2$, for agent 2. This contract choice will depend on agent 1's information θ_1 as well as on his report to the principal. Accordingly, we denote the contract by $x_2(m_2^v \mid \theta_1, m_1^u)$ and $a_2(m_2^v \mid \theta_1, m_1^u)$. Given assumptions A.1 and A.2, it can be shown that agent 1 will always choose an interval partition for agent 2. Thus, agent 2 is asked to report the message m_2^v if $\theta_2 \in (\theta_2^{v-1}, \theta_2^v]$ with $\theta_2^0 = \underline{\theta}_2$. The following optimization program represents the contract design problem for agent 1 when he has received the contract $x_1(c, m_1^u)$ from the principal, his type is θ_1 and he has reported m_1^u.

$$\text{P.3}_{\text{sub}} \quad \max_{\substack{\{\theta_2^v\},\{a_2^v\} \\ \{a_1^v\}}} \sum_{v=1}^{k_2} [x_1(c(m_2^v \mid \theta_1, m_1^u), m_1^u) - L_1(\theta_1) \cdot a_1(m_2^v \mid \theta_1, m_1^u)]$$

$$\times [F_2(\theta_2^v) - F_2(\theta_2^{v-1})],$$

subject to: for $1 \leq v \leq k_2$,

$$\text{(i)} \quad \theta_2^v \geq \theta_2^{v-1}, \qquad a_2^v \leq a_2^{v-1},$$

$$\text{(ii)} \quad B(a_1(m_2^v \mid \theta_1, m_1^u), a_2(m_2^v \mid \theta_1, m_1^u)) \geq \bar{B},$$

[30] Note that $\alpha^{-1}(\mu^u \cdot \theta_1) \leq \theta_1$ for $\theta_1 \leq \tilde{\theta}_1^u$, while $\alpha^{-1}(\mu^u \cdot \theta_1) \geq \theta_1$ for $\theta_1 \geq \tilde{\theta}_1^u$. This follows from the fact that $\alpha(\theta)/\theta$ is increasing in θ, which, in turn, is implied by assumption A.2.

[31] At times it will be notationally convenient to represent $x_1(c, m_1^u)$ more compactly as $x_1^u(c)$.

$$\text{(iii)}\quad c(m_2^v \mid \theta_1, m_1^u) \equiv \int_{\theta_2^{v-1}}^{\theta_2^v} [e_1(a_1(m_2^v \mid \theta_1, m_1^u)) + e_2(a_2(m_2^v \mid \theta_1, m_1^u) + h_2(\theta_2)\cdot a_2(m_2^v \mid \theta_1, m_1^u)]\,\mathrm{d}F_2(\theta_2).$$

For brevity, the incentive compatibility and participation constraints have been omitted in $P.3_{\text{sub}}$; instead agent 2's virtual cost has been substituted directly for his expected payment. We denote by $\Gamma(\theta_1 \mid x_1(m_1^u, c))$ the optimal value of $P.3_{\text{sub}}$. The principal's problem is then similar in spirit to P.2, except that he also has to choose a reporting rule $\lambda_1 : \Theta_1 \to M_1$. Unfortunately, there is no guarantee that the optimal reporting rule will induce an interval partition. We define $\Theta_1^u \equiv \lambda_1^{-1}(m_1^u) = \{\theta_1 \in \Theta_1 \mid \lambda_1(\theta_1) = m_1^u\}$. The principal's contract design problem can then be represented as

$$\text{P.3}\qquad \min_{\lambda_1(\cdot),\, x_1(\cdot)} \sum_{u=1}^{k_1} \int_{\Theta_1^u} \sum_{v=1}^{k_2} [c(m_2^v \mid \theta_1, m_1^u) + x_1(m_1^u, c(m_2^v \mid \theta_1, m_1^u))] \times [F_2(\theta_2^v) - F_2(\theta_2^{v-1})]\,\mathrm{d}F(\theta_1),$$

subject to: for all u, $1 \le u \le k_1$, and $\theta_1 \in \Theta_1^u$,

(i) $m_1^u \in \operatorname*{argmax}_{1 \le u \le k_1} \Gamma(\theta_1 \mid x_1(m_1^u, c))$,

(ii) $\Gamma(\theta_1 \mid x_1(m_1^u, c)) \ge 0$.

We are now in a position to compare the performance of the cost center and the two-tier hierarchy under limited communication.

Theorem 2. Suppose communication is limited to message sets of arbitrary sizes k_1 and k_2. Given assumptions A.1 and A.2, a cost center dominates centralized contracting.

Proof. See appendix.

The proof of Theorem 2 amounts to a generalization of the construction shown in the above example. Starting with the interval partition $\{\theta_1^u\}_{u=1}^{k_1}$ for agent 1 in the optimal centralized contract, the principal designs a menu $\{x_1(m_1^u, c)\}_{u=1}^{k_1}$ that induces the same partition, i.e., agent 1 will report

m_1^u whenever $\theta_1 \in (\theta_1^{u-1}, \theta_1^u]$. Each contract in the menu is linear in the observed cost and the cost-share parameters β^u satisfy

$$\frac{1}{\beta^u} = \mu^u = 1 + \frac{L_1'(\tilde{\theta}_1^u)}{L_1(\tilde{\theta}_1^u)} \cdot \frac{F_1(\tilde{\theta}_1^u)}{f_1(\tilde{\theta}_1^u)},$$

for some type $\tilde{\theta}_1^u$ in the interval $(\theta_1^{u-1}, \theta_1^u]$. Furthermore, $\tilde{\theta}_1^u$ can be chosen so that every type θ_1 of agent 1 will choose production assignments which provide a better approximation to the second-best production assignments than those resulting from centralized contracting. Monitoring of the cost c enables the principal to ameliorate the control loss sufficiently so that the flexibility gain associated with delegated decision-making dominates the residual control loss. Finally, our basic example suggests that 'in general' this dominance relation is strict.

When communication is unlimited, we showed that the principal does not lose anything by basing agent 1's compensation on the aggregate cost rather than the individual line items $e_1(a_1)$, $e_2(a_2)$, x_2 (Corollary to Theorem 1). A contract of the form $x_1(\theta_1, c)$ can align precisely agent 1's and the principal's preferences. With limited communication, though, a cost center arrangement involves a residual control loss, as represented by the horizontally hatched areas in fig. 3. This suggests that contracting on individual line items of the cost c may be valuable when communication is limited. We demonstrate this again in the context of our basic example, when the agents are restricted to k_1 and k_2 messages respectively.

Suppose first the principal offers agent 1 a menu of contracts based only on the aggregate cost c. Furthermore, suppose each contract is linear in c; that is,[32]

$$x_1^u(c) = b^u - \beta^u \cdot (c - c^u) \quad \text{for} \quad 1 \leq u \leq k_1.$$

An incentive compatible menu of contracts has the property that β^u is decreasing in u, and type θ_1 will select the contract with index u if $\theta_1 \in (\theta_1^{u-1}, \theta_1^u]$. As argued in connection with Theorem 2, type θ_1 of agent 1 will then make a take-it-or-leave-it offer to agent 2 at a price of $\alpha^{-1}(\mu^u \cdot \theta_1)$. If β^u is chosen optimally, there exists $\tilde{\theta}_1^u \in (\theta_1^{u-1}, \theta_1^u]$ such that $\alpha^{-1}(\mu^u \cdot \tilde{\theta}_1^u) = \tilde{\theta}_1^u$,

[32] We recall that linear contracts are optimal when communication is unlimited. Furthermore, as argued above in connection with Theorem 1, linear incentive schemes have desirable robustness properties when the agency problem is subject to 'noise'. Nonetheless, the following discussion only pertains to the value of line term reporting relative to incentive schemes that are linear in the aggregate cost c. Conceivably, the optimal incentive scheme is nonlinear when communication is limited, and it remains an open question whether line term reporting has value in this case.

i.e., there exists one type in the interval $[\theta_1^{u-1}, \theta_1^u]$ who will implement the second-best production assignments.[33]

If the principal monitors the individual components $p \cdot a_1$, $p \cdot a_2$, and x_2 of the cost c, agent 1 can be given an incentive scheme of the following form:

$$x_1^u(c, a_1) = \hat{b}^u - \gamma^u \cdot (c - \hat{c}^u) + \rho^u \cdot a_1 \quad \text{for} \quad 1 \leq u \leq k_1 .$$

Faced with the above incentive scheme, agent 1 will offer agent 2 a price of $\alpha^{-1}((\theta_1 - \rho^u) \cdot \varepsilon^u)$, where $\varepsilon^u \equiv 1/\gamma^u$. For a geometric illustration consider again fig. 3. Consider the interval $[\theta_1^1, \bar{\theta}_1]$, i.e., $u = 2$. Line item reporting will be valuable, if the parameters ρ^2 and ε^2 can be chosen in such a way that the curve $\alpha^{-1}((\theta_1 - \rho^2) \cdot \varepsilon^2)$ is everywhere between the diagonal and the curve $\alpha^{-1}(\mu^2 \cdot \theta_1)$ on the interval $[\theta_1^1, \bar{\theta}_1]$. In particular, this requires that the three curves meet at the point $\tilde{\theta}_1^2$, i.e., $\alpha^{-1}((\tilde{\theta}_1^2 - \rho^2) \cdot \varepsilon^2)) = \tilde{\theta}_1^2$. As a consequence, agent 1 will choose production assignments that are uniformly closer to the second-best assignments (represented by the diagonal) than those resulting from a cost center arrangement without line item reporting.

Algebraically, the parameters ρ^u and ε^u have to satisfy

$$\varepsilon^u > \mu^u \quad \text{and} \quad \rho^u = \tilde{\theta}_1^u \cdot \frac{\varepsilon^u - \mu^u}{\varepsilon^u} . \tag{13}$$

Recalling that the function $\alpha(\theta)/\theta$ is increasing in θ (by assumption A.2), it is readily verified that (13) implies

$$\alpha^{-1}(\mu^u \cdot \theta_1) \leq \alpha^{-1}((\theta_1 - \rho^u) \cdot \varepsilon^u) \leq \theta_1 \quad \text{for} \quad \theta_1 \in [\tilde{\theta}_1^u, \theta_1^u],$$

while the opposite inequalities hold for $\theta_1 \in [\theta_1^{u-1}, \tilde{\theta}_1^u]$.

Thus, when the principal monitors the individual line items, it is preferable to lower the cost-share parameter on the aggregate cost c from β^u to γ^u. As argued before, this has the effect of inducing agent 1 to assign more production to agent 2. At the same time, agent 1 is given a unit subsidy of ρ^u for the production of a_1. In the resulting compensation scheme, agent 1 is given a stronger incentive to contain the expenditures in connection with a_2, i.e., $p \cdot a_2 + x_2$, relative to the expenditure $p \cdot a_1$. In summary, line item reporting is valuable in this context because it allows the principal to fine-tune agent 1's incentives so as to further ameliorate the control loss.

[33]It can be verified that if $\alpha^{-1}(\mu^u \cdot \theta_1) < \theta_1$ for all $\theta_1 \in [\theta_1^{u-1}, \theta_1^u]$, then the principal could do better by lowering β^u, and accordingly adjust b^u and c^u.

4. Profit centers

We now consider the more general setting in which the benefit level (output or gross profit) for the principal is not determined exogenously. In centralized contracting with unlimited communication, the principal then seeks to maximize the difference between the gross benefit B and all costs subject to the usual incentive and participation constraints. By the same reasoning as in P.1 and (5) above, it can be shown that the grand contracting problem reduces to

$$\max_{a_1, a_2} \mathrm{E}_\theta \left[B(a_1, a_2) - \sum_{i=1}^{2} [e_i(a_i(\theta_1, \theta_2)) + h_i(\theta_i) \cdot a_i(\theta_1, \theta_2)] \right].$$

Again, this problem can be solved pointwise, i.e., for each environment (θ_1, θ_2), the second-best production assignments solve:

$$\max_{a_1, a_2} \left[B(a_1, a_2) - \sum_{i=1}^{2} [e_i(a_i) + h_i(a_i, \theta_i)] \right]. \tag{14}$$

In this setting, a natural extension of the cost center arrangement is for the principal to choose the output $\bar{B}(\theta_1)$ to be produced on the basis of agent 1's message θ_1. Agent 1's compensation scheme would again take the form $x_1(c, \theta_1)$, where c is the cost incurred to produce the output level $\bar{B}(\theta_1)$. As shown above, monitoring of the cost c allows the principal to create incentives for agent 1 to distribute the production assignments between himself and agent 2 in an unbiased way. However, the cost center arrangement would still result in a distortion of the overall output level $\bar{B}$ produced. This follows from the observation that 'generically' the solution $(a_1^*(\theta_1, \theta_2), a_2^*(\theta_1, \theta_2))$ to the optimization problem is such that $B(a_1^*(\theta_1, \theta_2), a_2^*(\theta_1, \theta_2))$ varies with θ_2. Using the Revenue Equivalence Theorem again, we conclude that a cost center cannot replicate the second-best performance if the output level $\bar{B}$ is solely based on agent 1's message θ_1.[34]

An alternative organizational arrangement is for the principal to delegate the output choice to agent 1. Agent 1 may then decide on $\bar{B}$ based not only on his information, but on the basis of information communicated by agent 2.

[34] We could have considered an alternative sequencing of the agents' messages. In particular, if we let agent 1 communicate with agent 2 before he reports to the principal, the output level could depend on both agents' messages. The problem with this scenario, however, is that it would amplify the adverse selection problem. The principal would then have to elicit two-dimensional information from agent 1, and as a result incur higher informational costs. See also the work of McAfee and McMillan (1992).

Conceivably, this arrangement might aggravate the control loss associated with delegated contracting, since agent 1 now has the opportunity to also distort the choice of the output level $\bar{B}$. To examine this issue, suppose that the principal evaluates agent 1 on the basis of the following measure of profit:

$$\pi \equiv B(a_1, a_2) - e_1(a_1) - e_2(a_2) - x_2 .$$

The principal then chooses an incentive contract of the form $\{x_1(\theta_1, \pi)\}$ for agent 1. In contrast to the cost center arrangement, the principal now no longer needs to observe the aggregate output $\bar{B}$. Given any report $\tilde{\theta}_1$ and his true type θ_1, agent 1 designs an incentive contract $\{a_2(\theta_2 | \theta_1, \tilde{\theta}_1), x_2(\theta_2 | \theta_1, \tilde{\theta}_1)\}$ for agent 2, solving the following optimization problem:

$$\text{P.4}_{\text{sub}} \quad \max_{\substack{a_1(\cdot), a_2(\cdot) \\ x_2(\cdot)}} \ \mathrm{E}_{\theta_2}[x_1(\tilde{\theta}_1, \pi(\theta_2 | \theta_1, \tilde{\theta}_1)) - L_1(\theta_1) \cdot a_1(\theta_2 | \theta_1, \tilde{\theta}_1)],$$

subject to: for all θ_2,

$$\text{(i)} \quad \theta_2 \in \operatorname*{argmax}_{\tilde{\theta}_2} \{x_2(\theta_2 | \theta_1, \tilde{\theta}_1) - L_2(\theta_2) \cdot a_2(\theta_2 | \theta_1, \tilde{\theta}_1)\},$$

$$\text{(ii)} \quad x_2(\theta_2 | \theta_1, \tilde{\theta}_1) - L_2(\theta_2) \cdot a_2(\theta_2 | \theta_1, \tilde{\theta}_1) \geq 0,$$

$$\text{(iii)} \quad \pi(\theta_2 | \theta_1, \tilde{\theta}_1) \equiv B(a_1(\theta_2 | \theta_1, \tilde{\theta}_1), a_2(\theta_2 | \theta_1, \tilde{\theta}_1)) - [e_1(a_1(\theta_2 | \theta_1, \tilde{\theta}_1)) + e_2(a_2(\theta_2 | \theta_1, \tilde{\theta}_1)) + x_2(\theta_2 | \theta_1, \tilde{\theta}_1)] .$$

The principal, in turn, chooses agent 1's incentive contract $x_1(\theta_1, \pi)$ to maximize his expected *net* profit, i.e., the profit π minus the payment x_1 to agent 1. Let $\pi(\theta_1, \theta_2) \equiv \pi(\theta_2 | \theta_1, \theta_1)$ and $\Gamma(\theta_1 | x_1(\tilde{\theta}_1, \pi))$ denote the value of the optimization program P.4_{sub}, induced by the scheme $x_1(\theta_1, \pi)$. The principal's problem in the profit center arrangement can be represented as:

$$\text{P.4} \quad \max_{x_1(\cdot, \cdot)} \ \mathrm{E}_{\theta}[\pi(\theta_1, \theta_2) - x_1(\theta_1, \pi(\theta_1, \theta_2))],$$

subject to: for all θ_1,

(i) $\theta_1 \in \underset{\tilde{\theta}_1}{\operatorname{argmax}}\, \Gamma(\theta_1 \mid x_1(\tilde{\theta}_1, \pi))$,

(ii) $\Gamma(\theta_1 \mid x_1(\theta_1, \pi)) \geq 0$.

In analogy to Theorem 1 above we obtain the following result:[35]

Theorem 1′. Given assumptions A.1 and A.2 and unlimited communication, a profit center can replicate the performance of the optimal revelation mechanism.

To establish this result, it suffices to consider incentive schemes that are linear in the profit measure π,

$$x_1(\theta_1, \pi) = b(\theta_1) + \beta(\theta_1)\cdot(\pi - \pi^*(\theta_1)). \tag{14}$$

The logic of the argument is then the same as in Theorem 1 above. The target profits $\pi^*(\theta_1)$ and the lump-sum payments $b(\theta_1)$ are chosen so as to ensure truthful reporting and satisfaction of the participation constraints. Thereafter, agent 1 will choose a_1 and a_2 [and thereby $B(a_1, a_2)$] so as to maximize

$$\beta(\theta_1)\cdot\pi - L_1(\theta_1)\cdot a_1.$$

It is readily checked that if the profit shares $\beta(\theta_1)$ are chosen according to the formula in (9), then agent 1 will exactly internalize the principal's objective function as given in (14). We note again that there is no need for the principal to monitor more than just the aggregate profit π.

The above arguments also extend to a world of limited communication. In direct analogy to (11), the centralized contracting problem with an endogenous output level and message spaces of size k_1 and k_2, respectively, becomes

$$\max_{\substack{\{\theta_1^u\},\{\theta_2^{uv}\}\\ \{a_i^{uv}\}}} \sum_{u=1}^{k_1}\sum_{v=1}^{k_2} \int_{\theta_1^{u-1}}^{\theta_1^u} \int_{\theta_2^{uv}}^{\theta_2^{uv}} [B(a_1^{uv}, a_2^{uv}) - h_1(\theta_1)\cdot a_1^{uv} - h_2(\theta_2)\cdot a_2^{uv}$$

$$- e_1(a_1^{uv}) - e_2(a_2^{uv})]\,\mathrm{d}F_2(\theta_2)\,\mathrm{d}F_1(\theta_1).$$

Agent 1's and the principal's optimization problems in a profit center are similar to those of the cost center in P.4$_{\text{sub}}$ and P.4. We then obtain the following result.

[35] We omit the proofs of the following two theorems since they are sufficiently similar to those of Theorems 1 and 2.

Theorem 2′. Suppose communication is limited to message sets of arbitrary sizes k_1 and k_2. Given assumptions A.1 and A.2, a profit center dominates centralized contracting.

A natural question at this point is whether the principal prefers a profit center to a cost center arrangement. From the above argument we know that with unlimited communication a profit center will strictly dominate a cost center, since the former can replicate the second-best solution, while the latter cannot. To obtain a preference for cost centers it will therefore be necessary to change the model in a way that makes it impossible to eliminate the control loss by setting up a profit center. One such approach would be to relax the multiplicative separability assumption. The persistence of a control loss may make it preferable for the principal to retain decision-making regarding the firm's output level.

5. Concluding remarks

The model examined in this paper shows that there need not be a loss of performance when an organization decentralizes its operations by delegating to intermediate agents the authority to make decisions and to contract with other agents. To prevent such delegation schemes from introducing new contracting frictions, though, it is essential that the principal monitor some financial performance measure, such as cost or profit. A suitably constructed compensation scheme based on those financial performance measures allows the principal to align the objectives of the intermediate agent with his own. When communication is limited, a responsibility center enjoys additional flexibility that the principal can exploit to his own advantage. This additional flexibility makes responsibility centers superior to centralized two-tier arrangements.

The analysis of our earlier paper, MMR (1991), shows that the principal can mitigate the control loss associated with a three-tier hierarchy by monitoring the agents' individual production contributions, as determined by the intermediate agent. While our analysis in section 4 identifies a particular situation where monitoring payments is more effective than monitoring actions, this issue needs to be researched in more generality. Monitoring payments has the general advantage that monetary units can be aggregated across different layers of the organization. This aggregation property may become essential in large organizations.

Given the elementary structure of our model, it is probably premature to attempt an empirical validation of our results. It will be interesting, however, to see whether subsequent studies (theoretical and empirical) will lend support to some of our findings. First, our analysis suggests that the relative advantage of setting up responsibility centers increases as communication becomes more limited. One would therefore expect that responsibility centers are prevalent in situations where the agents' private information is of technical nature, making it particularly costly to communicate that information.

A second direction for empirical study is the use of line item reporting for the purpose of performance evaluation. Our analysis suggests that when communication is unlimited and the principal can offer a 'complete' menu of incentive schemes, it will be sufficient to evaluate the performance of the responsibility center manager on the basis of aggregate cost or profit. In contrast, we find that it becomes desirable to attach different weights to individual line items when communication is limited. It would be useful to compare these findings with institutional practice.

With regard to the choice of cost versus profit center, our analysis offers only a simple prediction at this juncture. A profit center appears preferable if the output level to be produced is endogenous, i.e., it varies with the information held by the agents in the center. In contrast, a cost center appears to be appropriate if the center's overall output is given exogenously. Our results show that in terms of the underlying agency model there is no difference between setting the optimal cost-share parameters for managers of cost centers and the profit-sharing parameters for managers of profit centers. In particular, the range of possible sharing parameters is the same for any given agency problem. It remains to be seen whether further studies can provide additional support for this characterization.

We view our analysis here as a step towards a richer theory of decentralization and responsibility centers. Such a theory can enhance our understanding of the frequently complex structure of responsibility centers.[36] A next natural step would be to develop a model that addresses how organizational subunits should be grouped into responsibility centers. For instance, what are the trade-offs in keeping separate revenue and cost centers as opposed to a single profit center?

Moving further afield, one would like to capture the costs and benefits of a nested responsibility center structure. A common example in this context is that the profit measure for a profit center is frequently based on the sum of costs incurred by cost centers that belong to the profit center. For obvious reasons, such a theory will require a more complex model involving additional agents and activities that need to be coordinated. We expect that the dichotomy analyzed in this paper will continue to play a major role in future work: delegation of production and contracting decisions is preferable when communication is costly. As the organization becomes more decentralized, though, it becomes increasingly difficult for the principal to contain the attendant control loss.

[36] The prevalence of such organizational structures has led to the perspective of a firm as a 'nexus of contracts'; see Williamson (1975, 1986), Jensen and Meckling (1976), and Ball (1988).

Appendix: Proofs

Theorem 1

Consider the linear contract for agent 1,

$$x_1(\theta_1, c) = b(\theta_1) - \beta(\theta_1)\cdot(c - c^*(\theta_1)), \tag{i}$$

where $b(\theta_1)$ is given by (9). The function $b(\theta_1)$ is specified below so as to meet the individual rationality and the incentive compatibility constraints. We argued in the text that if agent 1 is truthful, the contract in (i) induces second-best production assignments [given by (6)]. We now construct the function $b(\theta_1)$ so that the expected payments are the same in both frameworks, and then verify global incentive compatibility.

The choice of $b(\theta_1)$ and $c^*(\theta_1)$ is based on the individual rationality and local incentive compatibility constraints. With a slight abuse of notation, we let $\Gamma_1(\tilde{\theta}_1, \theta_1)$ denote agent 1's expected payoff in $P.3_{sub}$ when his type is θ_1 and he sends message $\tilde{\theta}_1$,

$$\Gamma(\tilde{\theta}_1, \theta_1) = \mathrm{E}_{\theta_2}[x_1(\tilde{\theta}_1, c(\theta_2|\theta_1, \tilde{\theta}_1)) - L_1(\theta_1)\cdot a_1(\theta_2|\theta_1, \tilde{\theta}_1)],$$

where, as defined in the text, $c(\theta_2|\theta_1, \tilde{\theta}_1) \equiv e_1(a_1(\theta_2|\theta_1, \tilde{\theta}_1)) + e_2(a_2(\theta_2|\theta_1, \tilde{\theta}_1)) + x_2(\theta_2|\theta_1, \tilde{\theta}_1)$.

Denote by $a_i^*(\theta_1, \theta_2)$ the second-best production assignments [i.e., the solution to (6)], and set

$$c^*(\theta_1) = \mathrm{E}_{\theta_2}[e_1(a_1^*(\theta_1, \theta_2)) + e_2(a_2^*(\theta_1, \theta_2)) + h_2(\theta_2)\cdot a_2^*(\theta_1, \theta_2)],$$

$$\tag{ii}$$

$$b(\theta_1) = \mathrm{E}_{\theta_2}[L_1(\theta_1)\cdot a_1^*(\theta_1, \theta_2) + \int_{\theta_1}^{\bar{\theta}_1} L_1'(t)\cdot a_1^*(t, \theta_2)\,\mathrm{d}t].$$

It is apparent that the resulting expected budget variance is zero and that agent 1's expected payment $b(\theta_1)$ is equal to his expected payment in grand contracting. So the only remaining task is to verify global incentive compatibility.

We invoke a result by Mirrlees (1981) showing that a mechanism is globally incentive-compatible provided:

(1) $\Gamma(\theta_1, \theta_1) \geqslant 0,$

$$(2)\qquad \Gamma(\theta_1,\theta_1)=\Gamma(\bar{\theta}_1,\bar{\theta}_1)-\int_{\theta_1}^{\bar{\theta}_1}\frac{\partial}{\partial\theta_1}\Gamma(t,t)\,\mathrm{d}t,$$

$$(3)\qquad \frac{\partial}{\partial\theta_1}\Gamma(\tilde{\theta}_1,\theta_1)\text{ is increasing in }\tilde{\theta}_1.$$

Obviously, requirements (1) and (2) are satisfied for the contract in (ii). From the Envelope Theorem we know that

$$\frac{\partial}{\partial\theta_1}\Gamma(\tilde{\theta}_1,\theta_1)=-\int_{\Theta_2}L_1'(\theta_1)\cdot a_1(\theta_2\,|\,\theta_1,\tilde{\theta}_1)\,\mathrm{d}F_2(\theta_2).^{37}$$

To verify condition (3) it is sufficient to check that $a_1(\theta_2\,|\,\theta_1,\tilde{\theta}_1)$ is decreasing in $\tilde{\theta}_1$. This property is established by the following 'revealed preference' argument.

Fix θ_1 and θ_2 and let $a_1(\tilde{\theta}_1)\equiv a_1(\theta_1\,|\,\tilde{\theta}_1,\theta_2)$ and $a_2(\tilde{\theta}_1)\equiv a_2(\theta_1\,|\,\tilde{\theta}_1,\theta_2)$ denote the optimal assignments made by agent 1 when his message is $\tilde{\theta}_1$. Similarly, let $c(\tilde{\theta}_1)\equiv c(\theta_2\,|\,\theta_1,\tilde{\theta}_1)$ and $x_2(\tilde{\theta}_1)\equiv x_2(\theta_2\,|\,\theta_1,\tilde{\theta}_1)$. The assumed optimality of $a_1(\tilde{\theta}_1)$ implies that for any $\hat{\theta}_1$,

$$\begin{aligned}&b(\tilde{\theta}_1)-\beta(\tilde{\theta}_1)(c(\tilde{\theta}_1)-c^*(\tilde{\theta}_1))-L_1(\theta_1)\cdot a_1(\tilde{\theta}_1)\\&\geq b(\tilde{\theta}_1)-\beta(\tilde{\theta}_1)(c(\hat{\theta}_1)-c^*(\tilde{\theta}_1))-L_1(\theta_1)\cdot a_1(\hat{\theta}_1).\end{aligned}$$

Alternatively,

$$\begin{aligned}&-\beta(\tilde{\theta}_1)[e_1(a_1(\tilde{\theta}_1))+e_2(a_2(\tilde{\theta}_1))+h_2(\theta_2)\cdot a_2(\tilde{\theta}_1)]-a_1(\tilde{\theta}_1)\cdot L_1(\theta_1)\\&\geq-\beta(\tilde{\theta}_1)[e_1(a_1(\hat{\theta}_1))+e_2(a_2(\hat{\theta}_1))+h_2(\theta_2)\cdot a_2(\hat{\theta}_1)]-a_1(\hat{\theta}_1)\cdot L_1(\theta_1).\end{aligned}$$

Rearranging and dividing by $\beta(\tilde{\theta}_1)$ yields

$$\begin{aligned}&e_1(a_1(\hat{\theta}_1))+e_2(a_2(\hat{\theta}_1))+h_2(\theta_2)\cdot a_2(\hat{\theta}_1)+\frac{L_1(\theta_1)}{\beta(\tilde{\theta}_1)}\cdot a_1(\hat{\theta}_1)\\&\geq e_1(a_1(\tilde{\theta}_1))+e_2(a_2(\tilde{\theta}_1))+h_2(\theta_2)\cdot a_2(\tilde{\theta}_1)+\frac{L_1(\theta_1)}{\beta(\tilde{\theta}_1)}\cdot a_1(\tilde{\theta}_1).\qquad\text{(iii)}\end{aligned}$$

[37] This follows from the fact that $c(\theta_2\,|\,\theta_1,\tilde{\theta}_1)$ depends on θ_1 only via $a_1(\theta_2\,|\,\theta_1,\tilde{\theta}_1)$ and $a_2(\theta_2|\theta_1,\tilde{\theta}_1)$, and that, by definition, the functions $a_1(\cdot\,|\,\theta_1,\tilde{\theta}_1)$ and $a_2(\cdot\,|\,\theta_1,\tilde{\theta}_1)$ are the pointwise maximizers of $\Gamma(\theta_1,\tilde{\theta}_1)$.

Similarly, the optimality of $a_1(\hat{\theta}_1)$ implies that

$$e_1(a_1(\tilde{\theta}_1)) + e_2(a_2(\tilde{\theta}_1)) + h_2(\theta_2)\cdot a_2(\tilde{\theta}_1) + \frac{L_1(\theta_1)}{\beta(\hat{\theta}_1)}\cdot a_1(\tilde{\theta}_1)$$

$$\geq e_1(a_1(\hat{\theta}_1)) + e_2(a_2(\hat{\theta}_1)) + h_2(\theta_2)\cdot a_2(\hat{\theta}_1) + \frac{L_1(\theta_1)}{\beta(\hat{\theta}_1)}\cdot a_1(\hat{\theta}_1). \qquad \text{(iv)}$$

Adding up inequalities (iii) and (iv), we obtain

$$\beta(\hat{\theta}_1)(a_1(\tilde{\theta}_1)) - (a_1(\hat{\theta}_1)) \leq \beta(\tilde{\theta}_1)(a_1(\tilde{\theta}_1)) - (a_1(\hat{\theta}_1)).$$

Since $\beta(\theta_1)$ is decreasing in θ_1, it follows that $a_1(\theta_1)$ is decreasing in θ_1.

Theorem 2

Suppose the principal offers a menu of contracts

$$x_1^u(c) = b_1^u - \beta_1^u\cdot(c - c^u), \qquad \text{(i)}$$

where the partition $\{\theta_1^u\}_{u=1}^{k_1}$ for agent 1 is the same as in the optimal centralized contract. The coefficient β_1^u is chosen such that

$$\beta_1^u = \frac{1}{\mu_1^u},$$

where

$$\mu_1^u = 1 + \frac{L_1'(\tilde{\theta}_1^u)}{L_1(\tilde{\theta}_1^u)}\cdot\frac{F_1(\tilde{\theta}_1^u)}{f_1(\tilde{\theta}_1^u)} \quad \text{for some} \quad \tilde{\theta}_1^u \in [\theta_1^{u-1}, \theta_1^u].$$

We now show that the production assignments under $x_1^u(c)$ result in lower costs for the principal than those chosen in the centralized arrangement.

Specifically, we show that the contract in (i) induces production assignments $a_i(\theta) \equiv a_i(\theta_1, \theta_2)$ such that for all $\theta_1 \in [\underline{\theta}_1, \bar{\theta}_1]$,

$$\int_{\underline{\theta}_2}^{\bar{\theta}_2} [h_1(\theta_1)\cdot a_1(\theta) + h_2(\theta_2)\cdot a_2(\theta) + \sum_{i=1}^{2} e_i(a_i(\theta))]\,\mathrm{d}F_2(\theta_2)$$

$$\leq \int_{\underline{\theta}_2}^{\bar{\theta}_2} [h_1(\theta_1)\cdot a_1^*(\theta) + h_2(\theta_2)\cdot a_2^*(\theta) + \sum_{i=1}^{2} e_i(a_i^*(\theta))]\,\mathrm{d}F_2(\theta_2), \qquad \text{(ii)}$$

where

$$h_1(\theta_1) \equiv L_1(\theta_1) + L_1'(\theta_1)\cdot\frac{F_1(\theta_1)}{f_1(\theta_1)},$$

and the functions $a_i^*(\theta)$ denote an optimal solution to the centralized program as stated in (11). Thus, $a_i^*(\theta) = a^{uv}$ if $\theta_1^{u-1} \leq \theta_1 \leq \theta_1^u$ and $\theta_2^{u,v-1} \leq \theta_2 \leq \theta_2^{uv}$. In a cost center arrangement, type θ_1 of agent 1 chooses a partition $\{\theta_2^v\}_{v=1}^{k_2}$ and production assignments $\{(a_1^v, a_2^v)\}_{v=1}^{k_2}$. For $\theta_1 \in [\theta_1^{u-1}, \theta_1^u]$, inequality (ii) can therefore be rewritten as

$$\sum_{v=1}^{k_2} [h_1(\theta_1)\cdot a_1^v + h_2^v\cdot a_2^v + \sum_{i=1}^{2} e_i(a_i^v)]\,\Delta F_2^v$$

$$\leq \sum_{v=1}^{k_2} [h_1(\theta_1)\cdot a_1^{uv} + h_2^{uv}\cdot a_2^{uv} + \sum_{i=1}^{2} e_i(a_i^{uv})]\,\Delta F_2^{uv}, \tag{iii}$$

where

$$\Delta F_2^v \equiv F_2(\theta_2^v) - F_2(\theta_2^{v-1}), \qquad \Delta F_2^{uv} \equiv F_2(\theta_2^{uv}) - F_2(\theta_2^{u,v-1}),$$

$$h_2^v \equiv \int_{\theta_2^{v-1}}^{\theta_2^v} h_2(\theta_2)\,\mathrm{d}F_2(\theta_2)/\Delta F_2^v, \qquad h_2^{uv} \equiv \int_{\theta_2^{u,v-1}}^{\theta_2^{uv}} h_2(\theta_2)\,\mathrm{d}F_2(\theta_2)/\Delta F_2^{uv}. \tag{iv}$$

Step 1: For any constant γ and given θ_1 define

$$W(\{\theta_2^v\}, \{a^v\}, \gamma) \equiv \sum_{v=1}^{k_2} [(h_1(\theta_1) + \gamma)\cdot a_1^v + h_2^v\cdot a_2^v + \sum_{i=1}^{2} e_i(a_i^v)]\,\Delta F_2^v.$$

We denote the minimizer of $W(\cdot,\cdot,\gamma)$ by $\{\theta_2^v(\gamma)\}_{v=1}^{k_2}, \{a^v(\gamma)\}_{v=1}^{k_2}$.[38] Furthermore, we define

$$V(\gamma) = \sum_{v=1}^{k_2} [h_1(\theta_1)\cdot a_1^v(\gamma) + h_2^v(\gamma)\cdot a_2^v(\gamma) + \sum_{i=1}^{2} e_i(a_i^v(\gamma))]\,\Delta F_2^v(\gamma),$$

[38]If there is more than one minimizer, any one of them can be selected.

where

$$\Delta F_2^v(\gamma) \equiv F_2(\theta_2^v(\gamma)) - F_2(\theta_2^{v-1}(\gamma)),$$

$$h_2^v(\gamma) \equiv \int_{\theta_2^{v-1}(\gamma)}^{\theta_2^v(\gamma)} h_2(\theta_2)\,\mathrm{d}F_2(\theta_2)/\Delta F_2^v(\gamma).$$

Therefore, $V(\gamma)$ is the principal's expected cost for a given θ_1 when the partition of Θ_2 and the production assignments are chosen so as to minimize $W(\cdot,\cdot,\gamma)$.

Claim. The function $V(\cdot)$ is increasing in γ for $\gamma > 0$ and decreasing for $\gamma < 0$.

Proof. To see this, we first verify that the function

$$R(\gamma) \equiv \sum_{v=1}^{k_2} a_1^v(\gamma)\Delta F_2^v(\gamma)$$

is decreasing in γ. This follows from the following 'revealed preference' argument. Given γ and γ^*, we obtain $W(\{\theta_2^v(\gamma)\},\{a^v(\gamma)\},\gamma) \geq W(\{\theta_2^v(\gamma)\},\{a^v(\gamma)\},\gamma^*)$ and $W(\{\theta_2^v(\gamma^*)\},\{a^v(\gamma^*)\},\gamma^*) \geq W(\{\theta_2^v(\gamma^*)\},\{a^v(\gamma^*)\},\gamma)$. Adding these two inequalities, we obtain

$$(\gamma^* - \gamma)\cdot \sum_{v=1}^{k_2} a_1^v(\gamma)\Delta F_2^v(\gamma) \geq (\gamma^* - \gamma)\cdot \sum_{v=1}^{k_2} a_1^v(\gamma^*)\Delta F_2^v(\gamma^*),$$

proving that $R(\gamma)$ is decreasing in γ. If we let

$$W^*(\gamma) \equiv W(\{\theta_2^v(\gamma)\},\{a^v(\gamma)\},\gamma),$$

then the function $V(\cdot)$ can be rewritten as

$$V(\gamma) = W^*(\gamma) - \gamma\cdot R(\gamma).$$

By the Envelope Theorem we have

$$\frac{\partial}{\partial\gamma} W^*(\gamma) = R(\gamma)$$

or

$$V(\gamma) = \int_0^{\gamma} [R(t) - R(\gamma)]\,\mathrm{d}t + W(0).$$

Since $R(\cdot)$ is decreasing in γ, it follows directly that $V(\cdot)$ is increasing (decreasing) in γ if $\gamma > 0$ ($\gamma < 0$).

Step 2: We next demonstrate that the production assignments $a(\theta)$, induced by the contract in (i) in the cost center arrangement, minimize $W(\cdot, \cdot, \gamma(\theta_1))$ for some $\gamma(\theta_1)$. At the same time, the solution to the two-tier hierarchy $a^*(\theta)$ minimizes $W(\cdot, \cdot, \delta(\theta_1))$ for some $\delta(\theta_1)$ such that $\delta(\theta_1) \geq \gamma(\theta_1)$ if $\gamma(\theta_1) \geq 0$ and $\delta(\theta_1) \leq \gamma(\theta_1)$ if $\gamma(\theta_1) \leq 0$. In light of Step 1, we then conclude that $V(\gamma(\theta_1)) \leq V(\delta(\theta_1))$ for all $\theta_1 \in [\underline{\theta}_1, \bar{\theta}_1]$ and, therefore, the inequality in (ii) holds for all θ_1. Assuming incentive compatibility of the contract in (i), type $\theta_1 \in [\theta_1^{u-1}, \theta_1^u]$ chooses $\{\theta_2^v\}$, $\{a^v\}$ so as to minimize

$$\sum_{v=1}^{k_2} [L_1(\theta_1) \cdot \mu_1^u \cdot a_1^v + h_2^v \cdot a_2^v + \sum_{i=1}^{2} e_i(a_i^v)]\,\Delta F_2^v.$$

Hence, the production assignments $a(\theta)$ in the cost center arrangement maximize $W(\cdot, \cdot, \gamma(\theta_1))$, where $\gamma(\theta_1)$ satisfies

$$\gamma(\theta_1) + h_1(\theta_1) = L_1(\theta_1) \cdot \mu_1^u. \tag{v}$$

In the centralized arrangement, the principal chooses $\{\theta_2^{uv}\}_{v=1}^{k_2}$, $\{a^{uv}\}_{v=1}^{k_2}$ in response to message m_1^u so as to minimize

$$\Delta F_1^u \cdot \sum_{v=1}^{k_2} [h_1^u \cdot a_1^{uv} + h_2^{uv} \cdot a_2^{uv} + \sum_{i=1}^{2} e_i(a_i^{uv})]\,\Delta F_2^{uv}.$$

Since $h_1^u = \int_{\theta_1^{u-1}}^{\theta_1^u} h_1(\theta_1)\,\mathrm{d}F_1(\theta_1)/\Delta F_1^u$ and $h_1(\cdot)$ is continuous and increasing in θ_1, there exists a $\tilde{\theta}_1^u \in (\theta_1^{u-1}, \theta_1^u)$ such that $h_1(\tilde{\theta}_1^u) = h_1^u$. Hence, $\{\theta_2^{uv}\}_{v=1}^{k_2}$ and $\{a^{uv}\}_{v=1}^{k_2}$ minimize $W(\cdot, \cdot, \delta(\theta_1))$, where $\delta(\theta_1)$ is given by

$$\delta(\theta_1) + h_1(\theta_1) = h_1(\tilde{\theta}_1^u). \tag{vi}$$

Comparing (v) and (vi), it remains to show that for a suitable choice of μ_1^u,

$$L_1(\theta_1) \cdot \mu_1^u - h_1(\theta_1) \leq h_1(\tilde{\theta}_1^u) - h_1(\theta_1), \tag{vii}$$

provided $L_1(\theta_1) \cdot \mu_1^u - h_1(\theta_1) \geq 0$, while the opposite inequality must hold if $L_1(\theta_1) \cdot \mu_1^u - h_1(\theta_1) \leq 0$. These inequalities will hold if we set $\mu_1^u = h_1(\tilde{\theta}_1^u)/L_1(\tilde{\theta}_1^u)$.

Step 3: It remains to show that the compensation scheme (i) is incentive-compatible. Following the same approach as in the proof of Theorem 1, it suffices to demonstrate that the function

$$-\frac{\partial}{\partial\theta_1}\sum_{v=1}^{k_2}\{\min_{(a_1,a_2)}[(\gamma_1^u + L_1(\theta_1))\cdot a_1 + h_2^{uv}\cdot a_2 + e_1(a_1) + e_2(a_2)]\}$$

is decreasing in u for all (θ_1, θ_2^v). Hence, the function

$$\sum_{v=1}^{k_2} L_1'(\theta_1)\cdot a_1(\theta_1, m_1^u, \theta_2^v)\Delta F_2^v \qquad \text{(viii)}$$

has to be decreasing in u, where $a_1(\theta_1, m_1^u, \theta_2^v)$ denotes the action choice for type θ_1 if he sent message m_1^u and agent 2 sent message m_2^v. Recalling (vi), (vii), and that $\gamma_1^u = h_1(\tilde{\theta}_1^u) - L_1(\tilde{\theta}_1^u)$, (viii) is exactly equal to

$$L_1'(\theta_1)\cdot R(L_1(\theta_1) - h_1(\theta_1) + h_1(\tilde{\theta}_1^u) - L_1(\tilde{\theta}_1^u)).$$

Since $h_1(\tilde{\theta}_1^u) - L_1(\tilde{\theta}_1^u)$ is increasing in u and $R(\cdot)$ is decreasing in its argument (by the above claim), the proof is complete.

References

Alchian, A. and H. Demsetz, 1972, Production, information costs and economic organization, American Economic Review 62, 777–795.

Amershi, A. and P. Cheng, 1990, Transfer pricing: The economics of dominant strategy equilibrium mechanisms under information asymmetry and moral hazard, Contemporary Accounting Research 7, 61–99.

Antle, R. and G. Eppen, 1985, Capital rationing and organizational slack in capital budgeting, Management Science 31, 163–174.

Baiman, S., 1990, Agency research in managerial accounting: A second look, Accounting, Organizations and Society 15, 341–371.

Baiman, S. and J. Noel, 1985, Noncontrollable costs and responsibility accounting, Journal of Accounting Research 23, 468–501.

Ball, R., 1988, The firm as a specialist contracting intermediary: Applications for accounting and auditing, Working paper (University of Rochester, Rochester, NY).

Calvo, G. and S. Wellisz, 1978, Supervision, loss of control, and the optimum size of the firm, Journal of Political Economy 86, 943–952.

Christensen, J. and J. Demski, 1990, Transfer pricing in a limited communication setting, Working paper (Yale University, New Haven, CT).

Cordiner, R., 1956, New frontiers for professional managers (McGraw-Hill, New York, NY).

Cremer, J., 1979, A partial theory of the optimal organization of a bureaucracy, Bell Journal of Economics 15, 396–403.

Demski, J., J. Patell, and M. Wolfson, 1984, Decentralized choice of monitoring systems, The Accounting Reveiw 59, 16–34.

Demski, J. and D. Sappington, 1987, Delegated expertise, Journal of Accounting Research 25, 68–89.

Demski, J. and D. Sappington, 1989, Hierarchical structure and responsibility accounting, Journal of Accounting Research 27, 40–58.

Geanakaplos, J. and P. Milgrom, 1985, A theory of hierarchies based on limited managerial attention, Working paper (Yale University, New Haven, CT).

Guesnerie, R. and J. J. Laffont, 1984, 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 25, 329–369.
Harrison, M., C. Holloway, and J. Patell, 1989, Measuring delivery performance: A case study from the semiconductor industry, Working paper (Stanford University, Stanford, CA).
Higgins, J., 1952, Responsibility accounting, in: The Arthur Andersen chronicle, Vol. 12 (Arthur Andersen, Chicago, IL).
Horngren, C. and G. Foster, 1991, Cost accounting: A managerial emphasis, 7th ed. (Prentice Hall, Englewood Cliffs, NJ).
Jensen, M. and W. Meckling, 1976, Theory of the firm: Managerial behavior, agency costs and ownership structure, Journal of Financial Economics 3, 305–360.
Jordan, J., 1989, Accounting-based divisional performance measurement: Incentives for profit maximization, Contemporary Accounting Research 6, 903–921.
Kaplan, R. and A. Atkinson, 1989, Advanced management accounting, 2nd ed. (Prentice Hall, Englewood Cliffs, NJ).
Kirby, A., S. Reichelstein, P.K. Sen, and T. Paik, 1991, Participation, slack, and budget-based performance evaluation, Journal of Accounting Research 29, 109–128.
Laffont, J.J. and J. Tirole, 1986, Using cost observation to regulate firms, Journal of Political Economy 94, 614–641.
Marschak, T. and S. Reichelstein, 1992, Network mechanisms, informational efficiency and hierarchies, Working paper (University of California, Berkeley, CA).
McAfee, P. and J. McMillan, 1992, Organizational diseconomies of scale, Working paper (University of California, San Diego, CA).
Melumad, N., 1989, Asymmetric information and the termination of contracts in agencies, Contemporary Accounting Research 5, 733–753.
Melumad, N. and S. Reichelstein, 1987, Centralization versus delegation and the value of communication, Journal of Accounting Research 23, 1–18.
Melumad, N., D. Mookherjee, and S. Reichelstein, 1991, Hierarchical decentralization of incentive contracts, Working paper (Stanford University, Stanford, CA).
Mirrlees, J., 1981, Optimal taxation, in: K. Arrow and M. Intriligator, eds., Handbook of mathematical economics, Vol. III (North-Holland, Amsterdam).
Mookherjee, D. and S. Reichelstein, 1992, Dominant strategy implementation of Bayesian incentive compatible allocation rules, Journal of Economic Theory 56, 378–399.
Myerson, R., 1981, Optimal auction design, Mathematics of Operations Research 6, 58–73.
Myerson, R., 1982, Optimal coordination mechanisms in generalized principal–agent relationships, Journal of Mathematical Economics 10, 67–81.
Penno, M., 1988, Controls and decentralization in a principal–agent setting, Working paper (University of Chicago, Chicago, IL).
Prince, T., 1975, Information systems for management planning and control, 3rd ed. (Irwin, New York, NY).
Rajan, M., 1988, Cost allocation in multi-agent games, Working paper (Carnegie Mellon University, Pittsburgh, PA).
Radner, R., 1990, The organization of decentralized information processing, Working paper (Bell Laboratories).
Ramakrishnan, R., 1990, Decentralization and responsibility center design with asymmetric information, Working paper (Columbia University, New York, NY).
Ronen, J. and K. Balachandran, 1988, An approach to transfer pricing under uncertainty, Journal of Accounting Research 26, 300–314.
Vaysman, I., 1992, A model of cost-based transfer pricing, Working paper (University of California, Berkeley, CA).
Williamson, O., 1975, Markets and hierarchies (Free Press, New York, NY).
Williamson, O., 1986, The economic institutions of capitalism (Free Press, New York, NY).