Advances in Theoretical Economics

Volume 1, *Issue* 1 2001 *Article* 4

Incentives and Coordination in Hierarchies

Dilip Mookherjee* Stefan Reichelstein†

*Boston University, dilipm@bu.edu

†UC Berkeley, reich@haas.berkeley.edu

Incentives and Coordination in Hierarchies

Dilip Mookherjee and Stefan Reichelstein

Abstract

The internal organization of large firms as well as procurement and regulation contexts frequently involve a hierarchical nexus of contracts, with substantial delegation of decision making across layers. Such hierarchical delegation of decision making creates problems of aligning incentives of vertically related agents, and coordinating the actions of different branches of the hierarchy. In a principal-agent setting with private information, it is shown that under certain assumptions (top-down contracting, observability of subcontracting outcomes, absence of limited liability constraints) the hierarchy can implement second-best allocations. Incentive problems are overcome via compensations that are linear in a measure of performance of the concerned department, defined as the difference between a measure of imputed revenues and procurement costs. The coordination problem is overcome by conditioning output targets and payments on cost reports submitted by other branches; despite this, agents' strategies are dominant with respect to the behavior of members of other branches. The result provides conditions for the lack of a 'control loss' from hierarchical decentralization of decision making, owing to incentive or coordination problems.

KEYWORDS: hierarchies, networks, organization theory, profit centers, incentives, coordination, control loss

1 Introduction

The internal organization of a modern firm is typically characterized by separation of ownership from control, with decision making authority extensively delegated to agents located at successive layers of a managerial hierarchy. These managers are responsible for organizing production and delivery of intermediate products, which includes decisions concerning outsourcing versus internal procurement, negotiation with suppliers, employment and supervision of subordinates. In procurement settings, purchasers or regulators frequently deal with 'prime' suppliers, each of whom is granted considerable autonomy over sourcing of components from subcontractors. These subcontractors in turn choose how to source their components from other suppliers, and so on. Marketing channels represent another example of a hierarchical network: producers sell their products through a network of distributors, who contract with retail agents, and the latter eventually deal with final customers.

A complete explanation for the prevalence of contractual hierarchies would need to consider both costs and benefits of hierarchies in comparison to other arrangements, such as centralized revelation mechanisms or nonhierarchical communication networks. Williamson (1967), Arrow (1974), Keren and Levhari (1983), Radner (1992, 1993), van Zandt (1996a,b, 1997) and Marschak and Reichelstein (1995, 1999) have argued that economies in communication and information processing costs may explain the primary benefits of hierarchies.[1] Our analysis abstracts from communication or information processing considerations, while presuming that they constitute the main benefit of a hierarchy. We ask instead whether hierarchical delegation of contracting and decision-making entails any incentive costs.[2]

In order to focus on incentive questions *per se*, we consider hierarchies which are *consistent* with the technology, permitting achievement of first-best allocations in a world without any incentive constraints.[3] Then we introduce incentive

[1]Theories of hierarchies that emphasize costs of information processing include Radner (1993), Radner and van Zandt (1992) and van Zandt (1997); much of this is surveyed in van Zandt (1996a, 1996b). Keren and Levhari (1983) presented an earlier model of 'bottlenecks' resulting from time delays in planning. Communication costs have been the focus of team theory, as studied recently by Marschak and Reichelstein (1999). Models examining the role of communication costs in the presence of incentive problems include Green and Laffont (1986, 1987), Melamud, Mookerjee, and Reichelstein (1992) and Laffont and Martimort (1998). Contract complexity is treated in Melumad, Mookherjee and Reichelstein (1997).

[2]The Revelation Principle states that optimal allocations can always be achieved by centralized revelation mechanisms, so there cannot be any benefits from an incentive standpoint *per se* from the use of a hierarchical mechanism under the conditions underlying the Revelation Principle, which are assumed to hold in this paper: absence of communication costs, information processing costs, incompleteness (or renegotiation) of contracts. Under these conditions, a hierarchy can at best entail no costs relative to an optimal centralized revelation mechanism.

[3]We focus on incentive problems in the context of *complete* contracts, and abstract from problems of contract incompleteness or renegotiation. The role of delegation in the presence of incomplete contracts and commitment problems is explored by Melumad and Mookherjee

considerations by supposing that agents are privately informed about their own productivity, and behave in a self-interested manner. In such a world, first-best allocations are no longer implementable by *any* mechanism. The Revelation Principle ensures that a centralized revelation mechanism can implement incentive constrained (second-best) allocations. The question posed is whether the hierarchy can equivalently implement the second-best.

Earlier work on this question has focused on a simple three layer vertical hierarchy with two agents beside the Principal (Baron and Besanko (1992), McAfee and McMillan (1995), Melumad, Mookherjee and Reichelstein (1992, 1995) and Laffont and Martimort (1998)).[4] McAfee and McMillan provided an example where the hierarchy could not achieve second-best outcomes, owing to a problem of 'double marginalization of rents', arising from the monopoly power of the intermediate agent over the bottom layer agent. In contrast, the Melumad-Mookherjee-Reichelstein papers showed that this problem can be overcome if (a) there are no limited liability constraints for intermediate agents; (b) the contracting sequence is top-down (i.e., the higher level agents contracts with the Principal before contracting or communicating with the bottom layer agent), and (c) either payments or production assignments to the bottom layer agent are verifiable by the Principal. Under these conditions the double marginalization problem can be overcome by subsidizing 'outsourcing' by the intermediate agent, and ensuring that the latter contracts with the Principal before contracting with the agent (allowing the monopoly rents from the subcontract to be taxed away *ex ante*). Moreover, assumptions (a)-(c) are necessary for a hierarchy to achieve second-best outcomes, under weak conditions on technology and cost distributions.

In this paper, we examine the same question in the context of arbitrary hierarchies that are consistent with the technology. Such hierarchies may be characterized by any number of vertical layers and horizontal branches. Our main result is that under the same assumptions required in the simple two agent context, such hierarchies can implement second-best outcomes. Since these assumptions are typically necessary in the two agent context, we essentially identify conditions that are both necessary and sufficient for an arbitrary hierarchy (consistent with the technology) to entail no incentive costs.

While treating the case of more than three vertical layers is relatively straightforward, the more challenging question concerns coordination of production across different horizontal branches. Intermediate agents at each branch of the hierarchy must design contracts for their subordinates that condition on messages to be submitted later by other branches. Agents in different branches therefore simultaneously design mutually interlinked contracts. The mechanism has three phases (in contrast to just one phase in a pure vertical hierarchy):

(1989), Poitevin (1995) and Aghion and Tirole (1997).

[4]Gilbert and Riordan (1995) address a different question, concerning the desirability of consolidating two agents supplying complementary products into a single one, in a regulatory framework.

first, contracts flow down the hierarchy; second, cost reports flow up; and third, output targets flow down. In general, therefore, agents need to predict the outcomes in all other branches of the hierarchy in order to make decisions, significantly complicating their relevant incentives.

The mechanism we construct combines elements of the earlier hierarchical mechanisms that overcome vertical control losses, with Groves-Vickrey-type mechanisms to ensure horizontal coordination. Each intermediate agent i can be thought of as manager of a 'profit center', whose revenues are given by the maximum-willingness-to-pay of the agent's immediate boss for the product delivered by i, and costs measured by the total payments to subordinates of i. Manager i's compensation is linear in the profit, with a bonus coefficient lying between 0 and 1. At the time of contracting with her boss, i reveals her own private information, thereby determining her profit bonus coefficient. This coefficient is calibrated to ensure that the double marginalization of rents is overcome: the effective subsidy for 'outsourcing' from subordinates overcomes the incentive to pay subordinates too little, while the 'fixed' component (i.e., which does not depend on profit) provides agents their monopsony rents.

The revenue measure of the profit center ensures horizontal coordination, by effectively internalizing the contribution of i's department to the rest of the organization. Analogous to Groves-Vickrey mechanisms, managers have incentives to report their department's cost truthfully. In particular, the reporting strategies are dominant with respect to the reports to be submitted by other departments, as are also their contract participation and selection decisions. The mechanism thus greatly simplifies the decision problem for agents: despite the complex interlinking of their respective contracts, agents in different branches do not need to predict each other's behavior. In particular, the prescribed pattern of behavior constitutes a *sequentially dominant strategy* for each agent. Bottom layer agents have a dominant strategy; the strategy of penultimate layer agents constitutes a best response to their subordinates' strategies, but are dominant with respect to strategies of all other agents, and so on. This eliminates the need for agents to know and understand how the rest of the hierarchy functions. At the same time, we show that the mechanism cannot be manipulated by more sophisticated agents: the prescribed patterns of behavior constitute a Bayesian 'solution', which corresponds to the notion of a Perfect Bayesian Equilibrium.

From an information processing perspective the mechanism we construct has certain advantages relative to a revelation mechanism. In the context of a multi-plant firm, for instance, where each plant has a quadratic cost function and uniformly distributed cost parameter, the Principal and higher level managers need only information concerning aggregate productivity shocks of the departments managed by their immediate subordinates; the latter are delegated the responsibility of allocating production among their subordinates in turn. The tasks of information processing are distributed among the hierarchy in exactly the fashion described in the hierarchical resource allocation model of van Zandt (1997).[5]

The dominant strategy features of our mechanism implies that the planning, reporting and production assignment rules are prior-independent. For changes in the organization's environment, represented by changes in the distribution over cost shocks in different departments, the only aspect of the mechanism that need modification are the formulae defining 'salary' and bonus coefficients. In particular, the internal reporting, target-setting and accounting systems do not need to be changed.

Our results have the following implications for the theory of industrial organization. First, provided the required conditions on contracting sequence, verifiability of subcontracts and unlimited liability of intermediate agents hold, our model questions the common notion that larger, more complex hierarchies are less efficient owing to 'control losses' with respect to incentives or coordination.[6] Moreover, our results imply that *all* hierarchies consistent with the technology are equally efficient from an incentive standpoint; the choice of hierarchical structure can then be based entirely on considerations of communication costs or information processing. Second, our model identifies different potential sources of hierarchical control losses based on incentive problems: (i) collusion between agents, i.e., either a departure from the required top-down contracting sequence, or nonverifiability of subcontracting outcomes by the Principal (e.g, as in Laffont and Martimort (1998) or Mookherjee and Tsumagari (2001)); (ii) limited liability constraints or risk-aversion of intermediate agents (as in McAfee and McMillan (1995) or Faure-Grimaud, Laffont and Martimort (1998)).

Coordination problems in a setting with adverse selection incentive problems have also been explored by Crémer and Riordan (1987). A key difference is that contracting in their model occurs at an *ex ante* stage, before agents have received their private information. Informational rents cannot therefore be earned by agents, and they show that first-best allocations can be implemented by an equilibrium of the contracting game.[7] The paper most closely

[5]The three stages also correspond to planning and 'responsibility budgeting' procedures commonly observed in large firms (Horngren and Foster (1991)).

[6]For instance, popular accounts of recent waves of 'corporate engineering' and 'downsizing' are described in terms of eliminating middle layers of management and contract intermediaries, facilitated by advances in information technology and increased product market competition (Hammer and Champy (1993)). It is often presumed that these gains result partly from a reduction in the control losses associated with delegation of decision making to intermediaries. Such control losses are also believed to account for limits to firm size, as large firms organized in the form of corporate hierarchies tend to behave more 'bureaucratically'. The control losses are typically associated with problems of motivating and supervising intermediaries, and coordinating decisions made by disparate branches of the hierarchy (Williamson (1967, 1985)). Williamson's analysis has been criticized by Mirrlees (1976) and Calvo-Wellisz (1978) who provided models based on moral hazard in which control losses do not create a limit to firm size.

[7]Since contracting occurs at an interim stage in our model, agents earn informational rents, and we obtain the analogous result that second-best allocations can be attained. There are a number of other differences as well. Their paper is concerned with implementation of Pareto efficient allocations, whereas we are interested in allocations that maximize the principal's

related to this one is our earlier work (Mookherjee and Reichelstein (1997)), where a similar three stage mechanism was studied for a narrower class of production environments. More importantly, our earlier work assumed that the Principal contracted directly with all agents, whose role was limited to filling in certain parameters in contracts for subordinates. In contrast, contracting authority is fully delegated to intermediate agents in this paper.

Section 2 illustrates the main results in the context of an example of a three layer hierarchy with two branches. Section 3 introduces the general model, then formally defines a hierarchical contractual mechanism, the solution concept employed, and then states the main result (whose proof is contained in the Appendix). Finally, Section 4 concludes.

2 Example: A Three-Tier Two-Branch Hierarchy

In this section we present the key ideas of the paper in the context of an example involving six agents who jointly produce a marketable output for a principal (denoted P). The technology is hierarchically decomposable in the following sense: the revenue accruing to P, denoted $B(q_1, q_2)$, depends on the quantities q_1, q_2 of two goods or services produced separately. The production of good i requires the collaboration of three agents A_{i0}, A_{i1}, A_{i2}, and is described by the production function $q_i = \bar{q}_i(a_{i0}, a_{i1}, a_{i2})$, where a_{ik} represents the (real-valued) productive contribution of A_{ik}.

P's payoff is $B - \sum_{i=1}^{2}\sum_{k=0}^{2} x_{ik}$, where x_{ik} is the transfer from P to A_{ik}. P and all agents are risk-neutral. The payoff of A_{ik} is $x_{ik} - \theta_{ik}a_{ik}$, where θ_{ik} is a cost parameter known privately by A_{ik}.[8] Agents are not subject to limited liability constraints, and their outside option payoffs are normalized to zero. The belief of P and the other agents concerning θ_{ik} is represented by a positive density $f_{ik}(.)$ on support $[\underline{\theta}_{ik}, \bar{\theta}_{ik}]$. Beliefs regarding costs of different agents are mutually independent and common knowledge. They satisfy a monotone hazard rate condition ($\frac{1}{\theta_{ik}} \frac{F_{ik}(\theta_{ik}}{f_{ik}(\theta_{ik})}$ is nondecreasing in θ_{ik}), where F_{ik} denotes the distribution function of θ_{ik}.[9]

No further restrictions are imposed on the production structure. Hence the model is flexible enough to be applied to internal organization, regulation or

residual income. They address the problem that communication between a pair of agents is bilateral and private, while we assume that agents can verify cost reports exchanged between others. They present uniqueness results, whereas our equilibrium has a sequential dominant strategy feature.

[8]The results can be extended to the context where production costs are not necessarily separable in θ_{ik} and a_{ik}, but satisfy a generalized single-crossing condition (see Melumad-Mookherjee-Reichelstein (1995) for a precise statement).

[9]This condition is stronger than the usual monotonicity condition in adverse selection and auction models, and implies the optimality of a linear menu of incentive schemes.

procurement, and includes contexts of competition and/or coordination. For instance, product i could be a downstream product produced by A_{i0}, following supply of intermediate inputs a_{ik} by upstream suppliers $A_{ik}, k = 1, 2$; the two suppliers may supply complementary inputs, or may be competing suppliers of the same input. Similarly, products 1 and 2 could be complementary (e.g., where a final output is produced upon assembling these two products) or competing (e.g. alternative plants producing the same product). In a procurement setting, agent A_{i0} could be a 'prime' contractor, while $A_{ik}, k = 1, 2$ are subcontractors. In the internal organization setting, the firm could be composed of two product divisions $i = 1, 2$, with A_{i0} the manager of division i, producing product i in collaboration with her subordinates $A_{ik}, k = 1, 2$. It is possible to incorporate the notion that a more able 'manager' (agent A_{i0}) increases the marginal product of other team members: the ability of the manager being represented by the parameter θ_{i0}, a more able manager is represented by a lower realization of θ_{i0}. *Ceteris paribus* this will induce A_{i0} to supply a higher level of managerial effort a_{i0}, which raises the marginal product of other team members if the production function exhibits complementarity between their respective contributions.

The separable structure of the production function (first analyzed by Leontief (1947)) implies that the marginal rate of substitution between the outputs of any two agents involved in production of a given product i is independent of production of other products, enabling optimal production allocations to be made in a multistage hierarchical fashion. We first consider a team setting (in the sense of Radner and Marschak (1972)) where all agents share a common objective function equal to their collective profit $B - \sum_{i=1}^{2}\sum_{k=0}^{2}\theta_{ik}a_{ik}$. Then agents do not have any incentive to misrepresent their private information concerning their cost parameter. Figure 1 depicts a hierarchical mechanism which implements optimal production decisions. P designates agent A_{i0} as the 'manager' of product i, responsible for receiving cost reports from the other two agents A_{i1}, A_{i2}, making a report of the 'aggregate cost' of product i to P, receiving a production target for product i, and subsequently allocating production assignments (a_{i0}, a_{i1}, a_{i2}) amongst herself and her two subordinates.

In the team setting, the three-layer hierarchical mechanism consists of the following stages:

(i) $A_{ik}, k = 1, 2$ reports θ_{ik} to A_{i0}.

(ii) A_{i0} computes the cost function $c_i(q_i; \theta_i)$ of product i (where θ_i denotes the vector $(\theta_{i0}, \theta_{i1}, \theta_{i2})$), by minimizing $\sum_{k=0}^{2}\theta_{ik}a_{ik}$ subject to the constraint $\bar{q}_i(a_{i0}, a_{i1}, a_{i2}) \geq q_i$. The resulting cost-minimizing production assignments will be denoted $a_{ik}(q_i, \theta_i)$.

(iii) A_{i0} reports θ_i to P.

(iv) P computes optimal production targets $q_i(\theta_1, \theta_2)$ by maximizing $B(q_1, q_2) - \sum_{i=1}^{2} c_i(q_i, \theta_i)$, and communicates this to A_{i0}.

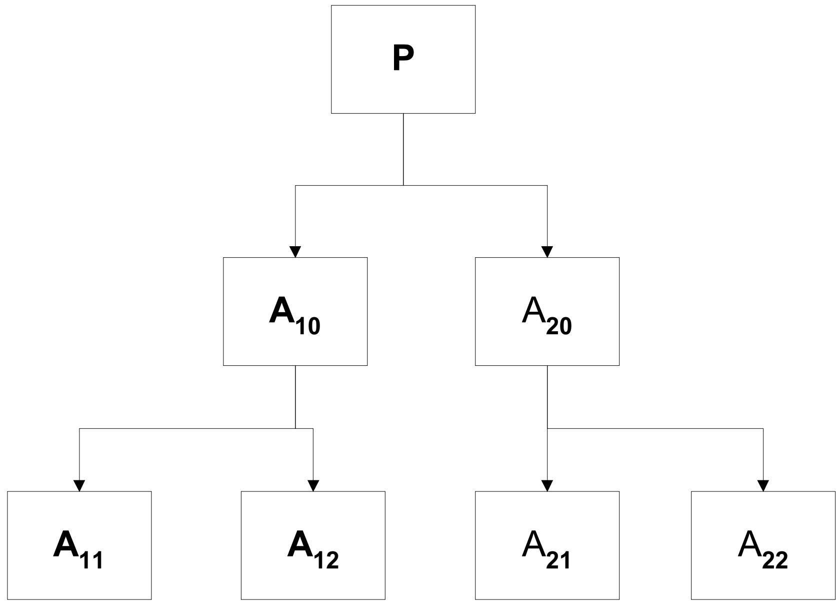

Figure 1: A Hierarchy with Three Layers and Two Branches

(v) A_{i0} communicates production targets $a_{ik}(q_i, \theta_i)$ to $A_{ik}, k = 1, 2$.

An alternative to such a three-layer hierarchical mechanism is a two-layer centralized revelation mechanism, where all agents communicate their cost reports to P, who subsequently allocates production assignments across all six agents and communicates these targets to them. Other nonhierarchical alternatives may include communication of cost reports by every agent to every other agent, each of whom subsequently solves the collective profit maximization problem and thereby decides on his own production assignment. As many authors (cited in the Introduction) have pointed out, one advantage of the three layer hierarchy is that it economizes on communication requirements: each agent communicates with at most two others. Moreover, it distributes the burden of computing optimal allocations across different agents who can work in parallel, enabling speedier decision making.

The question then naturally arises whether these economies of communication costs and information processing can continue to be realized in the presence of incentive problems. In such a setting first-best allocations are no longer attainable, owing to the fact that agents will earn informational rents. The appropriate benchmark of optimality is different in the presence of an incentive problem, as the cost to P of paying the agents their informational rents needs to be incorporated. In this setting, the Revenue Equivalence Theorem applies (see Myerson (1981)): the expected profit realized by the Principal from any Bayesian equilibrium of any mechanism depends only on the induced production assignments $a_{ik}(\theta)$ and the interim rents R_{ik} awarded to the highest cost types of $A_{ik}, i = 1, 2; k = 0, 1, 2$. In particular, P's expected profit equals the expected difference between gross benefits and the sum of 'virtual' costs and rents of the highest cost types:

$$\mathcal{E}_\theta\big[B(\{\bar{q}_i(a_{i0}(\theta), a_{i1}(\theta), a_{i2}(\theta))\}_{i=1,2}) - \sum_{i=1}^{2}\sum_{k=0}^{2}\{h_{ik}(\theta_{ik})a_{ik}(\theta)) + R_{ik}\}\big]. \quad (1)$$

The virtual cost parameter h_{ik} equals the cost θ_{ik} plus the informational rent $\frac{F_{ik}}{f_{ik}}$. A second-best mechanism sets the rents R_{ik} of the highest cost types to zero, and selects production assignments in any given state θ to maximize the difference between benefit and virtual cost in that state:

$$B(\{\bar{q}_i(a_{i0}, a_{i1}, a_{i2})\}_{i=1,2}) - \sum_{i=1}^{2}\sum_{k=0}^{2} h_{ik}(\theta_{ik})a_{ik}. \quad (2)$$

By the Revelation Principle, we know that second-best allocations can be implemented by a centralized revelation mechanism. The key question then is: *can they be equivalently implemented by a three layer hierarchical mechanism as well?*

2.1 Pure Vertical Control Problem: $B = B_1(q_1) + B_2(q_2)$

It is helpful to examine this problem in a setting where P's benefit function is additively separable between the production of the two products, so there is no need to coordinate the decisions of the two teams. Then effectively there are just three agents and a single branch. Since this case has already been studied in Melumad-Mookherjee-Reichelstein (1995), we outline the main steps of the analysis briefly. The problem reduces entirely to one of *vertical control*: the manager A_{i0} of product i is delegated authority over contracting and communicating with subordinates A_{i1}, A_{i2}. P's ability to monitor these subcontracts is typically limited: at best she can observe some of their implications in terms of production assignments or payments; it is rarer for P to also observe the actual subcontract or messages exchanged between them. Under these circumstances a self-interested manager A_{i0} can strategically misrepresent the cost function of product i to P, and bias production assignments between herself and her subordinates in order to inflate her own rents. This problem is essentially that of the *double marginalization of rents* arising from the exercise of monopoly power by the manager over contracting with subordinates.

Given the separability of P's revenue function, we can focus on the contracting problem with producers of product i in isolation. A *three-layer hierarchical mechanism with top-down contracting* for product i is defined to be a multi-stage game (depicted in Figure 2) where

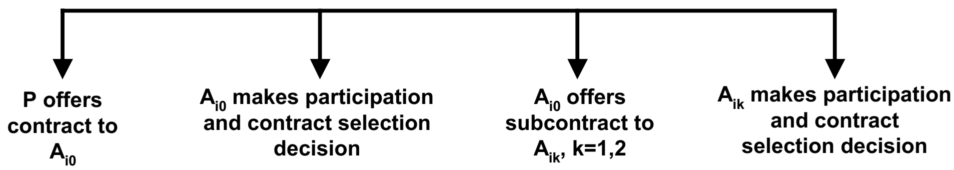

Figure 2: Sequence of events in three-layer hierarchy with additively separable departments

(i) at the first stage P offers A_{i0} a contract menu $x_i(\tilde{\theta}_i, q_i|C_i)$, where $\tilde{\theta}_i$ denotes a message to be sent by A_{i0} signifying choice of a contract from the menu, q_i is the output eventually delivered by A_{i0}, and $C_i \equiv x_{i1} + x_{i2}$ is the total 'cost', or payments to subordinates resulting from the subcontract between A_{i0} and her subordinates.

(ii) at the next stage A_{i0} responds to P's contract offer either by accepting or not accepting it; in the former case she selects a contract from the menu by reporting $\tilde{\theta}_i$, and designs a menu of subcontracts $x_{ik}(\tilde{\theta}_{ik}|M_{i0}, M_{i1}, M_{i2})$, $a_{ik}(\tilde{\theta}_{ik}|M_{i0}, M_{i1}, M_{i2})$, $k = 1, 2$, where $\tilde{\theta}_{ik}$ denotes a message by A_{ik} signifying a choice from the menu offered, and $M_{ik}, k = 0, 1, 2$ denotes cost report messages subsequently exchanged between the three agents.

(iii) Each subordinate A_{ik} responds with a decision whether or not to participate in the subcontract; in the former case he also selects $\tilde{\theta}_{ik}$ and reports M_{ik}.[10] Simultaneously A_{i0} selects the cost report M_{i0}.

(iv) If both agents agree to participate the game proceeds (otherwise it is terminated with no production or payments), production and payments of subordinates are determined as per the subcontract, and finally A_{i0} determines her own productive contribution a_{i0} (which along with production of subordinates determines the level q_i of the product delivered).

The top-down hierarchical mechanism has the following attributes:

Contracting Sequence: A_{i0} must respond to P's contract offer with an acceptance and contract choice from the offered menu *before* receiving any communication from subordinates. This implies that the participation constraints for the manager must hold in an *interim* rather than *ex post* sense. Note that P need not submit her own cost report M_{i0} to her subordinates before they have decided whether or not to accept the subcontract, so we allow for the problem of an informed (sub)principal in the sense of Maskin and Tirole (1990).

Contract Observability P can verify the total payment to subordinates resulting from the subcontract; no other aspect of the subcontract need be observed. A

[10] The reader may observe that two reports for each agent are redundant in this setting. They will, however, turn out to be relevant in more complex hierarchies so we shall retain this structure in order to maintain consistency in the exposition.

natural interpretation of this assumption in the internal organization context is that x_{ik} is a payment from P's account to A_{ik} which is 'authorized' by his manager A_{i0}. Alternatively in a procurement setting, P needs to verify the prime contractor's outsourcing cost.[11] It also simplifies to initially assume that subordinates observe the contract negotiated by A_{i0} with P; in the absence of this assumption it will suffice for subordinates to know that A_{i0}'s compensation is linear in the payments made to them.[12]

In this context *a three layer hierarchical mechanism with top-down contracting implements second-best outcomes.*[13] We outline the argument briefly. The menu offered by P to A_{i0} is the following family of contracts linear in a measure of 'profit' $\pi_i \equiv B_i(q_i) - \sum_{k=1}^{2} x_{ik}$:

$$x_i(\tilde{\theta}_{i0}, \pi_i) = \gamma_i(\tilde{\theta}_{i0}) + \beta_i(\tilde{\theta}_{i0})\pi_i.$$

The profit bonus coefficient $\beta_i(\tilde{\theta}_{i0}) \equiv \frac{\tilde{\theta}_{i0}}{h_{i0}(\tilde{\theta}_{i0})}$, which lies between 0 and 1, is non-increasing in $\tilde{\theta}_{i0}$. The manager's fixed compensation γ_i is a non-decreasing function of $\tilde{\theta}_{i0}$. Reporting a lower value of $\tilde{\theta}_{i0}$ is thus tantamount to the manager self-selecting a more 'high-powered' incentive contract.

The managerial contract overcomes the problem of double marginalization of rents in the following way. Suppose that A_{i0} has indeed selected the contract corresponding to her true type θ_{i0}. Then in designing the subcontract her problem reduces to selecting (upon applying the Revelation Principle to this problem, and using θ to denote the vector $(\theta_{i1}, \theta_{i2})$) a production allocation $a_{ik}(\theta)$ and payments $x_{ik}(\theta)$ to maximize

$$\mathcal{E}_\theta[\beta_i(\theta_{i0})\{B_i(\bar{q}_i(a_{i0}(\theta), a_{i1}(\theta), a_{i2}(\theta))) - \sum_{k=1}^{2} x_{ik}(\theta)\} - \theta_{i0}a_{i0}(\theta)]$$

subject to incentive and participation constraints for A_{i1}, A_{i2}. Since the subordinate agents must be paid their informational rents, and since $\frac{\theta_{i0}}{\beta_i(\theta_{i0})} \equiv h_{i0}(\theta_{i0})$, this reduces to the unconstrained maximization of

$$\mathcal{E}_\theta[B_i(\bar{q}_i(a_{i0}(\theta), a_{i1}(\theta), a_{i2}(\theta))) - \sum_{k=0}^{2} h_{ik}(\theta_{ik})a_{ik}(\theta)]$$

which is exactly the objective function of P. The manager fully internalizes the Principal's objective, and thus selects second-best production assignments, while minimizing rents of the subordinate agents. Finally, the fixed salary component

[11] Of course, we must assume there is no scope for any hidden side-payments (such as bribes) between the manager and subordinates.

[12] This issue is elaborated in footnote 14 below.

[13] The implementation notion employed is that every perfect Bayesian equilibrium outcome of the game will result in second-best profit for P.

$\gamma_i(\theta_{i0})$ is set to ensure that A_{i0} selects $\tilde{\theta}_{i0} = \theta_{i0}$ and the interim rents of her highest cost type equals zero. Hence the second-best outcome results.[14]

It is useful to understand the role of the key assumptions underlying the preceding result, which are typically necessary as well:[15]

Observability of Subcontracting Outcomes: This is necessary for P to overcome the problem of double marginalization of rents in the hierarchy. If P cannot observe any aspect of the subcontract, A_{i0}'s compensation must be independent of the subcontract. For any given production target q_i, A_{i0} will tend to allocate production assignments in order to minimize $\sum_{k=1}^{2} h_{ik}(\theta_{ik})a_{ik} + \theta_{i0}a_{i0}$ in state $(\theta_{i0}, \theta_{i1}, \theta_{i2})$, instead of $\sum_{k=0}^{2} h_{ik}(\theta_{ik})a_{ik}$. The informational rents paid to subordinates are treated as costs by A_{i0}, whereas the information rents accruing to A_{i0} herself are counted as a benefit. So A_{i0} tends to under-procure from the subordinates and inflate her own rents. Another interpretation of this problem is that A_{i0}'s monopsony power over procurement results in her offering too low a procurement price, in turn eliciting insufficiently low supplies from A_{i1}, A_{i2}. When subcontracting costs are observable, this problem is overcome by offer of a procurement subsidy by P to A_{i0}: for every additional dollar incurred in paying subordinates, A_{i0}'s compensation is lowered by β_i, which is lower than one dollar, thus corresponding to a per-dollar subsidy of $1 - \beta_i$. In the absence of cost observability, it would also suffice for P to monitor a_{i0} the productive contribution of the manager: then the procurement subsidy could be set equal to her informational rent $[h_{i0}(\theta_{i0}) - \theta_{i0}]a_{i0}$.[16]

Top-down Contracting and Absence of Limited Liability Constraint: These assumptions ensure that the managerial contract is subject only to incentive and (interim) participation constraints for the manager. Otherwise the contract for

[14] The informed principal problem (resulting from the fact that the manager's type θ_{i0} is unknown to her subordinates at the time of contracting) is also overcome as a result of the managerial contract that makes the manager's compensation a linear function of the total payments to the subordinates. This effectively makes the manager 'risk-neutral' with respect to these payments, as are the subordinates themselves. The analysis of Maskin and Tirole (1990) shows that in such environments the informed principal problem vanishes: the only reason for the manager to delay revelation of her own type to her subordinates is to enter into some risk-sharing arrangement with them with regard to the realization of their payments. With risk-neutrality there is no scope for gains from such forms of risk-sharing. It is evident from this that the subordinates do not really need to observe the exact contract received by their manager: it suffices that it is common knowledge that the manager is risk-neutral with respect to the payments to the subordinates.

[15] The precise conditions under which they are necessary include continuous differentiability of the revenue function (which ensures some degree of substitutability between contributions of different agents), and indispensability of both agents in the second-best allocation (i.e., each agent produces with positive probability).

[16] With this alternative formulation of the procurement subsidy, the second best is implementable even if cost functions are not multiplicatively separable, but instead satisfy a general single-crossing property.

A_{i0} would be subject to an additional constraint of the form

$$x_{i0}(\theta) - \theta_{i0} a_{i0}(\theta) \geq 0$$

for *all* θ. This could result directly from the inability of A_{i0} to assume a positive liability (i.e., be remunerated below cost) in any state. Alternatively *ex post* participation constraints would have to be imposed if A_{i0} could communicate and contract with subordinates *before* responding to the Principal's offer. In such a context, A_{i0}'s *ex post* rents would have to be non-negative in all states of the world θ, implying in turn that *interim* rents will be strictly positive owing to incentive constraints. Intuitively, in order to shrink interim rent R_{i0} of A_{i0} to zero, as required by the second-best allocation, A_{i0} must be able to costlessly bear risk (or positive liability) arising from the dependence of her performance measure on costs of subordinates. This is exactly the result of McAfee and McMillan (1995): the informational asymmetry between P and A_{i0} now expands to include information concerning the costs of subordinates A_{i1}, A_{i2}, and the resulting rents cannot be 'taxed away' upfront at the time of contracting.[17]

2.2 Vertical Control-cum-Coordination: Non-Additively-Separable Revenue Function

We now introduce the problem of coordinating production across the two divisions, by dropping the assumption of additive separability of B in q_1 and q_2. The optimal level of production of one product now depends on the cost of the other product; hence manager A_{i0} can no longer make decisions concerning production of q_i in isolation from information concerning cost conditions of product j. However, the assumption of weak separability (or hierarchical decomposability) of the firm's revenue function implies that this interdependence is limited to decisions concerning the *scale* of production: conditional on the overall level of the service q_i to be delivered, A_{i0} can still decide how to allocate productive contributions between herself and her subordinates in isolation from the other branch.

The three-layer hierarchical mechanism must now involve the (division i) manager signing subcontracts for her subordinates that are conditioned on the production target eventually assigned for their division (equivalently on the reported cost of the other division j). After receiving cost reports from subordinates, the manager must subsequently make a 'divisional' cost report to the Principal. In general this involves reporting the entire cost vector $\theta_i \equiv (\theta_{i0}, \theta_{i1}, \theta_{i2})$, which enables P to compute the cost function of product i. Having received cost reports from both divisions, P can decide on the desired scale of production for both products, and communicate these targets to the divi-

[17] In a similar vein, the recent theory of delegation of Faure-Grimaud, Laffont and Martimort (1998) is based on risk-aversion of the supervisor.

sional managers. Thereafter production assignments and payments within each division are determined according to the conditional contracts previously signed.

One frequently mentioned attribute of a hierarchical reporting mechanism is that lower level managers can aggregate their information concerning costs of alternative subordinates into a summary statistic of cost of their own department as a whole. Then economies can be realized in communicational costs and information processing burdens on higher level managers. In the formulation adopted above, an intermediate manager is communicating the entire vector of subordinates' cost parameters to P. Aggregation of cost reports into a single dimensional cost aggregate is indeed possible when the production function $\bar{q}_i$ is homothetic. Then A_{i0} need only report a unit cost report for product i to P (as illustrated in the multi-plant production planning example in Section 3.6 below). In case the production function is not homothetic, such aggregation is not possible and intermediate managers must generally submit multidimensional cost reports. Since multidimensional incentive problems are generally complicated, it is worthwhile to continue with the general case to show how these problems can nevertheless be overcome.

We introduce the following notation for the two distinct cost reports issued by a manager: an *internal* cost report m_i issued to subordinates for the purpose of allocating production within the division, and an *external* cost report $M_i \equiv (r_{i0}, r_{i1}, r_{i2})$ issued to P for the purpose of deciding the overall scale of production (where r_{ik} denotes the reported cost parameter for A_{ik}). As for the subordinate agents, we shall continue to use M_{ik} to denote their cost report to A_{i0}.

Formally, a *three-layer hierarchical mechanism with top-down contracting and bottom-up cost reporting* is the following multistage game (depicted in Figure 3):

(i) P offers A_{i0} a contract menu $q_i(\tilde{\theta}_{i0}, M_1, M_2), x_i(\tilde{\theta}_{i0}, M_1, M_2|C_i)$, where $\tilde{\theta}_{i0}$ is the message to be sent by A_{i0} signifying choice of a contract from the menu, q_i is the output target for product i, and C_i continues to denote 'cost' incurred in division i (such as costs or production assignments).

(ii) A_{i0} responds to P's contract offer either by accepting or not accepting it; in the former case she selects a contract by reporting $\tilde{\theta}_{i0}$, and then designs a menu of subcontracts for subordinates $x_{ik}(\tilde{\theta}_{ik}, M_j|m_i, M_{i1}, M_{i2})$, $a_{ik}(\tilde{\theta}_{ik}, M_j|m_i, M_{i1}, M_{i2})$, $k = 1, 2$, combined with an external cost reporting rule $M_i(m_i, \{\tilde{\theta}_{ik}, M_{ik}\}_{k=1,2})$, where $\tilde{\theta}_{ik}$ denotes a message by A_{ik} signifying a choice from the menu offered, and $m_i, M_{ik}, k = 1, 2$ denotes cost report messages subsequently exchanged between the three agents. Here m_i denotes a report made by A_{i0} to her subordinates concerning her own cost, while M_{ik} is the cost report of subordinate A_{ik}.[18]

[18] Note the dependence on the external cost report M_j of the other division, necessary for coordinating production between the two divisions. Otherwise the subcontract is designed as

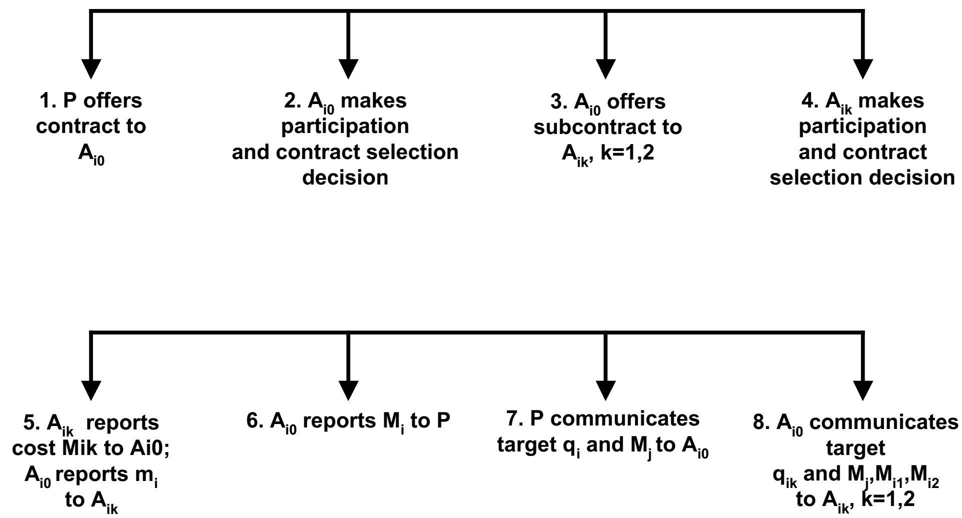

Figure 3: Sequence of events in three-layer hierarchy with non-additively separable departments

(iii) Each subordinate A_{ik} responds with a decision whether or not to participate in the subcontract; in the former case he also selects $\tilde{\theta}_{ik}$ and the report M_{ik}. Simultaneously A_{i0} selects the cost report m_i. If all managers and subordinates have agreed to participate the game proceeds (otherwise it is terminated with no production or payments).

(iv) $A_{i0}, i = 1, 2$ makes an external cost report $M_i \equiv (r_{i0}, r_{i1}, r_{i2})$ for division i to P (as per the subcontract already negotiated with subordinates), which determines output targets q_1, q_2 for each division (as per the managerial contracts already negotiated). P then communicates the output target q_i and the cost report of the other division M_j to all members of division i.

(v) Production assignments and payments to subordinates are now determined according to the conditional subcontracts selected at the first stage; given these and the divisional output target, the required production contribution a_{i0} of the manager is determined. Finally, payments to the manager are made according to the contract signed with P.[19]

Observability and timing assumptions are otherwise as in the case where B is additively separable. Note, however, that the necessity of coordinating production across the two divisions complicates the contracting considerably:

in the previous sub-section, to allocate production assignments within the division efficiently. The reporting rule $M_i(.)$ is independent of the cost report M_j of the other division, since the manager of different divisions report simultaneously to the Principal.

[19]Stages (iv) and (v) thus do not involve any strategic decisions, being determined entirely by decisions made earlier in the game.

divisional managers must now select subcontracts that are conditioned on outcomes in other divisions. The production assignment and payment of every agent depends on contracting and communication elsewhere in the organization. Each agent must also be able to verify the external cost reports of all divisions, as their own contract is conditioned on these variables.

Nevertheless, *the hierarchical mechanism continues to attain second-best outcomes.* The reasoning is as follows (we postpone a formal statement of the result and detailed proof to Section 4). The mechanism presented in the previous subsection is adapted in the following way. P offers A_{i0} the following menu of linear contracts:

$$x_i = \gamma(\tilde{\theta}_{i0}|M_j) + \beta_i(\tilde{\theta}_{i0})\pi_i$$

and

$$q_i(M_1, M_2) \in \arg\max_{q_1,q_2}[B(q_1, q_2) - \sum_{i=1}^{2} c_i(q_i|M_i)].$$

The divisional profit measure is now

$$\pi_i = R_i(q_i|M_j) - \sum_{k=1}^{2} x_{ik}$$

where R_i denotes an imputed measure of divisional 'revenue':

$$R_i(q_i|M_j) = \max_{q_j}[B(q_1, q_2) - c_j(q_j|M_j)]$$

equal to the maximum willingness of P to pay for the output delivered by division i, given the reported cost function for the other division j. The external cost report $M_j \equiv (r_{j0}, r_{j1}, r_{j2})$ is a vector of reported costs of the three producing agents in division j, which determine its cost function

$$c_j(q_j|M_j = (\{r_{jk}\})) \equiv \min_{a_{j0},a_{j1},a_{j2}} \sum_{k=0}^{2} r_{jk}a_{jk} \quad \text{subject to} \quad \bar{q}_j(a_{j0}, a_{j1}, a_{j2}) \geq q_j.$$

The problem of vertical control is now additionally complicated by the fact that each divisional manager A_{i0} is delegated the responsibility of making an external cost report M_i for purposes of production coordination, after having exchanged internal cost reports (m_i, M_{i1}, M_{i2}) with subordinates. And since the Principal does not observe the communication within each division, there is a multidimensional problem of ensuring that each manager report the entire vector of internal cost parameters truthfully (i.e., $M_i \equiv (r_i, r_{i1}, r_{i2}) = (m_i, M_{i1}, M_{i2})$).

Fortunately, however, this problem can be overcome by the mechanism described above, owing to a dominant strategy feature of the above mechanism. Conditional on having selected a contract corresponding to her true type ($\tilde{\theta}_i = \theta_{i0}$), the objective of manager of division i reduces to

$$\mathcal{E}_{M_j,\theta_{i1},\theta_{i2}}[R_i(q_i(M_1, M_2)|M_j) - \sum_{k=0}^{2} h_{ik}(\theta_{ik})a_{ik}]$$

upon incorporating the incentive and participation constraints for the subordinates, and the fact that the unit effective cost of any personal contribution is the virtual cost h_{i0} rather than the actual cost θ_{i0} owing to the profit bonus coefficient β_i in her compensation formula. Here M_j — the cost report of the other division — is treated as a random variable; subcontracts within division i can be conditioned on the eventual realization of M_j. So the manager's control variables are the conditional production assignments $a_{ik}(\theta_{i0}, \theta_{i1}, \theta_{i2}|M_j)$, and the divisional cost report $M_i(\theta_{i0}, \theta_{i1}, \theta_{i2})$ which affects the eventual production target q_i. Utilizing the definition of the divisional revenue measure R_i, the manager of division i seeks to maximize pointwise (i.e., for any given realization of $(M_j, \theta_{i1}, \theta_{i2})$) the following objective function:

$$\max_{q_j}[B(q_i(M_i, M_j), q_j) - c_j(q_j|M_j)] - \sum_{k=0}^{2} h_{ik}(\theta_{ik})a_{ik}.$$

In effect, thus, manager A_{i0} again internalizes the Principal's objective perfectly, conditional on any cost report M_j submitted by the other division j. It follows that each manager has a dominant strategy incentive (i.e., conditional on any M_j) to truthfully report its division's cost to the Principal (i.e., A_{i0} reports $M_i = \{(h_{ik}(\theta_{ik})\}_{k=0,1,2})$ and select second-best production assignments within division i. This applies irrespective of the dimensionality of the cost vector M_i.

Given that divisional managers report their virtual cost function truthfully, the aggregate production levels are also chosen by the Principal according to the second-best criterion: they maximize the difference between revenues and aggregate virtual cost. So production decisions are second-best. Finally, the fixed payments γ_i of each manager have to be calibrated to ensure they select a contract corresponding to their true type, and earn the same *interim* rents as in an optimal revelation mechanism. Owing to the assumption of top-down contracting and unlimited liability, this can be accomplished in the usual manner. At the time of contracting, therefore, each manager continues to bear the risk of a negative *ex post* liability if her subordinates happen to realize a high cost, but not with respect to cost realizations of other divisions.

A nice feature of the mechanism is that agents in any division do not have to form conjectures regarding the outcomes in the other division, despite the fact that their contracts are conditioned on these outcomes. The equilibrium has a *sequential dominant strategy* property: bottom-layer agents are playing dominant strategies; conditional on this, intermediate managers are also playing dominant strategies (i.e., the managers' strategies are dominant with respect to the cost reported by the other division, and a best response to the strategies of their subordinates). This renders inessential the need for agents in any division to monitor contracts and motivation of agents in divisions apart from their own.[20]

[20] Of course agents in division i need to be able to *ex post* verify the cost report M_j submitted by the other division, since their own contracts are conditioned on this variable. But they do

In summary, therefore, previous results concerning implementability of second-best outcomes by a three-layer hierarchical extends to the three-layer six-agent context where coordination and vertical control problems arise simultaneously. Observe that the same conditions continue to be necessary and sufficient for this to happen: top-down contracting, absence of limited liability for intermediate managers, and P's ability to monitor costs or production assignments within divisions. In the next Section, we show that this result extends to arbitrary hierarchical structures. There a formal analysis is presented which clarifies the exact assumptions concerning observability, sequence of moves, and the implementation concept for the hierarchical contracting game.

3 General Model

There are n productive agents, with agent i contributing $a_i \in \mathcal{A}_i$, an interval of the real line, with the principal's gross benefit function given by $B(a_1, \ldots, a_n)$, and net payoff by $B - \sum_{i=1}^{n} x_i$. All the assumptions and notation of the previous section are retained: agent i's payoff is $x_i - \theta_i a_i$, where the realization of θ_i is known privately by i. It is common knowledge among all agents and the principal that θ_i is distributed on the support $[\underline{\theta}_i, \bar{\theta}_i]$ with a positive density $f_i(.)$, where $\frac{F_i}{\theta_i f_i}$ is nondecreasing in θ_i, and $\theta_1, \ldots, \theta_n$ are mutually independent. The corresponding distribution function is denoted by F_i. It will be notationally convenient to refer to the Principal as agent 0, for whom $a_0 \equiv 0 \equiv \theta_0$.

We first explain the hierarchical structure of the technology. A *hierarchy* is an oriented tree for P and the n agents, with P at the root of this tree. Agent j is said to be a subordinate of Agent i, if the path from P to j leads through i. Agent i is said to be a direct boss of j (and j a direct subordinate of i) if in addition there is no third agent k such that k is subordinate to i and j is subordinate to k. Otherwise i is said to be an indirect boss of j (and j is an indirect subordinate of i. The set of direct subordinates of i is denoted $s(i)$, and the entire set of subordinates of i is denoted $S(i)$. Agent i can therefore be thought of as responsible for *department* $d(i) \equiv \{i\} \bigcup S(i)$, *i.e.,*which includes himself and all his subordinates. The assumption of hierarchical decomposability of the technology requires every department $d(i)$ to collectively produce a product q_i, in isolation from the rest of the organization, i.e., based only on the efforts of the members of $d(i)$. This motivates the following definition.

Definition 1 *The hierarchy $\mathcal{H}$ is said to be* **consistent** *with the production function B if it admits the following decomposition:*

(i) For each agent i, there is an output q_i delivered by the corresponding department $d(i)$ which is given by a production function $q_i = \bar{q}_i(a_i; q_{s(i)})$,where

not need to observe the contracts in other divisions, or form expectations concerning decisions made by agents in other divisions, in order to decide how they must themselves behave.

$q_{s(i)}$ denotes $\{q_j\}_{j\in s(i)}$, the vector of outputs produced by the departments managed by the immediate subordinates $s(i)$ of i.[21] *If i has no subordinates (i.e. is a bottom-layer agent), then q_i is a function of a_i alone.*

(ii) The Principal's gross benefit can be expressed as a function of outputs of the departments of top layer agents: $B \equiv q_0 = \bar{q}_0(q_{s(0)})$.

Second-best mechanisms select production assignments to maximize the expected difference between P's gross benefits and the sum of virtual production costs of all agents. Given the consistency of the hierarchy with the technology, second-best production assignments can be hierarchically decomposed as follows. For a bottom layer agent i, use M_i^* to denote the agent's virtual cost parameter $h_i(\theta_i) \equiv \theta_i + \frac{F_i(\theta_i)}{f_i(\theta_i)}$. For an agent i at a higher level, it is defined recursively by $M_i^* \equiv (h_i(\theta_i), M_{s(i)}^*)$, where $M_{s(i)}^*$ is the vector of virtual cost parameters of subordinates. Then the virtual cost of department $d(i)$ for delivering one unit of output is given by:

$$C_i(q_i; M_i^*) \equiv \min_{a_i, q_{s(i)}} [h_i(\theta_i)a_i + \sum_{j\in s(i)} C_j(q_j; M_j^*)] \quad \text{subject to} \quad \bar{q}_i(a_i; q_{s(i)}) \geq q_i. \tag{3}$$

Since a bottom-layer agent has no subordinates, $M_i^* = h_i(\theta_i)$ and $C_i(q_i; h_i(\theta_i)) = h_i(\theta_i)a_i(q_i)$, where $a_i(q_i)$ is the inverse of i's production function $\bar{q}_i(a_i)$. Let the second-best assignments which solve problem (3) be denoted by $a_i^*(q_i; M_i^*)$, $q_j^*(q_i, M_i^*)$.

3.1 Contracts

Define a *sibling* of agent i to be any other agent j who is a direct subordinate of i's direct boss. Next denote by $T(i)$ the set of i's siblings and also those of all his (direct and indirect) bosses. This set is defined recursively as follows. For a top layer agent $i \in s(0)$, $T(j) \equiv s(0) - i$ is the set of all other top-layer managers. For an agent i at the next layer, it is the set of all top-layer managers apart from his boss, and i's siblings. In general, therefore, $T(j) \equiv T(i) \bigcup [s(i) - j]$, where i is j's direct boss. The set $T(j)$ comprises all those agents in the hierarchy at higher or the same level in other departments whose cost reports will affect output targets and payments of j.

Each agent i in the hierarchy submits an (external) cost report M_i to his direct boss: this is a vector of cost parameters for all members in the department reported to i's boss. Specifically, $M_i \equiv (m_i, \tilde{M}_{s(i)})$, where $\tilde{M}_{s(i)}$ denotes the vector of reports on behalf of subordinates of i, and m_i is a report of the cost of her own contribution a_i. The vector $\tilde{M}_{s(i)}$ is effectively a listing of cost parameters of all subordinates — direct and indirect — of i. In the case where

[21] In what follows, we shall use this convention of denoting subvectors by a subscript for the set of relevant agents.

departmental production functions are homothetic, these multidimensional cost reports can be aggregated into single-dimensional unit cost reports.

The subcontract stipulates i's report M_i as a function of the reports $M_{s(i)}$ that she will receive from her subordinates. Besides, the subcontracts specifies payments and production targets for all direct subordinates $s(i)$, based on the external cost reports $M_{s(i)}$ they submit to i, and the cost reports of other departments (i.e., members of $T(i)$) which will eventually determine the target for i's department as a whole. We shall use μ_i to denote the set of external cost reports submitted by agents in $T(i)$. Note that M_i cannot be conditioned on μ_i, since i's cost report must be issued *prior* to learning the realization of cost reports μ_i of other higher level departments. However, with the sequence of events described below, the output targets and compensations can be conditioned on the realization of μ_i.

A subcontract designed by agent i for her subordinates $s(i)$ therefore comprises:

(a) a cost reporting rule $M_i(M_{s(i)}))$

(b) output target rule for each direct subordinate $j \in s(i)$: $q_j(\tilde{\theta}_j, M_j|\mu_j)$

(c) compensation rule for each direct subordinate $j \in s(i)$: $x_j(\tilde{\theta}_j, M_j, x_{S(j)}|\mu_j)$ where $\tilde{\theta}_j \in [\underline{\theta}_j, \bar{\theta}_j] \bigcup \{NP\}$ denotes the participation and contract selection decision made by j. If $\tilde{\theta}_j = NP$, j decides to drop out, whereupon $q_k = x_k \equiv 0$ for all agents k in the hierarchy.[22] Otherwise j decides to participate, and submits a cost report $\tilde{\theta}_j$ which selects a contract from the menu offered to j.[23]

In what follows, we shall use Γ_j to denote the contract offered to j. Agent j subsequently designs a mechanism for all her direct subordinates, and so on. Contracts flow down the hierarchy in this fashion, until an agent j at the bottom layer is reached. For such an agent (with boss i), the form of the contract is modified to: $\Gamma_j = (M_i(M_{s(i)})), q_j(\tilde{\theta}_j, M_j|\mu_j), x_j(\tilde{\theta}_j, M_j|\mu_j))$ reflecting the absence of any subordinates. In what follows, we shall denote the set of all bottom-layer agents by J.

[22]This is natural if every agent is essential in the production process, as with a constant-elasticity-of-substitution revenue function. If some agents are not essential, then too we can impose the property that the entire project shuts down in case any agent happens to not participate without loss of generality, since this event will occur off the equilibrium path. A more realistic formulation would allow only those agents who decide not to participate to drop out, while others continue to perform. This can also be incorporated into the model by extending the (cost report) message space to incorporate participation decisions of subordinates. We avoid this in the interests of simplifying the notation.

[23]The compensation payment for j depends on $x_{S(j)}$, the payments eventually made to all of j's subordinates. In the case that cost functions are not multiplicatively separable between θ_j and a_j, but satisfy the relevant single-crossing conditions, they can be conditioned instead on observation of a_j.

3.2 Sequence of Moves

Figure 4 depicts the sequence of events. First P offers menus of contracts $\Gamma_{s(0)}$ to top layer agents. These agents then choose their own contracts, cost reporting rules, and offer menus of contracts $\Gamma_{s(i)}$ to their direct subordinates. Their subordinates do the same in turn, and thus contracts flow down the hierarchy, until a bottom-layer agent $i \in J$ is reached, who selects a contract.

In the second phase, the cost report of bottom-most agents are sent to their immediate superiors. The external cost report of each superior in turn is determined by the reporting rule already selected in the first phase. External cost reports flow up the hierarchy in this way, until they reach P. The final phase involves production assignments flowing down, in accordance with the reports submitted at the second phase and the contracts signed at the first phase.

3.3 Information Structure

For contracts to be enforceable, it is necessary that every agent i (and third party contract enforcers) be able to verify *ex post* the relevant cost messages on which the profit measure of i's department is based. The mechanism therefore necessitates maintaining an internal cost accounting system which makes public to all agents the cost reports submitted and costs incurred by each department.

An additional issue concerns *ex ante* observability of contracts and reports of higher layer agents, before an intermediate layer agent decides whether to participate in the mechanism and which contract to select. This will turn out to be irrelevant when we consider the sequential dominant strategy solution concept (explained further below). But it needs to be made explicit in the Bayesian analysis. In that context we seek to minimize 'informed (sub)principal' problems, and consider the Perfect Bayesian Equilibrium concept which is usually defined for multistage games where moves at all previous stages are observable to agents moving at later stages (see Fudenberg and Tirole (1991)). Hence in the Bayesian analysis we shall assume that every $j \in s(i)$ observes the contract Γ_m offered to every $m \in T(j)$. This assumption ensures, for instance, that a boss does not have more information concerning contracts offered to his siblings than do his subordinates. The only piece of private information possessed by a boss i thus concerns the realization of his private cost parameter θ_i.[24]

3.4 Strategies

Let the information possessed by an agent i be denoted $\mathcal{I}_i$ when it is i's turn to move. And let D_i denote i's decision with respect to participation and reporting

[24] It will turn out that standard contracts will make agent compensations linear in compensations of their subordinates, so that the informed principal problem will not be an issue once the Principal offers standard contracts to top-layer managers. So this observability assumption will not be necessary even in the Bayesian analysis.

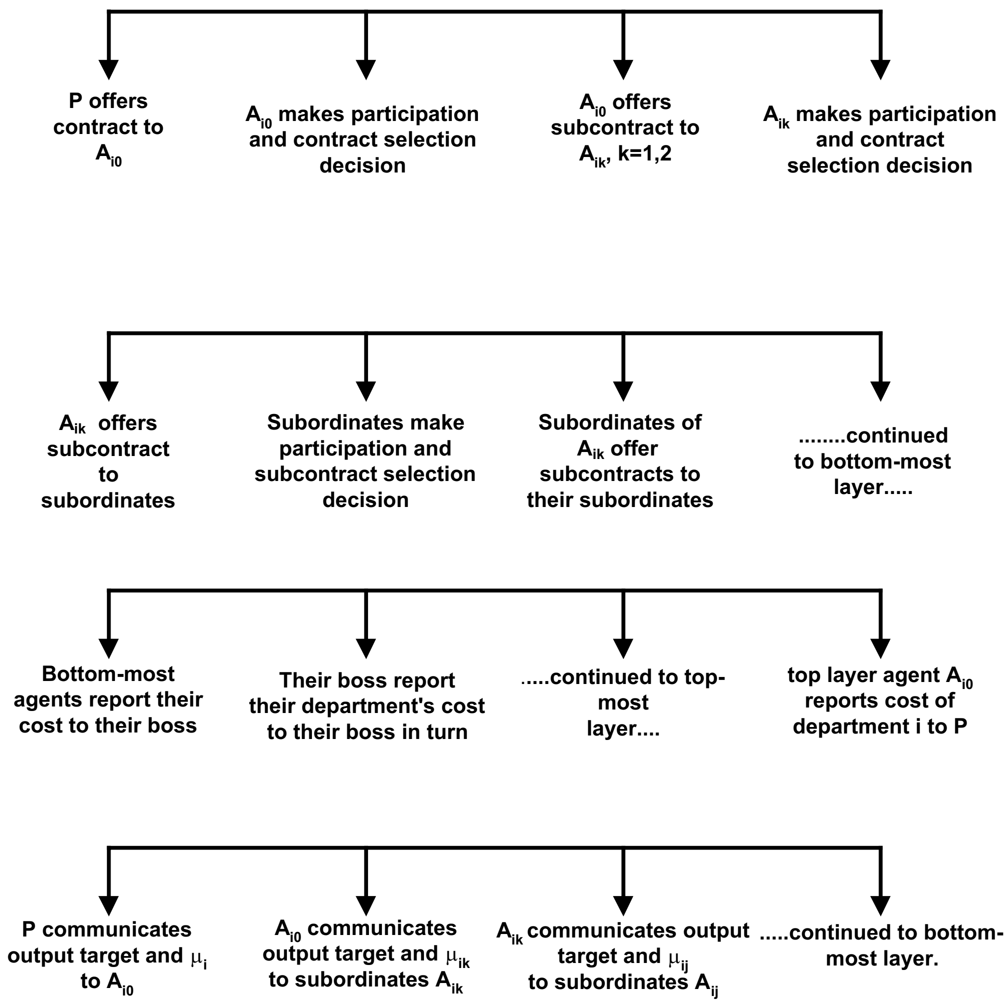

Figure 4: Sequence of events in a multi-layer/multi-branch hierarchy

of $\tilde{\theta}_i$ which serves to select a contract from the menu offered by i's boss. Then a *strategy* for $i \notin J$ is a function $\sigma_i(\theta_i, \Gamma_i|\mathcal{I}_i) \longrightarrow (D_i, \Gamma_{s(i)})$. For a bottom layer agent $i \in J$ it reduces to $\sigma_i(\theta_i, \Gamma_i|\mathcal{I}_i) \longrightarrow D_i$. Finally, P merely selects a menu of contracts to offer to each top layer agent.

3.5 Standard Contracts

We now introduce a specific class of contracts, which we call *standard contracts*. We shall argue below that agents will indeed have incentives to offer these contracts to their subordinates. First, the cost reporting rule takes the following form:

$$M_i(M_{s(i)}) = (m_i, M_{s(i)}) \tag{4}$$

i.e., i truthfully submits the report of subordinate costs exactly in the form that was submitted to i by her subordinates, adding to it a report of her personal cost m_i of contributing a_i.[25] Note in particular that i's report m_i of her personal cost parameter is independent of $M_{s(i)}$, so subordinates cannot influence m_i. In other words, i commits to reporting a given m_i at the time of offering a standard contract to subordinates.

Output and compensation rules are based on a measure of departmental profit, the difference between a measure of revenue and cost. For any top layer agent i, the revenue measure is

$$R_i(q_i|\mu_i) = \mathrm{Max}_{q_{s(0)-i}}[B(q_i, q_{s(0)-i}) - \sum_{l \in s(0)-i} C_l(q_l; M_l)] \tag{5}$$

where it may be recalled $\mu_i \equiv \{M_l\}_{\{l \in s(0)-i\}}$. Revenue functions for subordinates $j \in s(i)$ are then defined recursively as follows:[26]

$$R_j(q_j|\mu_j) = \mathrm{Max}_{a_i, q_{s(i)-j}}[R_i(\bar{q}_i(a_i, q_j, q_{s(i)-j})|\mu_i) - m_i a_i - \sum_{l \in s(i)-j} C_l(q_l; M_l)] \tag{6}$$

where $\mu_j \equiv (\mu_i, \{M_l\}_{\{l \in s(i)-j\}})$. The 'revenue' awarded to the department thus represents the 'maximum willingness to pay' of the boss i for the output of j's department when i seeks to maximize the profits of department $d(i)$, as measured by the difference between revenues R_i and costs reported by other subordinates $(s(i) - j)$ and i herself.

Given the revenue function, the output target for j is selected to maximize the difference between revenue and reported cost:

$$q_j(M_j|\mu_j) \in \arg\max_{\tilde{q}_j}[R_j(\tilde{q}_j|\mu_j) - C_j(q_j; M_j)]. \tag{7}$$

[25] The upper case report M_i denotes an (external) cost report submitted by i to i's boss, pertaining to cost of i's department as a whole. A lower case report m_i denotes an (internal) cost report submitted by i to i's subordinates concerning her personal cost of contributing a_i.

[26] To simplify notation, we suppress dependence on m_i, as the latter is stipulated as part of the contract.

Finally, the compensation scheme is structured as in the example of Section 3. Agent j is offered a menu of linear contracts

$$x_j = \gamma_j(\tilde{\theta}_j|\mu_j) + \beta_j(\tilde{\theta}_j)\Pi_j. \tag{8}$$

where the profit Π_i of i's department is defined by

$$\Pi_j = R_j(q_j|\mu_j) - \sum_{l \in S(j)} x_l, \tag{9}$$

the bonus coefficient is

$$\beta_j(\tilde{\theta}_j) = \frac{\theta_j}{h_j(\theta_j)}, \tag{10}$$

and the fixed payment is

$$\gamma_j(\tilde{\theta}_j|\mu_j) = \tilde{\theta}_j \hat{a}_j(\tilde{\theta}_j|\mu_j) + \int_{\tilde{\theta}_j}^{\bar{\theta}_j} \hat{a}_j(t|\mu_j)dt - \beta_j(\tilde{\theta}_j)\hat{\Pi}_j(\tilde{\theta}_j|\mu_j), \tag{11}$$

where $\hat{a}_j(\theta_j|\mu_j)$ equals the expected value of second-best assignment a_j^* conditional on: (i) j being of type θ_j, and (ii) the virtual cost reports μ_j made by members of $T(j)$ are truthful. The expectation is taken with respect to the types $\theta_{S(j)}$ of j's subordinates. Similarly, $\hat{\Pi}_j(\theta_j|\mu_j)$ denotes the corresponding conditional expectation of the profit measure Π_j.

3.6 Illustration of Standard Contracts in a Multi-plant Production Planning Example

Consider a firm producing a homogenous good across multiple plants, with a revenue function $B = bq - \frac{c}{2}q^2$, where q denotes the total quantity produced. There are n plants; if the manager of plant i contributes a_i the plant produces $q_i = (2a_i)^{\frac{1}{2}}$, so that $q = \sum_{i=1}^n q_i = \sum_{i=1}^n (2a_i)^{\frac{1}{2}}$. Plant i incurs a cost $\theta_i.a_i$, where θ_i is distributed uniformly on the interval $[0,1]$. This technology is consistent with *any* hierarchy, where the total output delivered by department D_i is given by $\bar{q}_i(a_i; q_{s(i)}) = (2a_i)^{\frac{1}{2}} + \sum_{j \in s(i)} q_j$.

It is convenient in this setting to represent the technology shock of plant i by its productivity $\eta_i \equiv \frac{1}{\theta_i}$. The virtual cost of manager i in supplying output q_i is $\frac{q_i^2}{\eta_i}$, and the second-best production assignment for plant i is $q_i = \frac{b\eta_i}{2+c\sum_j \eta_j}$, resulting in aggregate firm output $\frac{b\sum_j \eta_j}{2+c\sum_j \eta_j}$. For any hierarchy, the second-best allocation can be hierarchically decomposed as follows. Agent i will submit an internal (plant) productivity report of η_i to her subordinates, and an external (departmental) productivity report of $N_i = \eta_i + \sum_{j \in s(i)} N_j$. The total output for the firm is set by the principal as a function of the departmental reports of top layer managers j: $q = \frac{b\sum_{j \in s(0)} N_j}{2+c\sum_{j \in s(0)} N_j}$, with output target for department

$d(j)$ set at: $q_j = \frac{bN_j}{2+c\sum_{j\in s(0)} N_j}$. Output targets then flow down the hierarchy according to the following recursive rule: given target q_i for the department, $a_i = q_i \frac{\eta_i}{N_i}, q_j = q_i \frac{N_j}{N_i}$, for any $j \in s(i)$.

The revenue measures in the standard contract are set as follows. For a top layer agent i, $R_i(\hat{q}_i|\mu_i)$ is the maximized (with respect to $q_j, j \in s(0), j \neq i$) value of $b(\hat{q}_i + \sum_{j\neq i} q_j) - \frac{c}{2}(\hat{q}_i + \sum_{j\neq i} q_j)^2 - \sum_{j\neq i} N_j^{-1} q_j^2$. Solving this maximization problem, i's revenue is seen to depend on the reports of other departments via the single parameter $\delta_i = \frac{N_{-i}+c(N_{-i}^2-\tilde{N}_{-i}^2)}{(2+cN_{-i}^2)^2}$, where N_{-i} denotes $\sum_{j\in s(b_i), j\neq i} N_j$, and $\tilde{N}_{-i}^2$ denotes $\sum_{j\in s(b_i), j\neq i} N_j^2$. In the case of a top layer agent we can therefore identify μ_i with δ_i, and the corresponding revenue function is given by

$$R_i(q_i|\mu_i) = b^2\delta_i + (1-2c\delta_i)[bq_i - \frac{c}{2}q_i^2].$$

The revenue function handed to a top layer department is thus a marked down version of the revenue function facing the firm as a whole, with a markdown factor of $(1-2c\delta_i)$, plus an upfront fixed payment.[27] For lower layer agents, a similar revenue function is used, with a modification of the upfront payment and the markdown factor. We use F_i and G_i to denote the upfront payment and markdown factor in department i's revenues, *i.e.,* so that

$$R_i(q_i|\mu_i) = F_i + G_i[bq_i - \frac{c}{2}q_i^2].$$

These parameters of the revenue function are defined recursively as follows: $F_i = F_{b_i} + b^2 G_{b_i}\delta_i$; $G_i = (1-2c\delta_i)G_{b_i}$, where b_i denotes the boss of i. The parameter μ_i entering the revenue function offered to any department i is therefore summarized entirely by these two variables F_i and G_i.

The corresponding output target is $q_i(N_i, G_i) = b[c + 2N_i^{-1}{G_i}^{-1}]^{-1}$, while the required input in plant i is $a_i(N_i, \eta_i, G_i) = b\eta_i[cN_i^{-1} + 2{G_i}^{-1}]^{-1}$.

Finally, payment rules are as follows. The bonus coefficient β_i is set identically equal to $\frac{1}{2}$ (since θ_i is distributed uniformly on $[0,1]$). Hence $x_i \equiv \gamma_i + \frac{1}{2}[R_i - \sum_{j\in S(i)} x_j]$. The fixed payment γ_i is equal to the difference between two functions: γ_{i1}, the expected production cost plus informational rent of i, and γ_{i2} the expected value of the incentive payment $\frac{1}{2}[R_i - C_i]$, both calculated at the optimal production assignment. These functions are computed as follows:[28]

$$\gamma_{i1}(\eta_i, G_i) = b^2\Bigg[\eta_i^2 E_{N_{s(i)}}(c\eta_i + cN_{s(i)} + 2G_i^{-1})^{-2} + \int_0^{\eta_i} t^2 E_{N_{s(i)}}(ct + cN_{s(i)} + 2G_i^{-1})^{-2}dt\Bigg],$$

[27]It is easily verified that $\delta_i > 0$, hence $(1-2c\delta_i) < 1$. Moreover, if $c \in (0,1)$ then $(1-2c\delta_i) > 0$.

[28]We continue to treat η_i as the uncertain cost parameter, and use $N_s(i)$ to denote $\sum_{j\in s(i)} N_j$.

and

$$\begin{aligned}\gamma_{i2}(\eta_i, G_i) &= \frac{b^2}{2} G_i E_{N_{s(i)}} \Big[\{c + 2(\eta_i + N_{s(i)}{}^{-1} G_i{}^{-1})\}^{-1} \\ &\quad - \frac{c}{2}\{c + 2(\eta_i + N_{s(i)}{}^{-1} G_i{}^{-1})\}^{-2} \Big] \\ &\quad - E_{N_{s(i)}} \left[\frac{N_{s(i)}}{c(\eta_i + N_{s(i)}) + 2G_i{}^{-1}} \right].\end{aligned}$$

Note here that the design of the compensation rule for i requires information only about *aggregate* productivity $N_{s(i)}$ of the department $d(i)$, the sum of the individual η_i's across all plants in the department. The boss of i uses this aggregate information to design a suitable aggregate output target and incentive scheme for i, and then delegates responsibility to i to allocate this target between i's own plant and those of i's subordinates.

3.7 Straightforward Behavior

We now describe behavior 'on the equilibrium path' which if followed by all agents would result in the implementation of the optimal second-best mechanism. For any non-bottom-layer agent $i \notin J$, the strategy is said to be *straightforward* if conditional on receiving a standard contract, type θ_i of i selects the following actions: she accepts the contract corresponding to $\tilde{\theta}_i = \theta_i$, and offers a standard contract (with personal cost report $m_i = h_i(\theta_i)$) to all subordinates. For the Principal, straightforward behavior entails selecting a standard contract for all top-layer agents. Finally for bottom layer agents, straightforward behavior requires acceptance, $\tilde{\theta}_i = \theta_i$ and $M_i = h_i(\theta_i)$ following offer of a standard contract. Thus, straightforward behavior specifies that an agent, who is offered a standard contract, selects from the menu according to his true type, reports virtual costs truthfully, and offers a standard contract to every subordinate in turn.

PROPOSITION **1** *For any hierarchy consistent with the technology, straightforward behavior by the Principal and all agents results in second-best outcomes.*

The reasoning behind this result is that under straightforward behavior all agents report their virtual cost parameters truthfully. Production assignments are second-best because they are selected to maximize the difference between the principal's benefit function and the (reported) aggregate virtual cost. Finally, conditional on (θ_j, μ_j), the expected value of Π_j equals $\hat{\Pi}_j$, so every agent j ends up receiving the same *interim* rent as in the optimal revelation mechanism.

3.8 The Solution Concept

We first explain the two solution concepts employed, starting with the Bayesian one first. Note that we can express the payoff (hereafter denoted V^i) of agent i as a function of his own type θ_i, own decisions $\tilde{D}_i$ and contract Γ_i; contracts and decisions of all members of his department $d(i)$, besides the cost report vector μ_i. Thus

$$V^i = V^i(\tilde{D}_i, \Gamma_i, \tilde{\Gamma}_{S(i)}, \tilde{D}_{S(i)}(\theta_{S(i)})|\mu_i, \theta_i).$$

At the time that intermediate managers choose their own contracts and decisions, they do not yet know μ_i, the cost reports to be issued later by other departments that department $d(i)$ has to coordinate with. So contracts within department $d(i)$ will be conditioned on μ_i, and beliefs of members of the department regarding μ_i become important.

In analogy to the Perfect Bayesian Equilibrium concept, we require that members of department $d(i)$ hold common exogenous beliefs concerning μ_i; in particular, these beliefs cannot be affected by the actions of its boss i.[29] Under this assumption, we can treat agent i's mechanism design problem in isolation from other departments, with payoff functions for every member of $d(i)$ obtained by taking the expectation of their underlying payoffs with respect to these beliefs over μ_i.

Start with an agent i at a penultimate layer, and i and his subordinates have common exogenous beliefs over μ_i. A *Bayesian solution to the* $\mathcal{P}_i(\Gamma_i|\theta_i)$ *problem relative to common beliefs* $\mathcal{B}_i(\mu_i)$ is a vector of decisions and contracts for departmental subordinates $D_i^*(\theta_i), \Gamma_{s(i)}^*(\theta_i), D_{s(i)}^*(\theta_{s(i)}|\theta_i)$, which maximizes the expected payoff of type θ_i, i.e., $E_{\{\theta_{s(i)},\mu_i\}}\ V^i(\tilde{D}_i, \Gamma_i, \tilde{\Gamma}_{s(i)}, \tilde{D}_{s(i)}(\theta_{s(i)})|\mu_i, \theta_i)$, subject to the constraint that the decisions $\tilde{D}_{s(i)}(\theta_{s(i)})$ constitute a Bayesian equilibrium of the game played among subordinates $s(i)$, given $(\tilde{D}_i, \tilde{\Gamma}_{s(i)})$ the decisions and contracts selected by i, and given the use of beliefs $\mathcal{B}_i(\mu_i)$ over μ_i in computing conditional expectation of payoffs of i and all her subordinates. Specifically, the constraint requires that for each $j \in s(i)$, $\tilde{D}_j(\theta_j)$ maximizes $E_{\{\theta_{s(i)-j},\mu_i\}}\ V^j(D_j, \tilde{\Gamma}_j|\tilde{\mu_j})$, where $\tilde{\mu_j}$ denotes $(\mu_i, \{\tilde{D}_k(\theta_k)\}_{k\in s(i)-j})$.

The requirement of a Bayesian equilibrium amongst members of $s(i)$ requires that their beliefs regarding each others actions will be consistent with the strategies actually chosen. Moreover, these depend on the contracts chosen by i for his subordinates. In contrast, beliefs about the reports to be made by other departments are taken as given by i and his subordinates. This corresponds naturally to the standard Bayesian mechanism design problem faced by i with respect to his subordinates, given common exogenous beliefs concerning μ_i (on which no consistency conditions have yet been imposed).

[29] In the absence of this restriction, it is conceivable that i's mechanism choice for her subordinates affects their beliefs concerning μ_i. The scope for manipulating subordinates' beliefs in this way may motivate i to deviate to non-standard contracts. We argue below that this restriction on off-equilibrium path beliefs is in exact analogy to the belief restrictions imposed by the concept of a Perfect Bayesian Equilibrium.

Now move to higher layers, and define the Bayesian solution recursively. Having defined it for all agents at a given layer and below, consider an agent i at the next layer. A *Bayesian solution to the* $\mathcal{P}_i(\Gamma_i|\theta_i)$ *problem relative to beliefs* $\mathcal{B}_i(\mu_i)$ is $D_i^*(\theta_i), \{\Gamma_j^*(\theta_i), D_j^*(\theta_j|\theta_i), \Gamma_{s(j)}^*(\theta_{d(j)}|\theta_i), D_{s(j)}^*(\theta_{d(j)}|\theta_i)\}_{j\in s(i)}$ which maximizes

$$E_{\{\theta_{S(i)},\mu_i\}} \quad V^i(\tilde{D}_i, \Gamma_i, \{\tilde{D}_j(\theta_j), \tilde{\Gamma}_j, \tilde{D}_{S(j)}(\theta_{d(j)}), \tilde{\Gamma}_{S(j)}(\theta_{d(j)})\}_{j\in s(i)}|\mu_i, \theta_i)$$

subject to the constraint that for each $j \in s(i)$, $\tilde{D}_j(\theta_j), \tilde{D}_{S(j)}(\theta_{d(j)}), \tilde{\Gamma}_{S(j)}(\theta_{d(j)})$ is a Bayesian solution to the $\mathcal{P}_j(\tilde{\Gamma}_j|\theta_j)$ problem relative to beliefs $\mathcal{B}_j(\mu_j)$ over $\mu_j = (\mu_i, M_{s(i)-j})$, where $\mathcal{B}_j(\mu_j)$ is generated by the (independent) product of beliefs $\mathcal{B}_i(\mu_i)$ over μ_i, and beliefs over $M_{s(i)-j}$ generated by the (composition of) strategies actually pursued in sibling departments $(\{\tilde{D}_{d(k)}(\theta_{d(k)})\}_{k\in s(i)-j})$ with their prior beliefs over $\theta_{d(k)}$.

As before, beliefs concerning reports μ_i submitted by *other* departments are parametrically given, while beliefs concerning actions of members *within* the department $d(i)$ are consistent with the strategies actually chosen. In other words, it represents the sequential mechanism design problem within any department, relative to common beliefs with regard to the external environment.

Definition **2 A Bayesian solution to the Principal's problem in the contractual hierarchy** $\mathcal{P}_0$ *is* $\{D_i^*(\theta_i), \Gamma_i^*, D_{S(i)}^*(\theta_{S(i)}|\theta_i),\ \Gamma_{S(i)}^*(\theta_{S(i)}|\theta_i)\}_{i\in s(0)}$ *which maximizes the principal's payoff*

$$E_\theta V^0\big(\{\tilde{D}_i(\theta_i), \tilde{\Gamma}_i, \tilde{D}_{s(i)}(\theta_{d(i)}), \tilde{\Gamma}_{S(i)}(\theta_{d(i)})\}_{i\in s(0)}\big)$$

subject to the constraint that, for each $i \in s(0)$, $\tilde{D}_i(\theta_i)$, $\tilde{D}_{s(i)}(\theta_{S(i)}|\theta_i)$, $\tilde{\Gamma}_{S(i)}(\theta_{S(i)}|\theta_i)$ *is a Bayesian solution to the* $\mathcal{P}_i(\tilde{\Gamma}_i|\theta_i)$ *problem relative to beliefs* $\mathcal{B}_i(\mu_i)$ *generated by the composition of prior beliefs over* $\theta_{S(k)}, \theta_k$ *with strategies pursued by members of other departments* $\{\tilde{D}_k(\theta_k), \tilde{D}_{s(k)}(\theta_{S(k)}|\theta_k)\}_{k\in s(0)}$.

This solution reduces exactly to the solution concept used in existing literature on three tier contractual hierarchies (e.g. Baron and Besanko (1992), Laffont and Martimort (1998), or Melumad-Mookherjee-Reichelstein (1992, 1995)), involving a sequence of nested maximization problems. It differs, however, from the concept of Perfect Bayesian Equilibrium (PBE). The main difference is that the Bayesian solution as defined above specifies outcomes only on the equilibrium path; off-equilibrium-path actions and beliefs are not made explicit.

Nevertheless there is a close connection with PBE. First, the analogy with PBE motivates our assumption that beliefs concerning μ_i are exogenously given and commonly held within department $d(i)$. This follows from the assumption that all members of $d(i)$ observe the same variables concerning other departments headed by agents in $T(i)$ (i.e., the contracts offered to these agents). The 'no-signaling-what-you-don't know' restriction (B(iii) in Fudenberg-Tirole

(1995, p. 332)) on beliefs in a PBE imply that actions of members of $d(i)$ cannot alter beliefs of other members concerning the realization of μ_i. Moreover, restriction B(iv) in the definition of a PBE (Fudenberg-Tirole (1995, p. 332)) requires that all members of $d(i)$ must share the same beliefs concerning μ_i.

Moreover, our solution concept ensures that an intermediate manager would not benefit by deviating to a different set of contracts for his immediate subordinates. Here the outcome of any alternative deviating contract is represented by a Bayesian equilibrium of the ensuing continuation game, with expectation taken with respect to the types of subordinates, and beliefs concerning cost reports of other departments. Hence agents do form correct conjectures about the outcomes of off-equilibrium-path continuation games. Our solution concept incorporates this form of sequential rationality, without explicitly specifying strategies and beliefs for every possible continuation game. It is evident from this that every PBE outcome must be a Bayesian solution.

The converse would also be true if the existence of an equilibrium in every continuation game could be guaranteed. This would be the case, for instance, if each agent was restricted to a finite strategy set and mixed strategies were considered. In our setting agents have infinite strategy sets and we focus on pure strategies. Our solution concept thus enables us to avoid the purely technical problems involved in specifying equilibrium actions and beliefs in all possible continuation games, while incorporating the requirement of sequential rationality central to the PBE concept.

Since the Bayesian solution concept requires agents to form beliefs that are consistent with strategies pursued elsewhere in the hierarchy, it is demanding in terms of the informational and computational requirements on agents. For instance, they have to intelligently predict the behavior of *all* other agents in the organization. It is therefore worthwhile to explore an alternative solution concept which is significantly less demanding, in which members of each department pursue strategies that are dominant with respect to μ_i, the cost reports of other departments. This motivates the following definition.

Definition **3** **A sequentially dominant strategy solution to the** $\mathcal{P}_i(\Gamma_i|\theta_i)$ **problem** *is* $D_i^*(\theta_i),\ \{D_j^*(\theta_j|\theta_i),\ \Gamma_j^*(\theta_i), D_{S(j)}^*(\theta_{S(j)}|\theta_j),\ \Gamma_{S(j)}^*(\theta_{d(j)}|\theta_i)\}_{j\in s(i)}$ *which maximizes for any* $\mu_i = \bar{\mu}_i$*:*

$$E_{\theta_{S(i)}}[V^i(\tilde{D}_i, \Gamma_i, \{\tilde{D}_j(\theta_j), \tilde{\Gamma}_j, \tilde{D}_{S(j)}(\theta_{d(j)}), \tilde{\Gamma}_{S(j)}(\theta_{d(j)})\}_{j\in s(i)}|\bar{\mu}_i, \theta_i)]$$

subject to the constraint that for each $j \in s(i)$*:* $\tilde{D}_j(\theta_j), \tilde{D}_{S(j)}(\theta_{d(j)}), \tilde{\Gamma}_{S(j)}(\theta_{d(j)})$ *is a Bayesian solution to* $\mathcal{P}_j(\tilde{\Gamma}_j|\theta_j)$ *given that all subordinates in* $S(i)$ *share the belief that the distribution over* μ_i *is degenerate and concentrated on* $\bar{\mu}_i$*. Finally, define a sequentially dominant strategy solution to the Principal's problem as a solution to agent* 0*'s problem.*

Note that for a bottom layer agent, every other agent in the organization is 'external', so these agents must be selecting a strategy which is 'truly' dominant, *i.e.* with respect to the strategies of *all* other agents. For higher layer agents, the chosen strategy should be dominant with respect to actions of members of *other* departments, but optimal in a Bayesian sense with respect to the types of members of her *own* department. Beliefs concerning the outcomes in other departments do not play a role at all: each intermediate manager faces a conventional (Bayesian) contract design problem with respect to her subordinates, in which the outcomes of other departments enters as a parameter — and the same contract happens to be optimal no matter what this parameter happens to be.[30] This obviates the need for agents to predict the behavior of agents outside their own department.

It is evident from the definition that for bottom layer agents, sequentially dominant strategies are dominant, and hence Bayesian. This property does not, however, extend to intermediate layer agents. For instance, consider an agent i at the penultimate layer. In a Bayesian solution to i's problem, the continuation strategies for i's subordinates need not constitute dominant strategies (with respect to μ_i). Conceivably agent i may prefer to select a mechanism for her subordinates in which their incentive compatibility constraints hold when expectation is taken with respect to μ_i, but not for each and every possible realization of μ_i. Hence sequential dominant strategy implementation need not imply Bayesian implementation.

This is a shortcoming of the sequential dominant strategy solution: it restricts intermediate agents to continuation strategies that possess the sequentially dominant strategy feature, rather than to the broader class of Bayesian continuation solutions. So a Bayesian agent may prefer to design a mechanism for subordinates that does not possess the sequential dominance property. In that case, sophisticated intermediate managers will want to deviate: the contracts they select for their subordinates will need to be monitored by the Principal for their adherence to the requirement of sequential dominance.

The best of all possible worlds is where a solution is simultaneously both a Bayesian and a sequentially dominant solution. Then there would be no need for the Principal to monitor the subcontracts selected by intermediate managers; at the same time, agents would not need to predict outcomes in departments other than their own. It turns out that straightforward behavior does have this convenient dual property.

PROPOSITION **2** *For any hierarchy consistent with the technology, straightforward behavior is a Bayesian as well as a sequentially dominant strategy solution to the Principal's problem.*

The proof of this result is given in the Appendix. The key idea has already been explained in the three-layer example in Section 3: standard contracts cause

[30] See Crémer and Riordan (1985) for a similar sequential solution concept in the context of a public goods problem. In their context, which is much simpler than ours, all but one player of the game is playing a dominant strategy, and the last player is playing a best response.

each agent in the organization to internalize the Principal's objective. The sequential dominance feature of the solution follows essentially from the fact that optimal Bayesian mechanisms can be implemented in dominant strategies in this setting (extending the logic of Mookherjee and Reichelstein (1992)). For a penultimate layer agent, the optimal Bayesian mechanism for her subordinates can be implemented in dominant strategies, so the restriction imposed on her mechanism design problem has no bite. The same logic extends to upper layer agents as well.

Note finally a corollary of Proposition 2:

Corollary to Proposition 2 *All hierarchies consistent with the technology result in equivalent performance (represented either by Bayesian or sequentially dominant solution to the Principal's problem).*

In the example of multi-plant production planning in the previous subsection, this implies that *every* hierarchy can achieve second-best outcomes. Consequently, the choice of hierarchical design can be based entirely on considerations of economizing communication costs or information processing costs; incentive or coordination considerations will not play any role.

4 Concluding Comments

Our analysis establishes a benchmark result asserting the irrelevance of hierarchical size and structure with regard to control loss problems. Under the stated conditions (which include consistency of the hierarchy with the technology, risk neutrality and unlimited liability, appropriate sequencing of subcontracts and cost verifiability), design of the hierarchy can be based entirely on considerations of limiting costs of information processing and communication. Alternatively, insofar as incentive considerations do play a role in determining hierarchical structure, one or more of these stated assumptions must be violated. This suggests suitable directions for future research, i.e., models where control losses do arise, and control loss considerations interact with information processing or communication costs in determining the size and structure of efficient contractual networks.

One significant assumption concerns verifiability of procurement cost of subordinates, which amounts to absence of a certain form of collusion: the subsidy on procurement will tempt the agent to artificially pad costs, in exchange for hidden side payments received from subordinates. Future research needs to address the nature of losses encountered when such forms of collusion cannot be prevented (as exemplified in Laffont and Martimort (1998) or Mookherjee and Tsumagari (2001)). Alternatively, risk aversion or bankruptcy constraints for intermediate agents can cause hierarchies to incur incentive costs, as efficient hierarchical design necessarily requires these agents to bear risk associated with fluctuations in productivity of their subordinates (as in the approach of Faure-Grimaud, Laffont and Martimort (1998)). Finally, we hope that a theory of hierarchies based on a trade-off between their information processing advantages and incentive costs will eventually emerge.

Acknowledgments

We are very grateful to Ilya Segal and three anonymous referees for extremely detailed comments on the earlier draft of this paper. We also thank Tim van Zandt and Mike Riordan for useful discussions, and seminar participants at Alicante, Boston University, Carnegie-Mellon University, Maryland, Oslo and the 1999 Studienzentrum Gerzensee Summer Theory Workshop.

Appendix:Proof of Proposition 2

We first establish the sequential dominance property of straightforward behavior. Use Ω_j to denote a standard contract for agent j.

LEMMA **1** *Consider any agent i who is offered a standard contract Ω_i. Then straightforward behavior (i.e.,where for any $j \in d(i)$: $D^*_j(\theta_j)$ is given by $\tilde{\theta}_j = \theta_j, m_j = h_j(\theta_j), M_j(M_{s(j)}) = C_j(h_j(\theta_j), M_{s(j)})$, and $\Gamma^*_k(\theta_j) = \Omega_k$ for all $k \in s(j)$) is a Bayesian solution to the $\mathcal{P}_i(\Omega_i|\theta_i)$ problem, conditional on degenerate beliefs concentrated on an arbitrary $\bar{\mu}_i$.*

Proof of Lemma 1: This is proven inductively.

Step 1: Consider first a bottom layer agent $j \in J$ who is offered a standard contract, with $\bar{\mu}_j$ given. His problem is to select $\tilde{\theta}_j, M_j$ to maximize

$$V^j((\tilde{\theta}_j, M_j), S_j|\bar{\mu}_j) = \gamma_j(\tilde{\theta}_j|\bar{\mu}_j) + \beta_j(\tilde{\theta}_j)R_j(q_j(M_j \mid \bar{\mu}_j)|\bar{\mu}_j) - \theta_j a_j(q_j(M_j \mid \bar{\mu}_j)). \tag{12}$$

Conditional on a given $\tilde{\theta}_j$, then, the report M_j must be chosen to maximize

$$R_j(q_j(M_j \mid \bar{\mu}_j)|\bar{\mu}_j) - H_j(\theta_j, \tilde{\theta}_j)a_j(q_j(M_j \mid \bar{\mu}_j)). \tag{13}$$

where $H_j(\theta_j, \tilde{\theta}_j) \equiv \frac{\theta_j}{\beta_j(\tilde{\theta}_j)}$. Given the definition of the q_j function in a standard contract, it follows that the optimal cost report associated with any $\tilde{\theta}_j$ is $M_j = H_j(\theta_j, \tilde{\theta}_j)$, irrespective of the value of $\bar{\mu}_j$.

Next consider the optimal choice of $\tilde{\theta}_j$, combined with the cost report $H_j(\theta_j, \tilde{\theta}_j)$. Let $\tilde{V}^j(\theta_j, \tilde{\theta}_j \mid \bar{\mu}_j)$ denote this payoff. The Envelope Theorem implies that $\tilde{V}^j$ is differentiable in θ_j, with derivative given by

$$\frac{\partial \tilde{V}^j(\theta_j, \tilde{\theta}_j \mid \bar{\mu}_j)}{\partial \theta_j} = -a_j(q_j(H_j(\theta_j, \tilde{\theta}_j) \mid \bar{\mu}_j)) \tag{14}$$

which is nondecreasing in $\tilde{\theta}_j$. Let $v^j(\theta_j \mid \bar{\mu}_j)$ denote $\tilde{V}^j(\theta_j, \theta_j \mid \bar{\mu}_j)$. Given the construction of the fixed payment γ_j, it follows that

$$v^j(\theta_j \mid \bar{\mu}_j) = \int_{\theta_j}^{\bar{\theta}_j} a_j(q_j(h_j(t) \mid \bar{\mu}_j)dt. \tag{15}$$

Comparing (14) and (15), and using the facts that (i) $H_j(\theta_j, \theta_j) = h_j(\theta_j)$, (ii) $\frac{\partial \tilde{V}^j}{\partial \theta_j}$ is nondecreasing in $\tilde{\theta}_j$, it follows from applying Theorem 6.3 in Mirrlees

(1986) that $\tilde{\theta}_j = \theta_j$ is optimal for type θ_j, for any given $\bar{\mu}_j$. Hence conditional on participating in a standard contract, it is a dominant strategy for j to select $\tilde{\theta}_j = \theta_j$ and $M_j = h_j(\theta_j)$. Finally, note that the payoff from participation in the standard contract is given by (15), which is nonnegative irrespective of the realization of $\bar{\mu}_j$. Hence straightforward behavior is a dominant strategy for j.

Step 2: Now consider any agent i at a higher layer, all of whose subordinates behave straightforwardly. Given a standard contract Ω_i, and given any $\tilde{\theta}_i, \theta_i$, the objective that i will seek to maximize while designing a mechanism for his subordinates is given by

$$E_{\theta_{S(i)}}[R_i(\tilde{q}_i|\bar{\mu}_i, \bar{\nu}_i) - H_i(\theta_i, \tilde{\theta}_i)\tilde{a}_i - \sum_{l \in S(j)} \tilde{x}_l] \tag{16}$$

where $\tilde{q}_i, \tilde{a}_i, \tilde{a}_l$ denote the departmental output, personal input and payments to subordinates (all functions of $\theta_{S(i)}, \bar{\mu}_i, \bar{\nu}_i$) resulting from the continuation game in $d(i)$. Agent i selects $m_i, M_i(M_{s(i)})$ and contracts $\Gamma_{s(i)}$ for immediate subordinates, and is subject to the constraint that continuation strategies of these subordinates must form a Bayesian solution to their respective problems, relative to the belief that $\mu_i = \bar{\mu}_i$, and beliefs concerning each others reports which are consistent with strategies actually chosen.

The key step is to note that we can obtain an upper bound to the value of the maximand in (16) by applying the Revelation Principle to the mechanism design problem faced by i. A standard argument establishes that the Revelation Principle does apply to our sequential Bayesian solution concept. It therefore follows from standard manipulations of incentive and participation constraints that an upper bound to the objective function (16) of type θ_i of agent i who has received a standard contract, and selected $\tilde{\theta}_i$, is given by the maximum value of

$$E_{\theta_{S(i)}}[R_i(q_i(\tilde{\theta}_i, M_i(\theta_{S(i)}))|\bar{\mu}_j) - H_i(\theta_i, \tilde{\theta}_i)a_i(\theta_{S(i)}) - \sum_{l \in S(j)} h_j(\theta_j)a_j(\theta_{S(i)}))] \tag{17}$$

subject only to the production feasibility constraint

$$q_i(\tilde{\theta}_i, M_i(\theta_{S(i)})|\bar{\mu}_j) = \bar{Q}_i(a_{d(i)}(\theta_{S(i)})) \tag{18}$$

where $\bar{Q}_i(a_{d(i)})$ denotes the output of department $d(i)$ resulting from the production assignments $a_{d(i)}$ within the department. The maximization is carried out with respect to choice of an external cost reporting rule $M_i(\theta_{S(i)})$ and production assignment rule $a_{d(i)}(\theta_{S(i)})$ for the department.

It is evident that the solution to this problem can be obtained as follows. Corresponding to any given $\tilde{\theta}_i$ and any reporting rule $M_i(\theta_{S(i)})$ — which jointly determine the output target for the department — optimal production assignments $a_l(\theta_{S(i)})$ must solve the second-best cost minimization problem for department $d(i)$, corresponding to a input cost of $H_i(\theta_i, \tilde{\theta}_i)$ for agent i, and $h_j(\theta_j)$ for every other member j. Consequently, given the output target rule $q_i(.)$ in a standard contract, the upper bound (relative to $\tilde{\theta}_i$) is associated with a cost report M_i given by input cost $H_i(\theta_i, \tilde{\theta}_i)$ for a_i, and $h_i(\theta_i)$ for all subordinates, followed by a cost-minimizing production assignment relative to the resulting production target. Finally, by an argument analogous to that used for a bottom layer agent, the optimal $\tilde{\theta}_i = \theta_i$. Hence an upper bound to the payoff of type θ_i of i in a Bayesian solution following receipt of a standard contract equals his second best informational rent.

Returning now to the sequential contracting game, this upper bound can be achieved by straightforward behavior on i's part, given that her subordinates are all behaving straightforwardly. This proves Lemma 1.

To conclude the proof of Proposition 2, we need to show that straightforward behavior is a Bayesian solution to every agent's decision problem following offer of a menu of standard contracts, when that agent and her subordinates share exogenous beliefs over μ_i which are not degenerate. This follows from noting that all steps in the argument for intermediate layer agents (Step 2) in the proof of Lemma 1 above carry over when expectation is taken with respect to the common beliefs concerning μ_i. ■

References

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

Arrow K. (1974), *The Limits of Organization*, W.W. Norton, New York.

Baron, David and David Besanko (1992) "Information, Control and Organizational Structure" *Journal of Economics and Management Strategy*, Vol 1(1), 367-384.

Bolton, Patrick and Mathias Dewatripoint (1994) "The Firm as a Communication Network", *Quarterly Journal of Economics*, 109, 4, 809-839.

Calvo, G. and S. Wellisz (1978) "Supervision, Loss of Control and the Optimal Size of the Firm" *Journal of Political Economy*, vol. 86, 943-952.

Cremer, J. and M. Riordan (1985) "A Sequential Solution to the Public Goods Problem" *Econometrica* vol. 53, no. 1, 77-84.

————- (1987) "On Governing Multilateral Transactions with Bilateral Contracts," *Rand Journal of Economics*, vol. 18, no. 3, 436-451.

Faure-Grimaud, A., J.J. Laffont and D. Martimort (1998), "A Theory of Supervision with Endogenous Transaction Costs," mimeo, IDEI, University of Toulouse.

Fudenberg, D. and J. Tirole (1991) "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53: 236-260.

——— (1995), *Game Theory*, Cambridge, Mass: MIT Press.

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

Hammer M. and J. Champy (1993) *Reengineering the Corporation.* New York: Harper Collins.

Horngren, Charles and George Foster (1991) *Cost Accounting: A Managerial Emphasis, Seventh Edition*, Englewood Cliffs, NJ: Prentice-Hall Inc.

Keren, M. and David Levhari (1983) "The Internal Organization of the Firm and the Shape of Average Costs" *Bell Journal of Economics*, Vol. 14, 474-486.

Laffont J. and J. Tirole (1986) "Using Cost Observation to Regulate Firms," *Journal of Political Economy* 619-641.

——— (1993) *A Theory of Incentives in Procurement and Regulation*, Cambridge: MIT Press.

Laffont J. and D. Martimort (1998) "Collusion and Delegation," *Rand Journal of Economics* 29(2), 280-305.

Leontief W. (1947), "Introduction to a Theory of the Internal Structure of Functional Relationships," *Econometrica*, vol 15, No. 4, 361-373.

Maskin, Eric and Jean Tirole (1990) "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values" *Econometrica* vol. 58, No. 2, 379-409.

——— (1992) "The Principal-Agent Relationship with an Informed Principal,II: Common Values" *Econometrica* vol. 60, no. 1, 1-42.

McAfee, Preston and John McMillan (1995) "Organizational Diseconomies of Scale" *Journal of Economics and Management Strategy*, 4(3), 399-426.

Marschak T. and S. Reichelstein (1995) "Communication Requirements for Individual Agents in Networks and Hierarchies," *The Economics of Informational Decentralization: Complexity, Efficiency and Stability*, J. Ledyard (ed.), Boston: Kluwer Academic Publishers.

——— (1999), "Network Mechanisms, Informational Efficiency and Hierarchies," *Journal of Economic Theory.*

Melumad, Nahum, Dilip Mookherjee and Stefan Reichelstein (1995) "Hierarchical Decentralization of Incentive Contracts" *Rand Journal of Economics*, 26(4), 654-672.

——— (1992) "A Theory of Responsibility Centers" *Journal of Accounting and Economics* vol. 15, 445-484.

——— (1997) "Contract Complexity, Incentives and the Value of Delegation," *Journal of Economics and Management Strategy* 6(2), 257-289.

Mookherjee, Dilip and Stefan Reichelstein (1992) "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules, *Journal of Economic Theory,* 56:378-399.

——————- (1997) "Budgeting and Hierarchical Control," *Journal of Accounting Research*, 35(2), 129-155.

Mookherjee, Dilip and Masatoshi Tsumagari (2001), "Delegated Management and Contract Intermediation," mimeo, Department of Economics, Boston University.

Mirrlees, J.A. (1986) "The Theory of Optimal Taxation", Chapter 24 in Kenneth J. Arrow and Michael D. Intriligator, Eds., *Handbook of Mathematical Economics*, Vol. III, Amsterdam: North Holland.

————- (1976) "The Optimal Structure of Incentives and Authority within an Organization" *Bell Journal of Economics* vol. 7, 105-131.

Myerson, Roger (1981) "Optimal Auction Design" *Mathematics of Operations Research"* vol. 6, 58-73.

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

Radner, Roy (1992) "Hierarchy: The Economics of Managing" *Journal of Economic Literature* vol. 30, 1382-1415.

———— (1993) "The Organization of Decentralized Information Processing," *Econometrica* vol. 61, no. 5, 1109-1146.

————- and J. Marschak (1972), *Economic Theory of Teams*

——— and Timothy van Zandt (1992) "Information Processing in Firms and Returns to Scale" *Annales d'Economie et de Statistique* vol. 25-26, 265-298.

van Zandt, Timothy (1996a) "Decentralized Information Processing in the Theory of Organizations," in *Contemporary Economic Development Reviewed, Volume 4: The Enterprise and its Environment*, Murat Sertel (Ed.), London: Macmillan Press.

————— (1996b) "Organizations with an Endogenous Number of Information Processing Agents," in *Organizations with Incomplete Information*, Mukul Majumdar (Ed.),Cambridge: Cambridge University Press.

————— (1997), "Real-Time Hierarchical Resource Allocation," mimeo, Department of Economics, Princeton University.

Williamson, Oliver (1967) "Hierarchical Control and Optimal Firm Size," *Journal of Political Economy* 123-138.

————— (1985) *The Economic Institutions of Capitalism* New York, Free Press.