Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective

DILIP MOOKHERJEE*

Separation of ownership from management, multidivisional firm organizations, delegation of production decisions to worker teams, delegation of pricing and advertising decisions to retail franchisers, reliance on intermediaries in trade or finance, and distribution of regulatory authority across different agencies represent examples of organizations that delegate and distribute decision-making authority instead of centralizing it. This paper reviews literature on costs and benefits of delegated decision making in hierarchical organizations or contracting networks with regard to problems of incentives and coordination. It starts by describing incentive and coordination costs of delegation in simple canonical examples of hierarchies where both information and incentives of different decisionmakers differ. One class of models pertain to contexts where the classical Revelation Principle applies, i.e., where costs of contractual complexity, information processing, or communication are absent, agents do not collude, and the mechanism designer can commit to the mechanism. Delegation may conceivably entail a loss of control and coordination arising from the divergence of information and incentives. Sufficient and necessary conditions for this loss to be mitigated entirely include risk neutrality, top-down contracting, and monitoring of transfers or production assignments between subordinates. The next class of models introduces communication costs that restrict the performance of centralized arrangements relative to delegation owing to a resulting loss of flexibility, which has to be traded off against possible control losses of delegation. Finally, consequences of collusion among agents is discussed, which typically enlarge the range of circumstances under which delegation can attain optimal second-best outcomes. The paper concludes with a discussion of the relevance of this theoretical literature to recently emerging empirical studies of industrial organizations where delegated decision making plays an important role: adoption of innovative human resource management practices, new information technologies and retail franchising.

1. Introduction

A classic topic in economic theory is the organization of economic activity. While

the roots of this topic go back to discussions of the appropriate role of the state in a market economy by the Mercantilists and Adam Smith's *Wealth of Nations*, it became the central issue in the 1930s debates concerning the possibility of efficient resource allocation in a socialist economy. In this debate, it was recognized that information relevant to efficient resource allocation is dispersed throughout agents in an economy. A decentralized market economy distributes decision-making authority to individual agents,

who are motivated by their self-interest and are coordinated by market prices. This was contrasted with a socialist economy where decision making authority is vested in a central planner, on the basis of information communicated by individual agents.

Leonid Hurwicz (1972a, 1972b, 1973, 1986) has provided excellent surveys of this early literature and its subsequent offshoots culminating in the contemporary theory of mechanism design. The comparative performance of these specific decentralized and centralized resource allocation systems were sought to be compared on the basis of the extent of communication and information processing they entail, and the incentives of individual agents they give rise to. Arguments pertaining to the superiority of the decentralized market-based system relied on the fact that their communications requirements are minimal (represented by the dimensionality of the message space for communication), the fact that it distributes tasks of information processing across agents in the economy, and is consistent with self-interested behavior of agents. On the other hand, a centralized system may be better equipped to combat weaknesses of the decentralized system in dealing with problems of externalities, public goods, increasing returns, and distributional equity. In environments where these latter problems are significant, hybrid systems combining elements of centralization and decentralization have been studied, e.g., in the literature on economic planning (Edmond Malinvaud 1967, Geoffrey M. Heal 1971).

Versions of the grand question of comparative economic systems continue to be relevant to a variety of modern areas in economics. An example is the internal organization of large firms, to which A. Michael Spence (1975) and Roy Radner (1992) provide excellent introductions. A large corporation is in many ways a miniature economy, in the organization of which the question of centralization versus decentralization is fundamental. Should decisions concerning production, procurement, pricing, distribution,

budgeting, and compensations be centralized in a corporate headquarter office akin to a socialist economy? Or should the corporation be organized into multiple divisions that operate as relatively independent cost, profit, or investment centers, whose managers are delegated substantial authority over production, pricing, procurement, and resource allocation decisions pertaining to their respective divisions? Should human resource management systems be based on delegation of problem-solving on production shop floors to teams of production workers? Comparative systems of organization of business enterprises—e.g., the evolution of the multidivisional structure in U.S. corporations in the early twentieth century, comparisons between American and Japanese business organizations, the adoption of interaction-oriented rather than control-oriented human resource management systems—have been described in these terms (e.g., Alfred D. Chandler 1962; Masahiko Aoki 1990; Casey Ichniowski and Kathryn Shaw 1999; Jon Gant, Ichniowski, and Shaw 2002; Ichniowski, Shaw, and Giovanna Prennushi 1997).

Similar issues arise in the literature on regulation of public utilities (David P. Baron 1985; Baron and David Besanko 1992; Richard J. Gilbert and Michael H. Riordan 1995). For instance, should regulation of all utilities be centered in a single federal regulatory authority or should it be distributed between several state regulatory bodies possessing overlapping jurisdictions with a federal regulator? Alternatively, should procurement of raw materials by a regulated utility company from an upstream firm be delegated by regulators to the utility or should the transaction with the upstream firm also be subject to regulation? Such organizational questions inherently raise similar issues of the optimal degree of decentralization of economic activity.

Modern industrial organization theory also raises questions concerning the organization

of contracting relationships and networks among different market agents. In procurement contracting, should the purchaser contract with a single "prime" contractor, delegating the responsibility of subcontracting with diverse upstream suppliers? Or should she contract directly with downstream and upstream suppliers, and personally coordinate their actions? Should a manufacturer delegate decisions over distribution, advertising, and pricing to wholesalers and retailers? Should investors contract directly with borrowers or delegate authority to financial intermediaries to negotiate loans and supervise the latter? Changes in the nature and extent of reliance on intermediaries appears to be occurring within firms as well as procurement contracting in recent years. Corporate "reengineering" involved elimination of middle layers of management in order to create more direct links between customers, top managers, and production centers (Michael M. Hammer and James Champy 1993, Erik Brynjolfsson and Lorin M. Hitt 2000). "Supply chain management" involves more extensive outsourcing and reduced reliance on intermediaries.

Comparisons between centralized and decentralized systems of economic organization therefore continue to be relevant to many areas of contemporary concern. In this article, I provide an overview of recent theoretical research on this question based on an application of mechanism design theory. I shall focus on comparisons between a centralized resource allocation "revelation" mechanism where a central "principal" makes decisions concerning production and distribution based on communication received from "agents," with mechanisms where decision-making authority is delegated to the latter. Most of the literature has focused attention where the pattern of delegation is hierarchical, i.e., where agents are organized as nodes in an inverted tree, with the "manager" of any divisional branch delegated authority over subordinate branches.

The purpose of this theory is to throw light on the costs and benefits of decentralization from an incentive standpoint and compare performance of different patterns of decentralization. For the most part, the literature I describe has abstracted from costs of communication or information processing. This has largely been driven by the need for tractability: existing theories of mechanism design typically focus on each of these phenomena in isolation. For example, team theory (Radner and Jacob Marschak 1972) or the message space literature (Hurwicz 1960, 1972b; Kenneth Mount and Stanley Reiter 1974; Thomas Marschak and Stefan Reichelstein 1995, 1998) concern themselves only with costs of communication. Theories of costly information processing (Radner 1993; Patrick Bolton and Mathias Dewatripoint 1994; Mount and Reiter 1995; Timothy van Zandt 1996, 1997) have abstracted entirely from incentive considerations. Nevertheless, I shall argue in section 6 below that, in order to have useful applications, the theory needs to address aspects of information technology and incentives simultaneously within a common framework. To this end, I shall include discussion of some models which incorporate communication costs as well as incentive problems.

A central result in the theory is the so-called "Revelation Principle," which relies for its validity on the absence of communication or information processing costs. Under some additional assumptions, this Principle establishes that centralized control cannot be dominated by any delegation arrangement. Specifically, it demonstrates that the outcome of any decentralized organization can be mimicked by a centralized organization in which the responsibility of each agent is merely to communicate their information to a central authority and await instructions on what to do. This argument corresponds to the claims of Oskar Lange and Abba Lerner in the 1930s socialism debate that incentive properties of the decentralized mechanism can be replicated by a suitably designed

centralized mechanism.[1] Conversely, one can view Hayek's counterarguments in the socialism debate that excluding considerations of communication cost or limited information processing capacity of the central planner is like throwing out the baby with the bathwater.

How then does this theory succeed in throwing any light on the centralization versus decentralization question? Its principal virtue is to provide an analysis of the potential costs of decentralization from an incentive standpoint, while abstracting from its principal virtues. These costs arise from the fact that agents that are delegated decision-making authority act in their own self-interest, rather than of the organization as a whole or of the central designer. Accordingly, it gives rise to a problem of "loss of control" or "abuse of power." In contexts where these incentive problems can be overcome by judicious design of the delegation arrangement (i.e., incentives of agents to whom power has been delegated), the presumption is that decentralization will be the superior organizational mode owing to its economy with respect to communication cost and distribution of information processing responsibility. Whereas if the incentive costs are substantial, the choice between centralization and decentralization will involve a trade-off between (incentive) costs and (communication, information processing) benefits.

To further explain the different branches of the theory and the organization of this paper, it helps to review the different assumptions underlying the Revelation Principle[2]:
- Absence of: (1) costs of communication between agents and principal, (2) information processing costs for the principal, and (3) contract complexity costs.
- Absence of collusion among agents.

- Ability of principal to commit upfront to a mechanism and not renegotiate it later.

In light of this Principle, there are two possible ways of constructing a theory of delegation and contractual hierarchies.

The first approach stays within the framework of the Revelation Principle and studies conditions under which delegation is an optimal mechanism, i.e., where a hierarchical mechanism with delegation replicates the performance of the best centralized mechanism. This approach has been pursued by Baron and Besanko (1992), Nahum D. Melumad, Dilip Mookherjee, and Reichelstein (1992, 1995), and Sergei Severinov (1999), among others. The main problem with this approach is that the conditions may be restrictive and, even if they are not, it cannot explain why delegation can be superior to centralization.

The second approach departs from the comfortable premises of the Revelation Principle. Different theories can be classified by the precise avenue of departure from the set of assumptions underlying the Revelation Principle:
  (a) Information processing costs (Radner 1993; Mount and Reiter 1995; Bolton and Dewatripoint 1994; van Zandt 1996, 1997)
  (b) Costly communication, contract complexity (Melumad, Mookherjee, and Reichelstein 1992, 1997; Jean-Jacques Laffont and David Martimort 1998; Ilya Segal 2001)
  (c) Collusion among agents (Sandeep Baliga and Tomas Sjostrom 1998; Laffont and Martimort 1998; Antoine Faure-Grimaud, Laffont and Martimort 2003; Gorkem Celik forthcoming; Mookherjee and Masatoshi Tsumagari 2004)
  (d) Incomplete Commitment and Renegotiation (Paul Beaudry and Michel Poitevin 1995; Poitevin 1995; Wouter Dessein 2002; Faure-Grimaud and Martimort 2001; Baliga and Sjostrom 2001), Incomplete Contracts (Philippe Aghion and Jean Tirole 1997; Andreas Roider 2003).

---

[1] For the most general statement of this Principle, see Roger B. Myerson (1982). For a wider discussion, see Oliver Williamson (1985).

[2] In what follows we refer to the "principal" as the de facto designer and residual claimant, and "agents" as all others with a production or information generation role.

In this paper, I will present an overview of the first approach, and avenues (b) and (c) of the second approach. The emphasis will be on papers that deal with the costs and benefits of delegation relative to centralization, in which incentive considerations play an active role. I thus exclude discussion of avenue (a) since it has so far abstracted from incentive considerations. I do not discuss the important avenue (d) based on incomplete commitment mainly because an excellent survey of this is already available in Poitevin (2000).

I also exclude the literature based on incomplete contracts which involves ad hoc restrictions on the set of allowable mechanisms.[3] These restrictions have been sought to be rationalized by some authors owing to the possibility of contract renegotiation. Segal (1999), for instance, explores conditions under which "simple" contracts which do not incorporate any contingencies achieve the same performance as optimal contingent contracts in a bilateral trading problem, when contracts can be renegotiated. The literature I discuss in the first part of this survey (as well as the second part dealing with collusion) poses the analogous question of when "simple" but commonly observed schemes such as delegated decision-making are optimal.

Clearly, there are some similarities between the literature surveyed here and the literature exploring the foundations of incomplete contracts. Both search for additional constraints on contracts (apart from standard noncooperative incentive compatibility constraints) that rationalize "simple" real-world mechanisms. Both are concerned about the allocation of decision rights, though they use somewhat different terminologies: decentralization rather than property rights. Nevertheless, the two literatures differ in some important details. The literature surveyed here concerns a static context

with a single principal and many production agents privately informed about their respective costs. Mechanisms may be constrained with regard to extent of communication among agents, or by the presence of collusive behavior. The literature on foundations of incomplete contracts deals instead with bilateral trades between two agents that make ex ante relation-specific investments under conditions of symmetric information that cannot be verified by third parties. Contracts may be subject to the possibility of ex post renegotiation.[4]

Returning to the theme of this essay, some additional caveats are in order. The literature I shall discuss does not seem relevant to the question of boundaries of the firm, where the incomplete contract framework seems more fruitful.[5] It therefore applies to delegation arrangements that may or may not be intrafirm and has little to say on whether and why firms should outsource some production rather than in-house. Agents in this framework cannot be distinguished by whether they happen to be within or outside a firm.

Moreover, my intention is to provide a selective guide to the literature that I am personally familiar with, rather than an exhaustive survey. So important papers may have been missed or glossed over. I also want to keep the overview brief and readable, so will eschew formalism of any kind and try to communicate the essence of key results and ideas. Those interested in more detailed statements of model assumptions and results should consult the original sources.

Section 2 sets out the canonical setting of production with multiple agents privately informed about their own costs, employed in most of the literature. Section 3 describes models belonging to the "first" approach

---

[3] See Tirole (1999) and the symposium on incomplete contracts in the *Review of Economic Studies* 1999 for a critical review of this literature.

[4] Of course, these comparisons pertain to polar types of models studied in the two literatures. Baliga and Sjostrom (2001) is an example of a principal–multiagent model where decentralization turns out to be optimal owing to the possibility of contract renegotiation.

[5] For an overview, see Oliver Hart (1995).

described above, which explore conditions for optimality of delegation within the traditional confines of the Revelation Principle. Section 4 then discusses avenue (b) within the second approach, based on costs of communication or contract complexity, while section 5 is devoted to implications of allowing agents to collude. Section 6 assesses the potential relevance of this literature to empirical industrial organization research. Finally, section 7 concludes with a summary of the principal insights and shortcomings of this literature, and suggests directions for future research.

2. The Canonical Setting of Joint Production with Adverse Selection

Most of the literature focuses on models where agents have private information about their costs, rather than moral hazard, and we will do the same. The typical model is usually of the following kind. There is a single principal (P), one or two producing suppliers (A_1, A_2), plus one manager/middleman/monitor M who plays no role in production. The gross benefit of P depends on joint output whose production depends on contributions or supplies of the two agents, described by the production function $q = f(q_1, q_2)$, where A_i produces q_i at private cost $C_i(q_i, \theta_i)$ satisfying a set of single-crossing conditions.[6] The parameter θ_i affects the productivity or cost of the agent and A_i is privately informed about realization of θ_i. The cost shocks θ_1, θ_2 are independently distributed, and satisfy a monotone hazard rate condition that allows one to ignore nonlocal incentive constraints in the classical setting. The manager or monitor M observes signals η_1, η_2 which are informative about θ_1, θ_2 respectively. P, A_1, A_2 are risk neutral, while M is risk neutral or risk averse. Outside

options for all agents are normalized to 0. Agents observe their costs prior to contracting and so earn information rents. The main trade-off in designing contracts or delegation is between productive efficiency and paying information rents to agents.

Many papers employ considerably restricted versions of this model, e.g., where there is one productive agent rather than two, or particular production functions— e.g., perfect complementarity between q_1, q_2, indivisibility of inputs or outputs, linear cost functions, or two point distributions for cost shocks and signals.

The role of M in the organization is informational "expertise" relative to the principal, acquired either through prior expertise or from ability to monitor cost realizations of the agents. The process by which M acquires this information is treated as exogenous; the theory explores the implications of such informational specialization. This contrasts with an earlier literature on contracting hierarchies (e.g., based on Guillermo A. Calvo and Stanislaw Wellisz 1978) focused on moral hazard in supervision, but not on delegation questions.[7] Some of the interesting questions addressed by the theory concern the relative desirability of delegating to M rather than the productive agents themselves. This concerns the widespread phenomenon of "management" as a specialized activity, distinct from "production" (as argued persuasively by Radner 1992). Accordingly, the organizational alternatives usually compared include the following (see figure 1):

---

[6] The key condition is that marginal production costs are increasing in θ_i, but these are usually supplemented by conditions that total costs and the rate of increase of marginal costs are also increasing. For the most general statement, see Melumad, Mookherjee, and Reichelstein (1995).

[7] In that literature, supervisors are not delegated any authority to contract with those they supervise. The focus was on whether moral hazard in supervision creates organizational diseconomies of scale. Indeed, Calvo and Wellisz and others found that explaining scale diseconomies through this channel was typically not as straightforward has had been originally imagined, e.g., by Williamson (1967). The literature on delegation described here in contrast does provide conditions for emergence of "control loss" in hierarchies under suitable conditions, so can be thought of as an alternative formalization of this phenomenon.
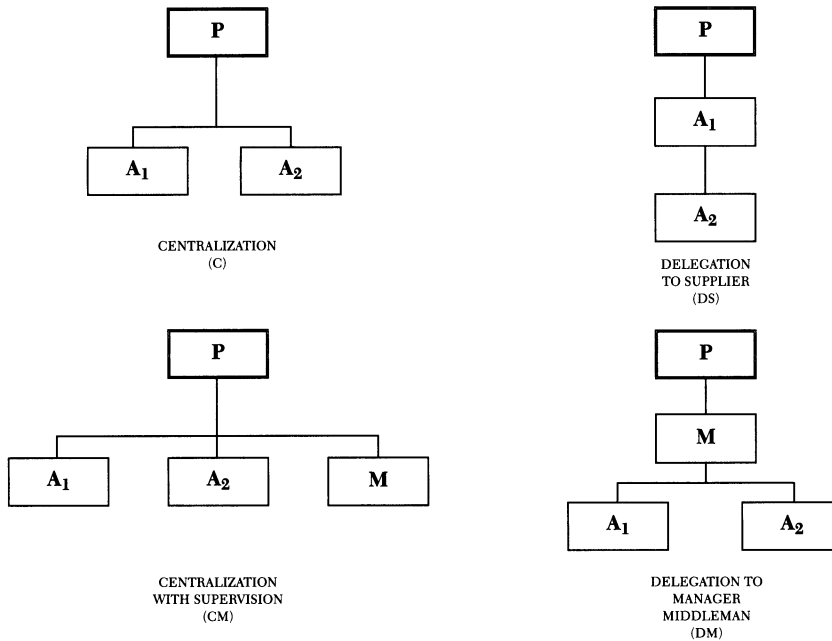
Figure 1. Organizational Alternatives

Centralization without Supervision (C): P retains all control, contracts and communicates with A_1, A_2

Centralization with Supervision (CM): P retains all control, contracts and communicates with A_1, A_2 and M

Delegation to Supplier (DS): P contracts only with A_1, delegates authority over contracting with A_2

Delegation to Manager/Middleman (DM): P contracts only with M, delegates authority to M over contracting with A_1, A_2.

Many papers pose the question of delegation versus centralization as DS versus C; others pose it as DM versus CM; yet others as DM versus C. Accordingly to understand the relation between different papers, we shall make this explicit in discussing their results using this terminology. It is also useful to clarify that we use the term "delegation" to connote a setting of hierarchical contracting with multiple vertical layers, where contracting and communication are restricted only between adjacent layers. For instance, in DS,

P does not contract or communicate with the subcontractor A_2 at all.

With only two productive agents, most of the literature compares the centralized two layer hierarchy with a three layer one where the agent at the intermediate layer is delegated authority over contracting with those at the third layer. The main focus is thus on problems of vertical control. Problems of horizontal coordination across different branches of a hierarchy inherent in delegation have therefore received little attention (with few exceptions, described in section 3.4). Issues concerning design of more complex hierarchies (e.g., span of control versus number of vertical layers, trading off horizontal coordination problems and vertical control loss), or comparison with nonhierarchical forms of delegation, must await future research.

Additional details concerning delegation concern exactly what is observed by the principal and the sequence of contracting in the hierarchy. Insofar as they help determine the

nature of control loss from delegation, these will be described in subsequent sections. Collusion between agents is modeled as hidden side-contracts between the agents, and further details of observability and timing are involved here, which will be explained in section 5.

3. First Approach: Conditions for Optimality of Delegation within the Traditional Framework

3.1 Conditions for DS = C

Consider first the optimality of delegating to a "prime" supplier the authority to subcontract with the other supplier, i.e., comparison of DS with C. When the Revelation Principle applies, we know that C weakly dominates DS, so the question is when DS achieves the outcome under the optimal centralized mechanism. In this case, delegation entails no control loss at all. Insofar as contracting with one agent rather than two is simpler or easier for the principal, this can provide a rationalization of delegation as a way of implementing the optimal outcome.

The potential control problem with delegation involves a form of "moral hazard": the "prime" supplier's incentive to allocate production and payments between himself and the subcontractor may differ from what the principal desires. In particular, the prime contractor has monopsony power over the subcontractors, and the principal in turn has monopsony power over the prime contractor. Both sets of relationships are subject to adverse selection, so monopsony will generate distortions in production and payments that raise contracting costs for the principal. The vertical control problem inherent in delegation therefore is essentially that of "double marginalization of rents" (DMR), stressed in the context of sequences of monopoly relationships in the industrial organization literature.

The monopsony power of the prime contractor (A_1, say) over the subcontractor (A_2) results in two problems. First, there tends to be "too little outsourcing" from subcontractors (or too low a supply price offered by A_1). This results from the attempt by A_1 to garner maximal informational rents for himself at the expense of A_2 and P. Roughly speaking, the volume of information rents are proportional to the level of production allocated to an agent, so A_1 allocates himself a larger share of production than either A_2 or P would desire. Second, A_1 is privately informed about subcontracting cost (what he has to pay A_2) vis-à-vis the Principal, owing to the lack of direct communication between P and A_2. This adds a dimension of adverse selection in A_1's relationship with P, over and above privacy of information about his own cost of production. There is a consequent "cascading" of information rents: first such rents are paid to A_2 by A_1, and then this is subject to an additional markup owing to privacy of A_1's information vis-à-vis P regarding subcontract costs. This is as far as the principal's welfare is concerned; from the standpoint of social efficiency only the production misallocation matters.

An important result in the literature is that these control problems can be eliminated in DS under the following conditions:

(3.1) **Observability of subcontract costs or allocation**: either q_1 the amount produced by the prime supplier, or x_2 the subcontracting cost incurred.

(3.2) **Top-Down Contracting**: P contracts with A_1 *before* A_1 communicates or contracts with A_2.

(3.3) **Risk-neutrality, absence of limited liability constraints**: for A_1

The main idea underlying this result is that condition (3.1) enables the principal to subsidize outsourcing or tax in-house production by A_1 to correct the productive misallocation between the two agents. Moreover (3.2) and (3.3) allow P to "tax" away upfront information rents deriving from privacy of the prime contractor's knowledge of subcontract cost.

Each of these three conditions are necessary for delegation to be optimal, barring

exceptional cases. The production misallocation can arise only in the presence of some substitutability between the inputs supplied by the two agents. Otherwise if there is perfect complementarity, there is no scope for a monopsony distortion between A_1 and A_2, and condition (3.1) is unnecessary (Baron and Besanko 1992; Gilbert and Riordan 1995). Otherwise, conditions (3.1) and (3.2) are necessary (given risk neutrality (3.3)) if the production function is smooth (Melumad, Mookherjee, and Reichelstein (MMR, hereafter) 1995). The necessity of risk-neutrality (3.3) is also suggested by MMR 1995, and reinforced by the literature to be described next.

3.2 Comparison of DM with C

Another way of posing the question of optimality of delegation is to compare DM (rather than DS) with C. R. Preston McAfee and John McMillan (1995) consider the case with only one productive agent, and a single monitor M who has no better information than P about the agent's cost. They focus on the costs of intermediation arising from delegation of procurement to M who is subject to limited liability. The main idea here is that M can earn information rents with respect to privacy of knowledge of subcontract costs, since there is no direct communication between the supplier and the principal. These rents cannot be taxed away upfront owing to limited liability constraints. This is very similar to the implications of violation of (3.2)—wherein A_1 can contract or communicate with A_2 before responding to the contract offered by P (MMR 1995).

Faure-Grimaud and Martimort (2001) replace limited liability by risk aversion of M and obtain an agency cost of delegation (with a single productive agent) that is qualitatively similar (though different in detail). Inability of M to costlessly bear risk prevents congruence of interests of M with P at the time M designs subcontracts for productive agents. This is somewhat akin to a classical moral hazard problem that trades off risk sharing and incentives with a single agent.

3.3 Effects of Supplier Consolidation: Comparing DM with C with Two Productive Suppliers

Now consider the related question of how DM relates to C in the case of two productive agents A_1, A_2. Assume M is perfectly informed ex ante about realization of the agents' costs: so there is no upfront uncertainty faced by M. Delegation to M implies that M rather than A_1, A_2 earn informational rents, owing to informational asymmetry between M and P concerning subcontract costs. The principal now contracts with a single supplier of *both* inputs who incurs supply cost equal to the sum of the two agents' cost. The comparison of C with DM is thus really a question of effects of consolidating the two suppliers into a single supplier.

The main result here is that *such consolidation benefits P if the inputs are perfectly complementary in production under some added distributional conditions* (Baron and Besanko 1992, Gilbert and Riordan 1995), *but hurts P if they are substitutes* (Severinov 1999; Mookherjee and Tsumagari 2004).[8] The intuitive idea underlying this result is the following. In C there are externalities between the two separate suppliers, which are internalized with consolidation. If they supply substitutes, then competition is suppressed and cost reports increase (on average) with consolidation. Consolidation also converts a pair of one-dimensional incentive problems into a single two-dimensional one (both in terms of dimensions of information and number of goods delivered). This tends to strengthen incentive constraints (owing to the ability of the single consolidated agent to coordinate reports). On both counts the

---

[8] The definition of substitutes and complements is in terms of the way that the demand for input delivered by one supplier varies with the cost reported by the other supplier. Mookherjee and Tsumagari show how this relates to notions of substitutes and complements in terms of the elasticity of substitution of the production function.

principal is worse off when the two inputs are substitutes. If they supply complements instead, the internalization of bidding externalities causes cost reports to fall on average with consolidation. But opposing this is the problem that the adverse selection problem is a higher dimensional one with consolidation. Under particular distributional conditions the higher dimensionality poses no additional problems, and then P attains superior expected profit from DM.[9]

That delegation scores above centralization here may occasion some surprise, since the Revelation Principle still applies. However the set of agents is not the same between DM and C: in the former there is an additional agent M who is better informed than P. The superiority of DM thus flows from access to this additional source of information unavailable in C by assumption. Indeed, the relevant centralized benchmark for DM is CM rather than C, where M belongs to the organization and P retains all control and contracts with M, A_1 and A_2. In such a setting M is relegated to the status of a supervisor or consultant, rather than being a manager with control rights over the suppliers. The Revelation Principle asserts that CM always weakly dominates DM. Moreover, in CM the principal can costlessly acquire all of M's information, given the absence of any collusion between M and the productive agents supervised. Whereas in DM the principal has to encounter the problem of controlling M. Hence in general P would be strictly better off retaining control (CM) rather than delegating (DM).[10] This implies that explanations of the widespread phenomenon of delegation to intermediaries uninvolved in actual production must perforce depart

from the traditional setting of the Revelation Principle. We shall return to this point below.

3.4 *More Complex Hierarchies: Organizational Diseconomies of Scale; Horizontal Coordination*

Now suppose there are more than two productive agents. Then there is a choice between different hierarchical patterns of delegation. Abstracting from the possibility of employing supervisors or intermediaries, the n producing agents could be organized in a linear vertical chain, with agent A_i authorized to contract with A_{i+1}. Or they could be organized into two horizontal departments, each of which contains a subset of producers, and involves a vertical linear chain within the department. Or we could have a hierarchy where each agent is authorized to contract with m subordinates (with k vertical layers). See figure 2 for examples. And so on: there are a large number of possible hierarchical structures, with varying spans of control (horizontal branches) and vertical layers. The questions that can be posed in this framework involve both questions of the design with a given number of agents (e.g., span of control versus vertical layers, how to group agents into departments, how to organize communication and contracting systems), as well as how the performance of the organization changes as the number of agents increases. The latter question is of interest to understanding limits to the size of firms or networks that arise solely from incentive and coordination problems.

Delegation is now potentially prone to problems both of vertical control loss (cascading across vertical layers) and coordination across horizontal branches. The latter problem is not addressed by any of the models described so far, though it has received some attention in the team-theoretic literature which has abstracted from incentive considerations (e.g., Marschak and Reichelstein 1995, 1998). Jacques Cremer and Riordan (1987) represents one

---

[9] The distributional condition is however quite restrictive: i.i.d. exponential cost shocks with a lower bound of 0. Extensions to a wider class of distributions remains an open question.

[10] This is easiest to see when M has perfect information about costs—then CM attains the first-best, which DM cannot.
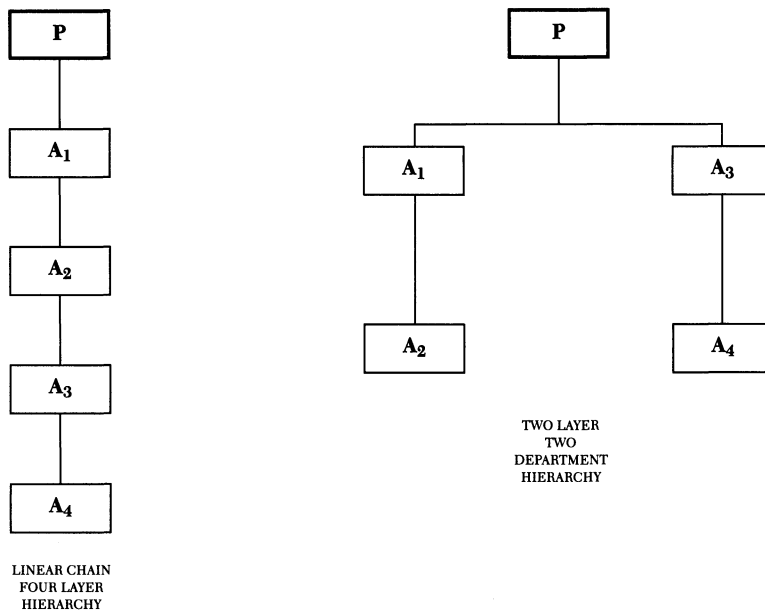
Figure 2. Hierarchies with More Than Two Agents

of the first attempts to deal with complex hierarchies (and also nonhierarchical contract networks), wherein conditions for implementability of first-best efficient allocations with ex ante contracting with risk-neutral agents were established.[11] In particular they describe a sophisticated system of contracting and communication that enables such a contracting network to resolve both incentive compatibility and coordination to achieve ex post efficient outcomes.

Extensions of the theory to implementation of optimal mechanisms for a principal (rather than ex post efficient outcomes) in the presence of adverse selection are considered in Mookherjee and Reichelstein (1997, 2001). They confine attention to contracting hierarchies that are "consistent" with the technology in the sense that one dimensional reporting would help the same hierarchy

achieve first-best outcomes in the absence of any incentive problems. In other words, the production technology has a constant returns recursive structure, wherein production is hierarchically decomposed into departments and subdepartments. In the absence of incentive problems, coordination across different departments can be achieved by a hierarchical communication system, where each manager receives (one-dimensional) cost reports from subordinates, aggregates them into a departmental cost, and reports this to his superior in turn. Cost reports flow up the hierarchy in this way, up to the very top. There aggregate output decisions are made, which are subsequently disaggregated into departmental targets, with these flowing subsequently down the hierarchy.

This mechanism can be extended to incorporate incentive problems as well, under exactly the same conditions (3.1)–(3.3) that enables DS to achieve optimal outcomes in the two agent case. Specifically, Mookherjee and Reichelstein show that *if assumptions (3.1)–(3.3) hold, any hierarchy "consistent"*

[11] Ex ante contracting refers to the situation where contracts are negotiated before agents receive any private information, so adverse selection problems do not arise.

with the technology (in the sense defined above) can achieve the same expected profit and production allocation as the optimal centralized mechanism. This implies that conditions (3.1)–(3.3) as well as "consistency" are sufficient (as well as necessary) for hierarchical delegation to implement optimal centralized outcomes. Once vertical control problems can be overcome with one level of delegation, they can be overcome with multiple vertical layers. Moreover horizontal coordination across departments can also be costlessly assured by the mechanism.

The mechanism involves three stages: first contracts flow down the hierarchy, then cost reports flow up and get aggregated along the way, determining production targets which flow down at the third stage. Specifically:

- at each layer an agent manages a "profit center"
- managerial incentive schemes are linear in a measure of profit of the center
- departmental costs equal aggregate payments authorized to subordinate departments
- departmental output is valued at willingness to pay (à la Groves-Vickrey) of the manager at next higher level
- managers self-select profit targets for themselves at the first (contracting) stage, then report cost (bid on projects) to their bosses at the second stage after receiving cost reports from their subordinates

Some of these features resemble performance budgeting systems described in management accounting textbooks (e.g., Charles T. Horngren and George Foster 1991). The mechanism distributes information processing tasks throughout the hierarchy, as managers aggregate cost reports for their respective departments and allocate production between subordinates and themselves. With incentive problems overcome, design of the hierarchy can be based on considerations of efficient distribution of information processing à la theories of Radner (1993) or van Zandt (1996, 1997). In particular, these considerations explain why the two extreme organizational

forms (a "flat" two layer centralized mechanism and a "steep" linear chain) would be dominated by intermediate forms with a number of vertical layers and horizontal branches. Note finally that the result provides a set of benchmark conditions under which there are no organizational diseconomies of scale.

3.5 *Shortcomings of the First ("Traditional") Approach*

As we have seen above, the conditions for optimality of delegation are restrictive. For instance, absence of risk aversion or limited liability constraints limits applicability to managerial hierarchies. Implementability of second-best allocations via delegation may be more relevant in procurement settings where the prime contractor is a large firm with deep pockets. On the flip side, the theory explains information rents of managers that cascade across vertical layers. If intermediate managers are more risk-averse than the owners, this makes it difficult to ensure coincidence of their objectives. These results suggest that costs will grow with the scale of the firm's operations, causing organizational diseconomies of scale, though explicit models of this phenomenon are still awaited.

A more serious problem with this approach is that it cannot ever explain why delegation may dominate centralization. This is simply a consequence of staying within the confines of the Revelation Principle.

Moreover, the constructed delegation mechanisms (e.g., under assumptions (3.1)–(3.2) of observability and top-down contracting) are vulnerable to collusion between agents. This is because the principal has to subsidize outsourcing from subcontractors, creating incentives for artificially exaggerating outsourced cost by the prime contractor.[12] The prime contractor also would have an incentive to contract and communicate with the subcontractor *before* responding to the principal's offer, which may be difficult for the principal to prevent.

Yet another problem is that this approach cannot explain delegation to middlemen or managers that play no direct productive role. As explained above, it is generally strictly better for P to retain control and treat M as an information provider. This stems partly from the assumption of noncollusive behavior between supervisor and productive agents.

This motivates interest in literature which confronts these problems, by departing from the confines of the Revelation Principle.

4. Benefits of Delegation: Costly Communication and Contract Complexity

Reconsider the choice of DS versus C, where there are two productive agents. The question is whether the principal should delegate contracting with A_2 to A_1. A commonly alleged advantage of decentralization is that it utilizes the benefits of local information. This presumes that centralization is unable to take advantage of such local information. In the framework where the Revelation Principle applies, however, communication is costless and centralized decision making can access the same information as any decentralized mode. Formalization of the "local information" advantage of decentralization necessitates incorporation of constraints on what agents can report to the principal, or alternatively on the extent of information that can be processed by the principal.

Models of costly information processing in the mould of Radner (1993), Mount and Reiter (1995), or van Zandt (1996, 1997) are still in their infancy and await extensions that incorporate incentive considerations. They model computations taking place in real time, and model advantages of decentralization in distributing information processing tasks throughout the organization, rather than concentrating it in one central

authority. While this is undoubtedly an appealing idea, it requires modeling limits on information processing of the principal and agents. Modeling incentive compatibility constraints in such a context is difficult, since it requires modeling how agents divide their limited information processing capacity between attempting to game the principal for their own self-interest, and carrying out the computations that they are appointed to perform.[13]

In order to avoid this foundational problem, an alternative approach is to introduce restrictions on communication between agents, rather than constrain their information processing capacities. Mechanisms are constructed under the assumption of "unbounded rationality," but subject to these communication restrictions. MMR (1992) compare optimal centralized and decentralized mechanisms with an exogenous restriction on the size of the message space that can be used by any agent to communicate with others. In particular, if the upper bound on the size of the message space is not large enough to permit agents to communicate everything that they know, centralized decision-making cannot access all the information that delegation mechanisms utilize. For instance, the mechanism DS allows production allocations between the two agents to be decided by A_1 on the basis of his information about his personal cost θ_1, which he can only partially communicate to the principal. The restrictions on the size of the message space, however, are not explicitly modeled, and would likely require some underlying model of "expertise" of suppliers not shared by P which restrict the vocabulary of communication between them, or the need to make decisions in real time where messages take time to be communicated.

The principal result in MMR (1992) is that: *DS strictly dominates C if (a) there is a*

---

[12] The prime contractor could claim to have outsourced more than the true amount, in return for a side payment from the subcontractor.

[13] Indeed, this decision itself is a higher order problem, quickly giving rise to an infinite regress of the form familiar in any model with "decision costs."

finite message space for each agent, whereas costs lie on a continuum; (b) assumptions (3.1)–(3.3) of cost observability, top-down contracting and risk-neutrality hold; and (c) the cost function of agent A_i is multiplicatively separable between quantity produced q_i and the cost shock θ_i. Note that this result applies irrespective of the size or nature of the message spaces, as long as they apply uniformly to both organizational regimes. Hence the comparison does not require any explicit model of communication costs. Note also this result pertains to DS and C with particular reporting structures (e.g., in C where agents either communicate simultaneously or sequentially in two stages). So this does not establish the global optimality of delegation among the class of all possible mechanisms subject to the same message space restrictions. Part of the problem is that there is no general characterization so far available (analogous to the Revelation Principle) of the set of all possible feasible mechanisms in the presence of communication constraints.[14]

Nevertheless, the result captures the idea that delegated decision making may be better able than most common versions of centralization to utilize "local" information. Production assignments are selected by A_1 in DS on the basis of his information about own cost θ_1, which is finer than can be communicated to P. In general, this "flexibility" could be abused by A_1 to pursue his own interest at the expense of the principal: the restriction on communication also limits the ability of P to calibrate the outsourcing subsidy in DS precisely enough to ameliorate the problem of double marginalization of rents. Nevertheless MMR show that the added control loss is always outweighed by the advantage of better informed production decisions, provided the principal can monitor

production assignments or payments between the agents ex post.

An alternative restriction which yields the same result concerns the number of contingencies in contracts. A contingency can be viewed as a statement of production targets and payments made to an agent, conditional on reports communicated by the latter. One measure of the complexity of a contract is the number of contingencies, since these have to be written (often with the aid of lawyers) and read by third party enforcers (e.g., courts) at some cost. Accordingly contracts with more contingencies involve higher costs of writing and verification.

Of course in some situations this is not a good measure of complexity, e.g., when it is possible to express the contract in the form of a simple mathematical formula (the simplicity of which may be contrasted with the number of elements in the graph of the function, the complexity measure being considered here). A deeper analysis of complexity should be based on measures of complexity or computability of the mathematical formula that represents the contract.

The performance of DS and C can be compared with identical restrictions on the number of contingencies in contracts in either regime. Restricting contracts to finite complexity imposes limits not just on the size of message spaces of agents, but also of decisions (concerning production and payments) that can be taken on the basis of such reports. This further limits the flexibility advantage of DS over C, compared with the case where message space size is restricted. Despite this, it can be shown that *the superiority of DS over C still prevails under the same conditions as where message space sizes are restricted* (MMR 1997). Production allocations can no longer be varied continuously by A_1; the contract can only specify a finite set of alternative production target configurations. Yet there is scope in DS for A_1 to choose flexibly among these different configurations based on his "local" information which cannot be accessed by a centralized mechanism.

---

[14] In particular, the result described above does not exclude the possibility that a centralized mechanism with a more general multistage communication mechanism may be able to match the performance of DS. This is an open question for future research.

On the other hand, if one or more of assumptions (3.1)–(3.3) do not hold then examples can be constructed whereby the ranking gets reversed, i.e., the exacerbation of the control loss outweighs the flexibility advantage, rendering C superior. Accordingly this theory succeeds in providing conditions when either of the systems (strictly) dominates the other. However there still remains the need to provide a deeper foundation of the notions of contractual complexity or communication costs employed in this branch of the literature.

5. *Collusion Among Agents*

We now describe a more recent and active strand of the literature evaluating delegation when agents collude, while all other traditional assumptions underlying the Revelation Principle are retained. Collusion affects the performance of both delegation and centralization. For instance we have already seen that the mechanisms constructed for DS that replicate outcomes of the best centralized mechanism under conditions (3.1)–(3.3) are vulnerable to collusion among the agents. Optimal centralized mechanisms are also frequently vulnerable to collusion.[15] The question is how the two regimes compare in the presence of collusion.

One possible intuition for the virtues of delegation is that it already incorporates side-contracting among the agents, unlike centralization. However we shall see below that this intuition is difficult to make precise: when one models hidden side contracts in a particular way, a version of the Revelation Principle reappears, whereby centralization can replicate the outcomes of any delegation arrangement. Nevertheless, one can then explore the idea that delegation and centralization are equivalent under a broader set of circumstances when agents collude. In other words, delegation may be an optimal response to the presence of collusion among agents—if side contracting cannot be prevented, the principal may as well authorize it explicitly.

An additional reason to study collusion was mentioned previously: many instances of delegation take the form of DM, where authority is delegated to expert intermediaries or managers who play no role in actual production. In the absence of collusion P can costlessly obtain M's information and use this to design contracts for suppliers personally. However, the agents then have a stake in bribing M to withhold his information. Collusion would therefore prevent costless acquisition of M's information under centralization. Delegation to M would not permit costless extraction of M's information either, since it would form a source of information rents for M in contracting with the principal. M will earn rents under either regime in the presence of collusion, so the superiority of centralization is no longer obvious.

5.1 *Modeling Collusion*

Collusion is typically modeled as a side contract between agents which is unobserved by P, and subject to asymmetric information within the coalition. Nevertheless many important details need to be specified concerning the nature of the side-contracting game. How are these contracts negotiated, how does their timing relate to the proposal and acceptance stages of the contract offered by the principal? What is the range of decisions the agents can coordinate on, and what does the principal observe? How is the side contract enforced?

The complexity of analysis of coalitional behavior under private information has motivated most authors to adopt simplifying assumptions that permit a tractable analysis. Most of the literature follows the trend set by Tirole (1986) by ignoring all enforceability constraints, by assuming long-term contracts or self-enforcing relationships among agents. Only restrictions imposed by asymmetric

---

[15] Consider for instance a second price auction in which bidders can coordinate their bids and enter into hidden side-contracts that reallocate the good among themselves. There often exist such side contracts that are interim Pareto superior from the standpoint of the bidders.

information within the coalition (besides limited liability, if applicable) are imposed. At the same time the contractual relationship between agents and principal is short term and restricted to that defined by the formal contract.

It is also common to assume that side contract allows agents to coordinate reports to P, and to reallocate production assignments and payments between themselves. This corresponds to auction contexts where coordinated bid-rigging and hidden "resale" of the good cannot be prevented by the seller. Since collusion occurs with asymmetric information, the Coase Theorem does not apply. Hence actual outcomes (e.g., production assignments) will depend on the allocation of bargaining power within the coalition. In turn this depends on who has the power to propose the side contract.

Laffont and Martimort (1998) assume that the organizational variant affects allocation of bargaining power: in DS A_1 makes a take-it-or-leave-it side contract offer to A_2, while in C a neutral third party designs the side contract for A_1 and A_2. This may seem natural in some ways: if the principal delegates authority for contracting with one of the agents to the other, it augments the bargaining power of the latter. Yet in the Laffont and Martimort (1998) formulation this is built in as an exogenous shift in bargaining power. An alternative would be to explain this shift endogenously. For instance, the reason that A_2 has more bargaining power vis-à-vis A_1 under centralization is that A_2 has the option of turning down the contract offered by A_1 and then playing the contract offered by the principal noncooperatively. Such an outside option is not available to A_2 in DS—thus conferring monopsony power to A_1 in the delegation arrangement.

The more recent literature models the endogenous effect of organizational structure on bargaining power within the coalition. Specifically, it is commonly assumed that one of the agents (denoted A^*, say) has the power to make a take-it-or-leave-it

side-contract to the others. If the principal delegates subcontracting to A^*, then the other agents do not have the option of rejecting the subcontract and dealing with the principal on their own. Under centralization in contrast, they have the opportunity of rejecting the subcontract and playing the principal's game noncooperatively. This raises their outside option vis-à-vis A^*, effectively gaining bargaining power. The extent to which they do so depends endogenously on the contract offered by the principal. The latter simultaneously defines the stakes of collusion for the agents, as well as their relative bargaining power.

A consequence of this formulation is that the difference between delegation and centralization boils down to this. If P delegates to A^*, then all other agents have no alternative to the subcontract offered by A^*. Under centralization, P offers a contract to all the agents, who now have the option of rejecting the side contract offered by A^* and responding to P's offer noncooperatively. Then centralization reduces to delegation if P offers a null contract to all agents but A^*. Therefore *a version of the Revelation Principle reappears: if the structure of the side-contracting game is the same in different organizational variants (in the sense of who has the right to design the side contract), then centralization is always weakly preferred to delegation.* Such a formulation cannot therefore permit delegation to dominate centralization. We are back to a question similar to that in the traditional first approach described above: is delegation costly relative to centralization?

5.2 Costs of Delegation with Collusion

In order to emphasize that centralization regime now incorporates collusion, we shall refer to the two variants of centralization (with and without M) by CMC and C respectively. As before, we can phrase the key question either as comparison of C and DS (should P delegate to one of two suppliers?), or of CMC and DM (should P delegate to a monitor/manager M?).

5.2.1 *C versus DS*

This question is posed by Baliga and Sjostrom (BS, hereafter) (1998) in a model with moral hazard and limited liability. Two suppliers A_1, A_2 jointly produce an indivisible output for P. The probability of "success" is increasing in the effort of each agent, which takes one of two possible values ("shirk" and "work"). Production occurs sequentially, and the effort of one agent (A_1) is observed by the other (A_2) but not vice versa. The principal cannot observe efforts of either agent, and is constrained to pay them nonnegative wages (with zero as their respective outside options). The combination of moral hazard with limited liability gives rise to "efficiency wage" information rents, and renders the Coase Theorem inapplicable. The stakes for collusion arise here whenever P tries to elicit A_2's information about A_1's effort: A_1 can bribe A_2 to withhold adverse information. BS do not model the actual process by which the side contract is negotiated, using instead a notion of an equilibrium side-contract as a Pareto-undominated side contract, given the principal's mechanism.

The main results of their paper are as follows: (i) In C it is always optimal to use simple contracts with no communication, where P pays A_i a wage w_i in event of success and 0 otherwise; (ii) For a large range of parameter values, delegation is optimal (i.e., the best simple contract has either w_1 or $w_2 = 0$); and (iii) There exist other cases for which delegation is not optimal (e.g., if neither agent A_i is "essential" enough that delegating to the other agent A_j would induce the latter to pay former a positive rent in event of success). The model thus captures the idea that in a large range of cases decentralization is an optimal response to collusion—the principal can implement the optimal centralized outcome under collusion by contracting with only one of the two agents and leaving that agent to subcontract with the other. The payment to the delegated agent is such that it motivates that agent to pay the other agent

exactly the same efficiency wage as the principal desires, and leave the same amount for the delegated agent as well. However (iii) shows that such an implementation is not globally possible.

A similar comparison is carried out in an adverse selection framework of the kind described in previous sections by Mookherjee and Tsumagari (MT, hereafter) (2004). They find in contrast to the BS paper that delegation to a supplier is *always* strictly dominated by centralization. The essential reason is that the shift of bargaining power to A_2 under centralization allows the productive distortion inherent in DS to be reduced. Recall that this distortion took the form of "insufficient outsourcing" from A_2. The need for A_1 to offer higher rents to A_2 in the centralized setting (where P offers a contract more favorable to A_2 than emerges in the solution to DS) forces A_1 to subcontract more to A_2, thus reducing the extent of the productive misallocation. In a sense, this is qualitatively similar to the cases studied by BS in which decentralization was not optimal: there centralization offers A_2 a positive efficiency wage which raises his effort, reducing the distortion arising in delegation when the delegated agent A_1 prefers to keep all the rents for himself and so does not pay any efficiency wage to A_2. The difference in results between the two papers could therefore reflect the fact that outputs and inputs are assumed to be divisible in the MT paper, enlarging the scope of distortions arising from A_1's monopsony power in delegation. Alternatively there are important distinctions between the moral hazard and adverse selection settings considered that drive their respective results.

5.2.2 *DM versus CMC*

Faure-Grimaud, Laffont, and Martimort (FLM, hereafter) (2003) and Celik (2002) both consider the case of one productive agent A and one supervisor M. The former is privately informed about his own cost. The two papers adopt different assumptions

about the nature of information available to M about this cost. FLM assume that the cost of the supplier takes two possible values, and M observes an informative signal which also takes two possible values. In contrast, Celik assumes an arbitrary finite number of possible costs, with the information of M represented by a connected partition over the state space. Specifically, M can narrow down the agent's cost to a (connected) subset of possible cost levels, but cannot distinguish between different costs within this subset. Apart from the information structure, FLM allows for risk-aversion of M, whereas Celik assumes risk neutrality, but this difference is less fundamental.

The results of these papers are strikingly different: *FLM find that delegation is always equivalent to centralization, whereas Celik finds that delegation is inferior in general (e.g., whenever the optimal contract for the agent in the absence of M is strictly monotone, and M is not perfectly informed about the agent's cost).* The intuition in Celik's paper seems similar to that in the MT paper: delegation to M is subject to a monopsony distortion, causing M to procure an insufficient amount from A that is not in the principal's interest. In particular, the optimal allocation involves incentive constraints that bind solely in the "downward" direction, where more productive types of the agent are indifferent between underreporting and reporting truthfully. Centralization allows the principal to raise A's outside option in bargaining over the subcontract with M, reducing the extent of this monopsony distortion. In particular, since the agent's reservation utility with respect to the side contract is strictly increasing in his productivity (owing to the agent's option of playing P's mechanism noncooperatively), it creates a set of "countervailing" incentives for the agent to over- rather than underrepresent its productivity. This relaxation of the downward incentive constraints permits a reduction in the extent of underproduction. The pattern of incentive constraints and associated productive distortions in the FLM context

are different, where "upward" incentive constraints bind with delegation. In such a context, creating countervailing incentives by raising the agent's outside option turn out not to be valuable for the principal.

Finally, MT (2004) compare CMC, DM, and C in the case of two productive agents A_1, A_2 and a perfectly informed M. Their motivation is to explore the idea that collusion may rationalize delegation to intermediaries uninvolved in production. This necessitates the presence of more than one productive agent.[16]

With M perfectly informed, there is effectively no asymmetric information within the coalition, so inducing changes in bargaining power within the coalition is not valuable for the principal and $DM = CMC$. Hence delegation is costless relative to the centralized alternative where P retains control and treats M as an information provider.

What is the value of hiring M when there are two productive agents? With M perfectly informed, DM reduces effectively to the case where the principal deals with a single consolidated agent that delivers both inputs at a cost equal to the sum of production costs of A_1 and A_2. Whereas if M is not hired, we have already explained that centralization (C) is the better alternative than delegating to one of the suppliers (DS). So the value of hiring (and delegating to M) is effectively the same question as the effect of consolidating diverse suppliers, except that now we must consider the presence of collusion among the suppliers under centralization. MT show that exactly the same results concerning consolidation effects (discussed in section 3.3 above) hold in the presence of collusion as well. Specifically, DM < C if the two agents produce substitutes, while DM > C if they produce complements and some additional distributional conditions are satisfied. In the latter case of supplier complementarity,

[16] With a single agent, the coalition of M and the agent behaves exactly the same way as the agent does, as a consequence of the Coase Theorem which applies when M is perfectly informed about the agent's cost.

therefore, DM is optimal among all the organizational modes considered: P is strictly better off hiring M, and then can delegate to him at no cost.[17] But in the substitutes case, it is strictly better for the principal to not hire M and contract personally with both suppliers.

6. *Applications to Industrial Organization*

The literature surveyed in this article has so far been primarily theoretical, without any significant interaction with more applied work on the internal organization of firms. In this section, I shall argue that the issues addressed by these theories are potentially relevant to applied work on industrial organization, and then discuss possible reasons why this potential has not yet been realized.

Part of the problem underlying the lack of interaction with applied work is that most of the evidence concerning internal organization of firms comprises case studies and business reports, instead of large scale empirical datasets. Nevertheless, some systematic empirical evidence has begun to become available in select contexts, which are discussed below.

6.1 *Human Resource Management Practices*

Ichniowski and Shaw (1999, 2003) discuss how U.S. businesses in the last twenty years have dramatically increased their use of "innovative" human resource management (HRM) practices. A detailed analysis of the nature of these practices and their productivity impact in the context of a sample of steel finishing production lines is provided in Ichniowski, Shaw, and Prennushi (1997) and Gant, Ichniowski, and Shaw (2002). Gant et al. (2002) provide a detailed description of different components of these practices, such as problem-solving teams,

incentive pay, flexible job design, information sharing among workers directly, training, careful hiring and selection, and employment security. They summarize by stating that

> . . . the underlying theme of the value of innovative HRM practices is that plants with innovative practices expect their employees to do more problem solving, both on and off the production line (Gant, Ichniowski, and Shaw 2002, p. 296).

Workers make decisions both individually and collectively as they work on the line, thus necessitating greater exchange of information within work teams, flexibility of job assignments and hiring of "smarter" workers. In contrast, traditional "control oriented" HRM practices concentrate decision-making responsibility among a few hierarchically superior agents, such as production foremen and supervisors. Ichniowski, Shaw, and Prennushi (1997) estimate the productivity impact of adoption of these practices to be statistically and quantitatively significant. They also find significant complementarities between different components of the HRM practices: each component by itself does not improve productivity.

The contrast between the innovative and traditional HRM practices corresponds closely to the distinction between mechanisms DS or DM and C described in previous sections. In the delegation mechanisms, production decisions are made by a team of agents A_1, A_2 and M on the basis of intrateam communication and group-based incentive pay for the team leader, who subsequently allocates production assignments and payments among team members. *If the key to the new HRM practices lie in delegation of production decisions to worker teams, their value cannot be explained by any conventional incentive-based theory of adverse selection or moral hazard that conforms to the Revelation Principle.* The outcome of any such decentralized mechanism can be replicated by a centralized mechanism which resembles traditional control-oriented

---

[17] However this result is driven by the assumption that M is perfectly informed about agents' costs. If M is imperfectly informed one expects delegation to M to be costly relative to CMC, but these costs ought to be small if M is "sufficiently" well-informed.

HRM practices in which each worker communicates and contracts separately with central management that makes all production decisions. Hence traditional incentive theory cannot explain the superior performance of innovative HRM practices. An appropriate theory would have to confront the issues raised in the literature described in this essay.

Naturally enough, it is evident that a theory explaining the superiority of innovative HRM practices would need to incorporate costs of communication, information processing, or contract renegotiation. In an empirical analysis of workplace reorganization in a large cross-section of 300 large U.S. firms during 1995–96, Timothy F. Bresnahan, Brynjolfsson, and Hitt (2002) find evidence of significant complementarities between adoption of new information technology (IT), innovative HRM practices, and hiring of skilled workers. Thomas N. Hubbard (2000) studies the effect of on-board computers in trucking, and its effects on the value of delegation. Massimo G. Colombo and Marco Delmastro (2004) examine the effect of communication technology on whether authority is delegated to plant managers. Theoretical explanations of these empirical findings would need to consider the implications of lowered costs of communication and information processing on the tradeoff between centralized and decentralized decision-making. Clearly, models that simultaneously incorporate information costs and incentives are necessary to address this question. The relative paucity of such models is one possible reason for the limited interchange between the theoretical and empirical literatures.

Nevertheless, the potential relevance of the theoretical literature is indicated by the model with communication cost described in section 4. Since centralized decision making necessitates greater exchange of information between production agents and decision makers, it might appear that lower communication cost should enhance the relative profitability of centralized decision making. Recall, however, that delegated decision making gives rise to a control loss, which has to be traded off against the enhanced flexibility of production decisions. The ability of the principal to moderate such control losses depends upon whether the principal can monitor decisions made by agents in a decentralized setting. Advanced IT enables such monitoring to be carried out by central management at much lower cost, as described by Brynjolfsson and Hitt (2000).[18] Hence the net effect of improved IT on the value of delegation is ambiguous in general, and consistent with observed complementarities between IT and innovative HRM practices. Needless to say, models addressing such questions explicitly are still awaited.

6.2 *Retail Franchising*

Benito Arruñada, Luis Garicano, and Luis Vázquez (2001) describe retail franchise agreements between automobile manufacturers and dealers in Spain. Retailers are delegated decisions concerning product pricing, purchases from manufacturers, advertising and after-sales service. This gives rise to a classic problem of "double marginalization of rents" across retail and manufacturing stages: decisions made by retailers affect profits earned by manufacturers. They also give rise to horizontal externalities across sales effort of different dealers: some customers gather information from one dealer and purchase from other dealers. To control these externalities, manufacturers retain certain monitoring rights:

---

[18] For example, they write that "Computer-based supply chain integration has been especially sophisticated in the consumer packaged goods industries . . . Because many consumer products have long shelf lives, retailers tended to buy massive amounts during promotional periods, which increased volatility in manufacturing schedules and distorted manufacturers' view of their market. In response, manufacturers sped up their packaging changes to discourage stock-piling of products and developed internal audit departments to monitor retailers' purchasing behavior for contractual violations" (Brynjolfsson and Hitt 2000, p. 28).

. . . they are authorized to directly inspect the inputs of the dealership (machinery, personnel etc.) and to measure dealer performance in terms of achievement of sales targets. Most contracts also allow manufacturers to monitor customer satisfaction, as measured by polls. Finally, most manufacturers may monitor dealers financial performance by regularly auditing their financial statements. Concerning termination rights, all contracts explicitly assign manufacturers the right to terminate the contracts when the dealer infringes certain duties (Arrunada, Garicano, and Vazquez 2001, p. 264).

Traditional incentive theory cannot explain why decisions over product pricing, purchases from manufacturers, advertising or after-sales service should be delegated to retailers by manufacturers who are clearly in a position to design the franchising relationship. When the Revelation Principle applies, delegated arrangements cannot outperform a centralized arrangement where all decision rights are retained by manufacturers. An obvious explanation of such a delegation arrangement is that it is either impossible, too expensive or time consuming for retailers to communicate all relevant local information to manufacturers. At the same time, in order to limit control loss, manufacturers retain the right to monitor dealers in a variety of ways.

Arrunada, Garicano, and Vazquez (2001) explain the prevalence of delegation by the presence of "two-sided moral hazard," i.e., the possibility of opportunistic ex post renegotiation of contracts by manufacturers. Consistent with this hypothesis, they find that manufacturers tend to delegate more to retailers when manufacturers are more reputable and thus less subject to renegotiation constraints. Nevertheless, the decisions that are subject to more or less delegation concern monitoring and enforcement of contractual terms. Decisions concerning pricing, advertising or after-sales service tend to be uniformly delegated, even amongst the most reputed manufacturers. The theoretical literature described in this essay will be needed to explain the widespread prevalence of such kinds of delegation even for reputable manufacturers for whom the moral hazard problem is of the "one-sided" variety.

7. Concluding Comments

In summary, the most important lacuna of existing theoretical incentive-based literature is that it focuses on costs rather than the benefits of delegation. The latter are difficult to incorporate into traditional contract theory. Perhaps the most important benefit of delegation is the distribution of information processing tasks, but no progress has occurred in theories that marry information processing costs with incentives. Some progress has been possible with communication costs and simple measures of contract complexity, but these need better foundations.

What have we learned from the existing literature? It identifies a number of potential costs of delegation: moral hazard for intermediaries owing to noncoincidence of their own objectives with the principal's, and their monopsony power over subordinates. These can result in production distortions (insufficient sourcing from subordinates), cascading of information rents across vertical layers, and problems of coordinating different horizontal branches.

If agents do not collude, these agency costs of delegation can be avoided if (and only if) the principal can monitor subcontract costs or quantities, if contracts flow down the hierarchy, and agents are risk-neutral. If any one of these conditions do not hold then agency costs cannot be avoided. The only significant problem pertains to vertical control loss; if they can be avoided (i.e., under the above mentioned conditions) then incentive considerations do not complicate horizontal coordination across branches of the hierarchy: "group"-based incentive contracts can be designed to costlessly internalize these horizontal externalities. On the other hand, managerial risk aversion or limited capacity for principals to monitor local conditions or agent decisions can cause significant control losses from delegation, that grow with the

size and complexity of the organization. This provides an explanation of organizational diseconomies of scale, i.e., why larger firms tend to be more "bureaucratic" and less able to control costs.

If agents collude, centralization is also subject to unobserved side contracting among agents, limiting the ability of the principal to moderate "control loss." However, centralization potentially allows greater control over side contracting outcomes by the principal offering outside options to subordinates that limit monopsony power of intermediaries. Depending on the precise distortions engendered, this added dimension of control may or may not be valuable. Overall, the presence of collusion among agents enlarges the range of circumstances where delegation implements optimal allocations.

There are numerous open questions and fruitful avenues for future research. I conclude by listing some of these.

First, a better understanding of effects of collusion is still needed. The few papers on this topic emerge with different results the intuitive basis for which is not very clear. One hopes a more unified perspective will emerge in due course. There is a need to explore implications of different formulations of side-contracting, e.g., more general assignment of bargaining power within coalitions, or alternative timing assumptions. Baron and Besanko (1999) provide an intriguing model in which agents themselves decide ex ante whether to consolidate themselves into a single entity, a decision which the principal observes and takes into account before offering a contract. In the models we described, the principal can anticipate a particular pattern of side contracting, but cannot observe whether or not the agents actually do side-contract. In contexts with more agents and vertical layers, the possibility of collusion-within-collusion further complicates the analysis.

Second, more effort needs to be devoted to explaining the potential benefits of delegation.

Models integrating information processing or communication costs with incentive considerations are needed to provide a full-blown theory of the trade-off between centralization and decentralization. This would render the theory useful in applied work assessing the effectiveness of innovative human resource management practices and their complementarity with new information technology.

A third possible avenue would consider applications and extensions to contexts involving more productive agents and a richer specification of the production technology. Questions concerning the optimal shape of hierarchies can then be addressed, e.g., trade-offs between span of control and number of vertical layers, how to group agents within departments, organizational diseconomies of scale, and the advantages of nonhierarchical organizations. One hopes that both theory and empirical datasets regarding these organizational attributes can be developed interactively, permitting better understanding of their productivity implications, and how they respond to changes in market competition or information technology.

REFERENCES

Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105(1): 1–29.

Aoki, Masahiko. 1990. "Toward an Economic Model of the Japanese Firm." *Journal of Economic Literature*, 28(1): 1–27.

Arrow, Kenneth. 1974. *The Limits of Organization.* New York: W. W. Norton.

Arrunada, Benito, Luis Garicano, and Luis Vazquez. 2001. "Contractual Allocation of Decision Rights and Incentives: The Case of Automobile Distribution." *Journal of Law, Economics, and Organization*, 17(1): 257–84.

Baliga, Sandeep, and Tomas Sjostrom. 1998. "Decentralization and Collusion." *Journal of Economic Theory*, 83(2): 196–232.

Baliga, Sandeep, and Tomas Sjostrom. 2001. "Optimal Design of Peer Review and Self-Assessment Schemes." *RAND Journal of Economics*, 32(1): 27–51.

Baron, David P. 1985. "Noncooperative Regulation of a Nonlocalized Externality." *RAND Journal of Economics*, 16(4): 553–68.

Baron, David P., and David Besanko. 1992. "Information, Control, and Organizational Structure." *Journal of*

Economics and Management Strategy, 1(2): 237–75.

Baron, David P., and David Besanko. 1999. "Informational Alliances." *Review of Economic Studies*, 66(4): 743–68.

Beaudry, Paul, and Michel Poitevin. 1995. "Contract Renegotiation: A Simple Framework and Implications for Organization Theory." *Canadian Journal of Economics*, 28(2): 302–35.

Bolton, Patrick, and Mathias Dewatripont. 1994. "The Firm as a Communication Network." *Quarterly Journal of Economics*, 109(4): 809–39.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt. 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics*, 117(1): 339–76.

Brynjolfsson, Erik, and Lorin M. Hitt. 2000. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *Journal of Economic Perspectives*, 14(4): 23–48.

Calvo, Guillermo A., and Stanislaw Wellisz. 1978. "Supervision, Loss of Control, and the Optimum Size of the Firm." *Journal of Political Economy*, 86(5): 943–52.

Celik, Gorkem. Forthcoming. "Mechanism Design with Collusive Supervision." *Review of Economic Studies*.

Chandler, Alfred D. 1962. *Strategy and Structure*. Cambridge: MIT Press.

Colombo, Massimo G., and Marco Delmastro. 2004. "Delegation of Authority in Business Organizations: An Empirical Test." *Journal of Industrial Economics*, 52(1): 53–80.

Cremer, Jacques, and Michael H. Riordan. 1987. "On Governing Multilateral Transactions with Bilateral Contracts." *RAND Journal of Economics*, 18(3): 436–51.

Dessein, Wouter. 2002. "Authority and Communication in Organizations." *Review of Economic Studies*, 69(4): 811–38.

Faure-Grimaud, Antoine, and David Martimort. 1999. "Political Stabilization by an Independent Bureaucracy." London School of Economics. Mimeo.

Faure-Grimaud, Antoine, and David Martimort. 2001. "On Some Agency Costs of Intermediated Contracting." *Economics Letters*, 71(1): 75–82.

Faure-Grimaud, Antoine, Jean-Jacques Laffont, and David Martimort. 2000. "A Theory of Supervision with Endogenous Transaction Costs." *Annals of Economics and Finance*, 1(2): 231–63.

Faure-Grimaud, Antoine, Jean-Jacques Laffont, and David Martimort. 2003. "Collusion, Delegation and Supervision with Soft Information." *Review of Economic Studies*, 70(2): 253–79.

Gant, Jon, Casey Ichniowski, and Kathyrn Shaw. 2002. "Social Capital and Organizational Change in High-Involvement and Traditional Work Organizations." *Journal of Economics and Management Strategy*, 11(2): 289–328.

Gilbert, Richard J., and Michael H. Riordan. 1995. "Regulating Complementary Products: A Comparative Institutional Analysis." *RAND Journal of Economics*, 26(2): 243–56.

Hammer, Michael M., and James Champy. 1993.

Reengineering the Corporation. New York: Harper Collins.

Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Oxford and New York: Oxford University Press and Clarendon Press.

Heal, Geoffrey M. 1971. "Planning, Prices, and Increasing Returns." *Review of Economic Studies*, 38(115): 281–94.

Horngren, Charles T., and George Foster. 1991. *Cost Accounting: A Managerial Emphasis*. Englewood Cliffs, N.J.: Prentice-Hall.

Hubbard, Thomas N. 2000. "The Demand for Monitoring Technologies: The Case of Trucking." *Quarterly Journal of Economics*, 115(2): 533–60.

Hurwicz, Leonid. 1960. "Optimality and Informational Efficiency in Resource Allocation Processes," in *Mathematical Methods in the Social Sciences*. K. Arrow, S. Karlin, and P. Suppes, eds. Stanford: Stanford University Press, 27–46.

Hurwicz, Leonid. 1972a. "On Informationally Decentralized Systems," in *Decision and Organization*. R. Radner and C. McGuire, eds. Amsterdam: North-Holland, 297–336.

Hurwicz, Leonid. 1972b. "On the Dimensional Requirements of Informationally Decentralized Pareto-Satisfactory Processes." Paper presented at the Decentralization Conference, Northwestern University.

Hurwicz, Leonid. 1973. "The Design of Mechanisms for Resource Allocation." *American Economic Review*, 63(2): 1–30.

Hurwicz, Leonid. 1986. "Incentive Aspects of Decentralization," in *Handbook of Mathematical Economics, Vol. 3*. Kenneth J. Arrow and Michael D. Intrilligator, eds. Amsterdam: North-Holland, 1442–81.

Ichniowski, Casey, and Kathryn Shaw. 1999. "The Effects of Human Resource Management Systems on Economic Performance: An International Comparison of U.S. and Japanese Plants." *Management Science*, 45(5): 704–21.

Ichniowski, Casey, and Kathryn Shaw. 2003. "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices." *Journal of Economic Perspectives*, 17(1): 155–80.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." *American Economic Review*, 87(3): 291–313.

Laffont, Jean-Jacques, and David Martimort. 1998. "Collusion and Delegation." *RAND Journal of Economics*, 29(2): 280–305.

Malinvaud, Edmond. 1967. "Decentralized Procedures for Planning," in *Activity Analysis in the Theory of Growth and Planning*. E. Malinvaud and M. Bacharach, eds. London: MacMillan, 170–208.

Marschak, Thomas, and Stefan Reichelstein. 1995. "Communication Requirements for Individual Agents in Networks and Hierarchies," in *The Economics of Informational Decentralization: Complexity, Efficiency, and Stability*. John O.

Ledyard, ed. Boston: Kluwer Academic, 311–46.

Marschak, Thomas, and Stefan Reichelstein. 1998. "Network Mechanisms, Informational Efficiency, and Hierarchies." *Journal of Economic Theory*, 79(1): 106–41.

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

Melumad, Nahum D., Dilip Mookherjee, and Stefan Reichelstein. 1992. "A Theory of Responsibility Centers." *Journal of Accounting and Economics*, 15(4): 445–84.

Melumad, Nahum D., Dilip Mookherjee, and Stefan Reichelstein. 1995. "Hierarchical Decentralization of Incentive Contracts." *RAND Journal of Economics*, 26(4): 654–72.

Melumad, Nahum D., Dilip Mookherjee, and Stefan Reichelstein. 1997. "Contract Complexity, Incentives, and the Value of Delegation." *Journal of Economics and Management Strategy*, 6(2): 257–89.

Mookherjee, Dilip, and Stefan Reichelstein. 1997. "Budgeting and Hierarchical Control." *Journal of Accounting Research*, 35(2): 129–55.

Mookherjee, Dilip, and Stefan Reichelstein. 2001. "Incentives and Coordination in Hierarchies." *Advances in Theoretical Economics*, 1(1).

Mookherjee, Dilip, and Masatoshi Tsumagari. 2004. "The Organization of Supplier Networks: Effects of Delegation and Intermediation." *Econometrica*, 72(4): 1179–1219.

Mount, Kenneth, and Stanley Reiter. 1974. "The Informational Size of Message Spaces." *Journal of Economic Theory*, 8(2): 161–92.

Mount, Kenneth, and Stanley Reiter. 1995. "A Theory of Computing with Human Agents." Northwestern University Department of Economics Working Paper.

Myerson, Roger B. 1982. "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems." *Journal of Mathematical Economics*, 10(1): 67–81.

Poitevin, Michel. 1995. "Contract Renegotiation and Organizational Design." CIRANO Working Paper 95-3.

Poitevin, Michel. 2000. "Can the Theory of Incentives Explain Decentralization?" *Canadian Journal of Economics*, 33(4): 878–906.

Radner, Roy. 1992. "Hierarchy: The Economics of Management." *Journal of Economic Literature*, 30(3): 1382–1415.

Radner, Roy. 1993. "The Organization of Decentralized Information Processing." *Econometrica*, 61(5): 1109–46.

Radner, Roy, and Jacob Marschak. 1972. *Economic Theory of Teams*. New Haven: Yale University Press.

Roider, Andreas. 2003. "Delegation of Authority as an Optimal (In)complete Contract." University of Bonn Discussion Paper.

Segal, Ilya. 1999. "Complexity and Renegotiation: A Foundation for Incomplete Contracts." *Review of Economic Studies*, 66(1): 57–82.

Segal, Ilya. 2001. "Communication Complexity and Coordination by Authority." University of California, Berkeley. Mimeo.

Severinov, Sergei. 1999. "Optimal Organization: Centralization, Decentralization or Delegation?" University of Wisconsin, Madison. Mimeo.

Spence, A. Michael. 1975. "The Economics of Internal Organization: An Introduction." *Bell Journal of Economics*, 6(1): 163–72.

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

Tirole, Jean. 1999. "Incomplete Contracts: Where Do We Stand?" *Econometrica*, 67(4): 741–81.

van Zandt, Timothy. 1996. "Decentralized Information Processing in the Theory of Organizations," in *Contemporary Economic Development Reviewed, Vol. 4: The Enterprise and its Environment*. Murat Sertel, ed. London: Macmillan Press, 125–60.

van Zandt, Timothy. 1997. "Real-Time Hierarchical Resource Allocation." Princeton University. Mimeo.

Williamson, Oliver. 1967. "Hierarchical Control and Optimal Firm Size." *Journal of Political Economy*, 85(2): 123–38.

Williamson, Oliver. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.