Département de sciences économiques

Cahier 2000-13

Innis Lecture 2000 : Can the Theory of Incentives Explain Decentralization?

POITEVIN, Michel

Université de Montréal

Département de sciences économiques
Université de Montréal
Faculté des arts et des sciences
C.P. 6128, succursale Centre-Ville
Montréal (Québec) H3C 3J7
Canada
http://www.sceco.umontreal.ca
SCECO-information@UMontreal.CA
Téléphone : (514) 343-6539
Télécopieur : (514) 343-7221

Ce cahier a également été publié par le Centre interuniversitaire de recherche en économie quantitative (CIREQ) sous le numéro 14-2000.

This working paper was also published by the Center for Interuniversity Research in Quantitative Economics (CIREQ), under number 14-2000.

ISSN 0709-9231

CAHIER 2000-13

INNIS LECTURE 2000 :

CAN THE THEORY OF INCENTIVES EXPLAIN DECENTRALIZATION?

Michel POITEVIN[1]

[1] Centre de recherche et développement en économique (C.R.D.E.) and Département de sciences économiques, Université de Montréal, and CIRANO

September 2000

The author would like to thank Chuck Blackorby for giving him the opportunity to deliver the 2000 Innis Lecture (Canadian Economics Association, Vancouver, June 2000), he greatly benefited from this exercise. He would also like to thank Frédérique Fell for her valuable research assistance. The author acknowledges financial support from CRSH (Canada) and FCAR (Québec).

RÉSUMÉ

Cet article présente, dans un cadre unifié, un survol de trois théories de la décentralisation de la prise de décision dans les organisations. Ces théories reposent sur la présence d'information privée et des incitations qui en découlent. La renégociation, la collusion et les limites à la communication représentent trois conditions suffisantes pouvant expliquer l'optimalité de la décentralisation.

Mots clés : design organisationnel, information privée, décentralisation

ABSTRACT

This survey presents within a single model three theories of decentralization of decision-making within organizations based on private information and incentives. Renegotiation, collusion, and limits on communication are three sufficient conditions for decentralization to be optimal.

Key words : organizational design, private information, decentralization

1 Introduction

Decentralization of decision-making is pervasive in many economic organizations. For example, shareholders typically decentralize management of the firm to professional managers. Many firms outsource part of their production process to external suppliers. We may think of personnel management, information systems, maintenance, parts, etc. A public regulator decentralizes production decisions to a regulated private enterprise. The structure of governments itself has some degree of decentralization since powers may be allocated to different levels of government. What then are the arguments in favor of more or less decentralization? For a long time, economists have suggested that communication was central to the debate.[1] The question was whether centralization or decentralization performs better in the presence of exogenously specified communication costs. Earlier papers neglected the incentives to reveal or withhold information.

More recently, incentives have entered the picture, namely, incentives to provide valuable private information to decision-makers. The basic argument in favor of decentralization is that it allows economizing on communication costs since the decision is delegated to the agent who possesses the most relevant information. The cost of doing so, however, is a potential loss of control by the principal, because the agent may not have its preferences perfectly aligned with the objectives of the principal, therefore leading to agency costs. A basic tradeoff emerges: more decentralization economizes on communication costs while also leading to a control loss on the part of the principal. The resolution of this tradeoff potentially can lead to a theory of organizational structure.

Despite this clear enough intuition, it has been a theoretical challenge to explain formally why decentralization of decision-making should take place in an organization. It can be shown that any decentralized organization can be replicated by a centralized one in which all agents report their private information to the principal, who then makes all decisions. In the centralized organization, the principal replicates what the agents would have done in the

[1] For example, see Marschak (1959) and Groves and Radner (1972).

decentralized organization conditional on their reported information. When the principal is committed to act as such, agents have no interest in lying, and they therefore report truthfully their private information. This implies that centralization is at least weakly preferred to any decentralized organization. This classic result is known as the "Revelation Principle."[2] The central question then is: can the theory of incentives explain decentralization of decision-making?

The Revelation Principle gives us a useful benchmark against which any theory of decentralization can be assessed. Any such theory must therefore start with relaxing some assumption(s) underlying this result. Three main avenues have been explored so far in the literature.[3]

A first assumption is that the principal is committed to obey the rules set out ex ante; namely, it is committed to act upon the reported information as specified in the initial contract that has been signed with all agents. If the principal cannot commit to doing so, renegotiation may occur, leading to the partial destruction of ex ante incentives (when such renegotiation is rationally anticipated by the agents). In some instances, the possibility for renegotiation may create a clear preference for decentralization.

A second assumption is that agents act non-cooperatively when reporting their private information to the principal. If they are allowed to collude and coordinate their reporting strategy, ex ante incentives may again be affected. Decentralization may be part of the solution for avoiding collusion.

A final assumption is that communication between the principal and the agents is absolutely costless. If such communication is costly, it may be worthwhile to decentralize decision-making to informed agents, thus avoiding costly communication.[4]

These three assumptions are related in some sense to communication costs. Renegoti-

[2] See, for example, Gibbard (1973), Green and Laffont (1977), or Myerson (1982).

[3] Relevant citations to each of these theories are given when they are formally presented.

[4] Recently, a fourth avenue has been explored, that of multiprincipals. These ideas are not presented here. The interested reader is referred to the survey of Laffont and Martimort (1997b).

ation is costless communication between the principal and the agents, collusion is costless communication between agents, and limits to communication suppose costly communication between the principal and the agents. The result for the weak superiority of centralization over decentralization can then be rephrased in terms of communication between the agents and the principal being costless when they report their information, with all other communication being infinitely costly. This seems to be an unreasonable assumption in many instances.

Relaxing this assumption about communication costs may provide a rationale for a decentralized organization. Delegation of decision-making to informed agents reduces the need for communicating and may help to avoid some of the costs associated with communication. There are, however, costs to decentralize in terms of loss of control and agency costs. Informed agents may not have all the incentives to perform in the best interests of the principal.

The purpose of this article is to survey the recent literature on decentralization within the context of a unified and simple model that underlines the role of communication costs for organizational design. In turn, the effects of renegotiation, collusion, and limits to communication on centralized and decentralized organizations are presented in detail, and the optimal organizational design is characterized for each case. This exercise provides a better understanding of what is known about the design of economic organizations, but also, and more important, about what is not known!

Before starting with the formal presentation, I would like to mention that I am leaving out large bodies of literature that have taken a different approach to the study of organizational structure. For example, many authors study the benefits and costs of vertical integration within the paradigm of "incomplete contracts."[5] There is also a literature in which is studied decision-making in teams or organizations when there is collective bounded rationality due to exogenously specified information-processing costs.[6] Most papers in this

[5] See among many others, Arrow (1975), Grossman and Hart (1986), and Williamson (1975).

[6] For example, see Marschak (1959) and Groves and Radner (1972).

literature do not explicitly consider incentives to report truthfully private information. I have decided not to cover these theories because I want to focus on the theory of incentives in explaining decentralization within the framework of complete contracts. Also, in writing this survey, I purposely chose to present with formal details the main arguments for a theory of decentralization. This allows the reader to get a better grasp of the main insights. The cost of doing so, however, is that I had to leave out many other deserving results that may be related but are not central to the understanding of the main arguments. I apologize to all authors who feel they are affected by this strategy.

In the next section I present the model. It is useful to adopt the simplest model possible in order to nest all theories within the same framework. I have therefore adopted the model of Laffont and Martimort (1997a), which itself is a special case of that of Mookherjee and Reichelstein (1992). In Section 3, the second-best optimal allocation is implemented by both the centralized and decentralized mechanisms. A discussion of the Revelation Principle follows. In the next three sections the cases of renegotiation, collusion, and limits to communication, respectively, are considered. A discussion of possible extensions and a conclusion follow.

2 The model

A principal must hire two productive agents to undertake a project. Each agent incurs a personal cost $C(\theta_i, q_i) = \theta_i q_i$, where q_i is the amount produced by agent i, and θ_i is a random shock that affects agent i's productivity. Let $\theta_i \in \left\{\underline{\theta}, \bar{\theta}\right\}$ and $\Delta\theta = \bar{\theta} - \underline{\theta} > 0$. Shocks are i.i.d.; that is, there is no correlation in the distribution of shocks of the two agents. The probability of $\underline{\theta}$ is ν. Agent 2 produces an intermediate input that is to be used by agent 1, who produces the final output. To simplify the analysis, the technology is assumed to be Leontieff; that is, $q_1 = q_2 = q$. Agent i is risk neutral and its preference over wages w and production q is $w - \theta_i q$. The principal derives a private benefit from the agents' production denoted by $B(q)$, with $B' > 0$ and $B'' < 0$. The utility of the (risk-neutral) principal is

$B(q) - w$.

There are four possible states of nature: $\theta = (\theta_1, \theta_2) \in \left\{(\underline{\theta}, \underline{\theta}), (\underline{\theta}, \bar{\theta}), (\bar{\theta}, \underline{\theta}), (\bar{\theta}, \bar{\theta})\right\}$. Throughout, I focus on symmetric allocations and I use the following notational convention:

$$\underline{z} = z(\underline{\theta}, \underline{\theta}); \hat{z}_1 = z(\underline{\theta}, \bar{\theta}); \hat{z}_2 = z(\bar{\theta}, \underline{\theta}), \bar{z} = z(\bar{\theta}, \bar{\theta}),$$

where z can be either w_i or q. When output is a function of only the sum of shocks, one has $\hat{q}_1 = \hat{q}_2$. The subscript is often dropped and $\hat{q} \equiv \hat{q}_1 = \hat{q}_2$. An allocation is a vector $\{q(\theta), w(\theta)\}$ for all θ, where $w = (w_1, w_2)$.

Two remarks are in order at this point.

1. This model is one of private values since the productivity parameter of one agent enters neither the cost function of the other agent nor the principal's benefit. This is obviously a simplifying assumption, but one that is not totally innocuous, especially when one introduces renegotiation or collusion.

2. Given the technology, there are obvious benefits to coordinating the production decision of the two agents. The need for coordination is a "necessary" condition for the analysis of organizational design to be interesting. It is an important consideration to take into account in any decision to decentralize or not. It is related to the problem of control loss mentioned in the Introduction.

The problem of implementing an allocation can be solved by designing a contract or mechanism that stipulates production and transfers. This contract is negotiated among the three players. The specifics of the contract determine the structure of the organization. The challenge is to characterize the optimal contract (or organizational form) under various informational and bargaining assumptions. Throughout the paper, I assume that the principal has the bargaining power in all negotiations.

Before going to the analysis of organizational design *per se*, I characterize the symmetric-information optimal allocation. This helps to better grasp the intuition of the effect of asymmetric information. To facilitate comparisons with other informational environments, I assume the following timing of events. First, each agent privately observes its productivity shock. The principal then offers a contract to both agents specifying transfers and production contingent on the realization of shocks. If this contract is rejected by at least one agent, the game ends with all players earning their reservation utility normalized to be zero. If both agents accept the contract, shocks become common knowledge and verifiable. Production and transfers take place and the game ends.

The optimal allocation maximizes the expected profit of the principal subject to the condition that agents are willing to participate in the production process. The optimal allocation is the solution to the following maximization problem.

$$\begin{array}{ll} \max\limits_{q(\theta),w(\theta)} & \mathrm{E}_\theta \left\{ B(q(\theta)) - \sum\limits_{i=1}^{2} w_i(\theta) \right\} \\ \text{s.t.} & \mathrm{E}_{\theta_j} \left\{ w_i(\theta) - \theta_i q(\theta) \right\} \geq 0 \quad \text{for all } \theta_i,\ i,j = 1,2, i \neq j \end{array}$$

The constraint is agent i's participation constraint, where the right-hand-side represents agent i's reservation utility. This constraint holds in expectation over the shocks of agent j since agent i must accept or reject the contract before learning the shock of agent j. The constraint must hold, however, for all values of θ_i. This is called an *interim* participation constraint.

In a first-best world, production is ex post efficient, the marginal benefit of the principal being equal to the realized marginal cost of production.

$$\begin{aligned} B'(\underline{q}^1) &= 2\underline{\theta} \\ B'(\hat{q}^1) &= \underline{\theta} + \bar{\theta} \\ B'(\bar{q}^1) &= 2\bar{\theta} \end{aligned}$$

where the "1" in superscript refers to first best.

Wages are such that all interim participation constraints are strictly binding. All agents earn zero expected rent.

$$
\begin{aligned}
U_i^1\left(\bar{\theta}\right) &\equiv \nu \hat{w}_{i2}^1 + (1-\nu)\bar{w}_i^1 - \bar{\theta}\left(\nu \hat{q}^1 + (1-\nu)\bar{q}^1\right) = 0 \\
U_i^1\left(\underline{\theta}\right) &\equiv \nu \underline{w}_i^1 + (1-\nu)\hat{w}_{i1}^1 - \underline{\theta}\left(\nu \underline{q}^1 + (1-\nu)\hat{q}^1\right) = 0
\end{aligned}
$$

Because of risk neutrality, only expected transfers matter, and there are many transfers that solve these two equations.

I now assume that productivity shocks are private information to each agent. The shock becomes the type of an agent.

3 Asymmetric information

Before we formally analyze different organizational structures, it is useful to characterize the second-best optimal allocation under the assumption that information is private. Agents privately observe their type, and types never become common knowledge. I delay the problem of implementation of this allocation to the next section.

The second-best optimal allocation solves the following maximization problem.

$$
\begin{aligned}
\max_{q(\theta), w(\theta)} \quad & \mathrm{E}_\theta \left\{ B(q(\theta)) - \sum_{i=1}^{2} w_i(\theta) \right\} \\
\text{s.t.} \quad & \text{for all } \theta, \tilde{\theta}_i,\ i, j = 1, 2, i \neq j : \\
& \mathrm{E}_{\theta_j}\left\{ w_i(\theta) - \theta_i q(\theta) \right\} \geq 0 \\
& \mathrm{E}_{\theta_j}\left\{ w_i(\theta_i, \theta_j) - \theta_i q(\theta_i, \theta_j) \right\} \geq \mathrm{E}_{\theta_j}\left\{ w_i(\tilde{\theta}_i, \theta_j) - \theta_i q(\tilde{\theta}_i, \theta_j) \right\}
\end{aligned}
$$

The second set of constraints represents incentive constraints that stipulate that agent i prefers to obtain the allocation designed for its type rather than the one designed for type j. It is a standard truth-telling constraint. As for the participation constraint, it holds for all types θ_i, and in expectation over types θ_j. It is an *interim* incentive-compatibility constraint.

The first-best optimal allocation does not satisfy the incentive constraint of efficient types and therefore cannot be the solution. This is easily explained. If the participation constraint of non-efficient types is strictly satisfied at their first-best allocation, it must be the case that efficient types can earn rents at that same allocation, since they can produce at a lower cost. The rent they would gain by pretending to be inefficient is the difference in productivity times the expected output of the non-efficient type; that is, $\Delta\theta(\nu\hat{q}^1 + (1-\nu)\bar{q}^1)$. To satisfy the incentive constraint, the principal can either raise the transfer of efficient agents by this amount and/or reduce $\hat{q}$ and $\bar{q}$. In effect, the principal does both until the incentive constraint of efficient types is satisfied.

Since efficient agents want to understate their productivity in order to produce less and still earn a relatively high wage, the principal must offer rents to the efficient agents when they report truthfully. To limit these rents, the principal reduces the production expected of less efficient types, therefore making their allocation less attractive to efficient types. This intuition is the essence of the solution. As shown in Laffont and Martimort (1998), second-best optimal production is implicitly defined by

$$\begin{aligned}
B'(\underline{q}^2) &= 2\underline{\theta} \\
B'(\hat{q}^2) &= \underline{\theta} + \bar{\theta} + \frac{\nu\Delta\theta}{1-\nu} \\
B'(\bar{q}^2) &= 2\bar{\theta} + 2\frac{\nu\Delta\theta}{1-\nu}
\end{aligned}$$

where the "2" in superscript refers to second best. If both agents are efficient, there is no distortion since second-best output is equal to first-best output. It is not necessary to distort $\underline{q}$ because it plays no role in reducing rents from mimicking. This is the standard "no distortion at the top" result. All other outputs are distorted downwards.

It is instructive to examine more closely the size of the distortion. It depends on the term $\nu\Delta\theta/(1-\nu)$. This term represents the shadow cost for the principal of the rents it leaves to efficient types. For example, when the principal increases $\bar{q}$ marginally, it must support the direct marginal cost of production equal to $2\bar{\theta}$ (since it must compensate agents), plus an indirect cost equivalent to the increase in expected rents that it must concede to efficient agents in order to satisfy their incentive constraint. This last term is proportional to the difference in productivity $\Delta\theta$ and the relative likelihood of an efficient agent $\nu/(1-\nu)$. The sum of the marginal cost of production and the marginal cost of rents is called the marginal "virtual cost" of the principal. This is the total marginal cost the principal must support when it increases $\bar{q}$. The marginal benefit of production is then equal to the marginal virtual cost. A similar reasoning can explain the expression for $\hat{q}$.

Wages are such that the interim participation constraint of an inefficient agent is strictly binding, while an efficient agent earns informational rents.

$$\begin{aligned} U_i^2\left(\bar{\theta}\right) &\equiv \nu\hat{w}_{i2}^2 + (1-\nu)\bar{w}_i^2 - \bar{\theta}\left(\nu\hat{q}^2 + (1-\nu)\bar{q}^2\right) = 0 \\ U_i^2\left(\underline{\theta}\right) &\equiv \nu\underline{w}_i^2 + (1-\nu)\hat{w}_{i1}^2 - \underline{\theta}\left(\nu\underline{q}^2 + (1-\nu)\hat{q}^2\right) = \Delta\theta(\nu\hat{q}^2 + (1-\nu)\bar{q}^2) > 0 \end{aligned}$$

Again, only expected transfers matter, and there are many transfers that solve these two equations.[7]

3.1 Implementation of the second-best optimal allocation

There are many mechanisms that can implement q^2. I focus on two such mechanisms: a centralized one and a decentralized one. The definition of centralization or decentralization is based on an "intuitive" interpretation of the game players play. Centralization corresponds to a mechanism in which the principal retains all rights to contract and produce. Decentralization corresponds to a mechanism where the principal delegates some of these rights to an agent.

• Centralization

We define "***centralization***" by the following mechanism or game.

[7] This implies that it is possible to implement the second-best solution in dominant strategies, that is, when all incentive constraints hold ex post (for any type of the other agent). See Mookherjee and Reichelstein (1992).

1. Agent i learns its private information θ_i.

2. The principal offers a contract to both agents. If both accept, the game continues. If not, the game ends and each player receives its reservation utility.

3. The two agents simultaneously send a message to the principal.

4. The principal makes a production decision and pays the agent.

In stage 2, the principal offers a contract to both agents. This contract has the following form: the principal commits to an allocation $\{q(m_1, m_2), w(m_1, m_2)\}$, where m_i is the message sent by agent i at stage 3. Without loss of generality, the message sent by an agent can be taken to be a report of its type, that is, $m_i \in \left\{\underline{\theta}, \bar{\theta}\right\}$. There is no obligation for an agent to report truthfully; it must be provided with incentives to do so. Once messages have been sent, the principal makes a production decision and pays the agents according to the contract.

The organization underlying this mechanism or game is called "centralized" because all production and contracting rights remain with the principal. It is the principal who decides on the level of production following the reports that agents have sent in. Agents, having no production or contracting rights, can only obey the principal's orders once they have decided to join the organization. Information flows to the top of the organization and is centralized with the principal, who then decides upon production and remuneration according to the contract agreed upon in stage 2.

The implemented allocation is the equilibrium allocation of this game. It is well known that this game implements q^2. Along the equilibrium path, the principal offers the contract supporting the allocation

$$\{q(m_1, m_2), w(m_1, m_2)\}_{m_1, m_2 = \underline{\theta}}^{\bar{\theta}} = \left\{(\underline{q}^2, \underline{w}^2), (\hat{q}^2, \hat{w}_1^2), (\hat{q}^2, \hat{w}_2^2), (\bar{q}^2, \bar{w}^2)\right\}.$$

The agents accept it, then simultaneously report truthfully. All players have the same expected payoff $U_i^2(\theta)$ as they do with the second-best optimal allocation. Centralization achieves the maximum that can be attained in a second-best world.

• Decentralization

There are many different games that may represent a decentralized organization. I adopt one that fits well most definitions of decentralization and that is sufficiently simple to illustrate in which circumstances decentralization dominates centralization. "***Decentralization***" is defined by the following game.[8]

1. Agent i learns its private information θ_i.
2. The principal offers a contract to agent 1. If it is accepted, the game continues. If not, the game ends and each player receives its reservation utility.
3. Agent 1 sends a message to the principal.
4. Agent 1 offers a contract to agent 2. If it is accepted, the game continues. If it is rejected, the game ends and each player receives its reservation utility.
5. Agent 2 sends a message to agent 1.
6. Agent 1 makes a production decision, and transfers take place.

Before I describe in more detail the nature of the contracts, it is useful to make some informational assumptions.[9] First, the principal can monitor output, but not the intermediate input q_2. Second, the message sent by agent 1 to the principal in stage 3 is not observed by agent 2. This is without consequence, since preferences are quasilinear and agent 2's preferences do not depend on θ_1. The informed-principal problem described by Maskin and Tirole (1990), therefore, does not arise here. Third, the principal cannot observe the interaction (contract and message) between agents 1 and 2. This assumption can only reduce the efficiency of the decentralized mechanism. The opposite assumption would serve only

[8]This is essentially the game proposed by Melumad, Mookherjee, and Reichelstein (1995) to study delegation within an organization.

[9]These are the assumptions Melumad, Mookherjee, and Reichelstein (1995) make.

the principal since it could use its contract with agent 1 to force an appropriate contract between agents 1 and 2.

The contract between the two agents negotiated in stage 4 specifies production and a transfer for agent 2 as a function of the message of agent 2; that is, $\{q(m_2), w_2(m_2)\}$. Note that since the principal cannot observe the interaction between the two agents, it is agent 1 that remunerates agent 2 out of its own pocket. The contract between the principal and agent 1 offered in stage 2 specifies a transfer for agent 1 as a function of its message and the output that agent 1 delivers to the principal; that is, $w_1(q, m_1)$. The transfer w_1 cannot depend on the message sent by agent 2, since the principal cannot observe it. It is then optimal to base the transfer on output, which may serve as a signal of the type of agent 2.

The organization underlying this mechanism or game is called "decentralized" because the principal delegates production rights and contracting rights with agent 2 to agent 1. Agent 1 is in charge of ordering to agent 2 the production of the intermediate input, and of producing and delivering the final output to the principal. Agent 1 is also in charge of hiring and remunerating agent 2, that is, contracting with agent 2. Agent 1 must pay agent 2 out of its own pocket, meaning that agent 1 is accountable for agent 2's budget. Agent 1 effectively becomes a responsibility or profit center.[10] The principal has delegated significant rights to agent 1. The exercise of these rights is nevertheless implicitly limited by the contract that the principal offers to agent 1.

This decentralized mechanism can also implement q^2.[11] Let the payment to agent 1 be the following.

$$\begin{aligned} w_1^d(q, \bar{\theta}) &= B(q) - \frac{\nu \Delta\theta}{1-\nu} q - \beta^d(\bar{\theta}) \\ w_1^d(q, \underline{\theta}) &= B(q) - \beta^d(\underline{\theta}) \end{aligned}$$

[10] Melumad, Mookherjee and Reichelstein (1992) provide a detailed theory of responsibility centers.

[11] The proof of this adapts that of Melumad, Mookherjee, and Reichelstein (1995) which is valid for a continuum of types.

where the superscript "d" stands for decentralization, and

$$
\begin{aligned}
\beta^d(\bar{\theta}) &= \mathrm{E}_{\theta_2|\theta_1=\bar{\theta}}\left\{B(q^2) - \bar{\theta}q^2 - \frac{\nu\Delta\theta}{1-\nu}q^2 - \theta_2 q^2\right\} - \nu\Delta\theta\bar{q}^2 \\
\beta^d(\underline{\theta}) &= \mathrm{E}_{\theta_2|\theta_1=\underline{\theta}}\left\{B(q^2) - \underline{\theta}q^2 - \theta_2 q^2\right\} - \nu\Delta\theta\hat{q}^2 - \mathrm{E}_{\theta_2|\theta_1=\bar{\theta}}\Delta\theta q^2.
\end{aligned}
$$

The principal "sells out" the firm to agent 1. When agent 1 sends the message $m_1 = \underline{\theta}$, it receives the full benefit of production minus a constant $\beta^d(\underline{\theta})$. Agent 1 then contracts with agent 2 with the same incentives that the principal would have had if it was contracting with agent 2. Agent 1 minimizes the rents conceded to agent 2 and therefore implements the second-best optimal output:

$$
\begin{aligned}
B'(\underline{q}^2) &= 2\underline{\theta} \ \text{ and } \ \underline{w}_2^d = \underline{\theta}\underline{q}^2 + \Delta\theta\hat{q}^2 \ \text{ if agent 2 is efficient;} \\
B'(\hat{q}^2) &= \underline{\theta} + \bar{\theta} + \frac{\nu\Delta\theta}{1-\nu} \ \text{ and } \ \hat{w}_{21}^d = \bar{\theta}\hat{q}^2 \ \text{ if agent 2 is inefficient.}
\end{aligned}
$$

When agent 1 sends the message $m_1 = \bar{\theta}$, it receives the full benefit of production reduced by a tax which is equal to the marginal cost of rents that the principal has to concede to agent 1 and by a constant payment $\beta^d(\bar{\theta})$. By taxing production in state $\bar{\theta}$, the principal induces agent 1 to reduce production. This effectively limits the informational rents the principal has to pay to agent 1. Production is implicitly defined by

$$
\begin{aligned}
B'(\hat{q}^2) &= \underline{\theta} + \bar{\theta} + \frac{\nu\Delta\theta}{1-\nu} \ \text{ and } \ \hat{w}_{22}^d = \underline{\theta}\hat{q}^2 + \Delta\theta\bar{q}^2 \ \text{ if agent 2 is efficient;} \\
B'(\bar{q}^2) &= 2\bar{\theta} + 2\frac{\nu\Delta\theta}{1-\nu} \ \text{ and } \ \bar{w}_2^d = \bar{\theta}\bar{q}^2 \ \text{ if agent 2 is inefficient.}
\end{aligned}
$$

The constant terms β^d ensure that the participation constraint of type $\bar{\theta}$ and the incentive constraint of type $\underline{\theta}$ of agent 1 are strictly satisfied.

The fact that agent 1 has already reported when contracting with agent 2 is not important, given the assumptions of private values and risk neutrality. An allocation implemented with interim participation and incentive constraints can also be implemented with ex post constraints for agent 2. The extra degrees of freedom in setting transfers in the second-best optimal allocation can be used to satisfy ex post constraints for agent 2.

Given this contract with agent 1, the principal can implement q^2. The expected payment to all players is the same under decentralization as it is under centralization.

$$\begin{aligned}
U_1^d\left(\bar{\theta}\right) &\equiv \mathrm{E}_{\theta_2|\theta_1=\bar{\theta}}\left\{w_1^d(q^2,\bar{\theta}) - \bar{\theta}q^2 - w_2^d\right\} = 0 \\
U_1^d(\underline{\theta}) &\equiv \mathrm{E}_{\theta_2|\theta_1=\underline{\theta}}\left\{w_1^d(q^2,\underline{\theta}) - \underline{\theta}q^2 - w_2^d\right\} = \Delta\theta\left(\nu\hat{q}^2 + (1-\nu)\bar{q}^2\right) > 0 \\
U_2^d\left(\bar{\theta}\right) &\equiv \mathrm{E}_{\theta_1|\theta_2=\bar{\theta}}\left\{w_2^d - \bar{\theta}q^2\right\} = 0 \\
U_2^d(\underline{\theta}) &\equiv \mathrm{E}_{\theta_1|\theta_2=\underline{\theta}}\left\{w_2^d - \underline{\theta}q^2\right\} = \Delta\theta\left(\nu\hat{q}^2 + (1-\nu)\bar{q}^2\right) > 0
\end{aligned}$$

For agent 1, the first term represents its transfer from the principal, the second term, its cost of production, and the last term, the transfer it must pay to agent 2. For agent 2, its payoff corresponds to its transfer from agent 1 minus its cost of production. It can be readily verified that these expected payoffs are the same as those for the solution to the second-best problem or those under centralization. We therefore have equivalence between the two organizational forms.

The resolution of the decentralization game also justifies in a sense why this game is called "decentralized." Hurwicz (1969) defines a decentralized system as one where communication is restricted to commodity-dimensional messages. The nature of the contract between the principal and the agent has that flavor, since the remuneration of agent 1 depends on the produced output. Implicitly, the principal uses the produced output as a signal of the agents' type.

3.2 The Revelation Principle

The Revelation Principle states that the centralized mechanism is always at least as good as any other mechanism. Take any mechanism $\mathcal{M}$ that implements an allocation (production and transfers from the principal). Consider, now, a centralized mechanism in which agents are asked to report their type to the principal. If the principal commits to the allocation implemented by mechanism $\mathcal{M}$, no agent has any incentive in lying, since telling the truth earns agents the same payoff as it does with mechanism $\mathcal{M}$. If agents were playing an equilibrium strategy with mechanism $\mathcal{M}$, it must be an equilibrium to tell the truth in the centralized mechanism. Hence, a centralized mechanism can always replicate the outcome of any other mechanism. The Revelation Principle implies that centralization is always weakly preferred to any decentralized mechanism. With the assumptions underlying the Revelation Principle, it is impossible to analyze the internal structure of organization.

There are three important assumptions underlying the Revelation Principle that allow a centralized organization to replicate the outcome of any decentralized mechanism.

1. Communication between the principal and agents is infinitely costly after reports have been sent. Renegotiation is precluded, even though there may be benefits to renegotiating.

2. Agents act non-cooperatively. They cannot communicate and coordinate their reporting strategies. Without this assumption, agents may be tempted to collude to extract more benefits from the principal.

3. Agents can freely communicate all their private information to the principal. If there are limits to communication, the implemented allocation may not reflect the full richness of the agents' information.

When one or more assumptions are relaxed, a theory for optimal organizational design emerges. I relax, in turn, these assumptions underlying the Revelation Principle, and then compare the centralized and decentralized organizations.

4 Renegotiation

I now assume that the principal cannot commit not to renegotiate the initial contract after agents have reported their information. This may affect the efficiency of the centralized and decentralized organizations. In both settings, production is distorted to limit the agents' rents, rents that are necessary to induce truthful reporting. Once agents have reported their information, however, such distortions are not necessary anymore. The principal may want to eliminate them by making a renegotiation offer to the agents. Agents rationally anticipate the prospect of renegotiation, and this affects their incentive to report truthfully in the first place. The principal must then take this into account when offering the initial contract. I analyze the possibility for renegotiation in the two settings.[12]

• Centralization

There are potentially many instances where renegotiation can be introduced in the centralized game. Beaudry and Poitevin (1995b) show that the only stage where renegotiation can have an effect is after the agents have sent their message, that is, after stage 3. Before stage 3, renegotiation has no effect, since nothing happened after the initial contract was accepted. The mere fact of accepting the contract cannot induce renegotiation, since the principal can always make an initial offer that is expected to be accepted. The acceptance decision of the agents, therefore, has no informational content. After stage 3, however, agents have sent their report, and this has some informational value on which renegotiation may be based. This is then the "right" place to introduce the possibility for renegotiation.

The centralization game is modified to take into account the fact that the principal may attempt to renegotiate with the agents following their report.

1. Agent i learns its private information θ_i.

2. The principal offers a contract to both agents. If both accept, the game continues. If

[12]Renegotiation in games of incomplete information was first studied by Holmström and Myerson (1983) and Dewatripont (1988).

not, the game ends and each player receives its reservation utility.

3. The two agents simultaneously send a message to the principal.

3.1 The principal offers a new contract to both agents. If both accept, the game continues, and this new contract becomes the outstanding contract. If not, the initial contract remains in force.

4. The principal makes a production decision and pays the agents.

There is no loss of generality in considering initial contracts that are not renegotiated along the equilibrium path, that is, renegotiation-proof contracts. I therefore look for contracts that are not renegotiated in stage 3.1 along the equilibrium path.

Given the single-crossing condition embodied in the agents' cost function and the assumption of private values, it is easy to show that the equilibrium renegotiation-proof allocation must be separating. Any initial pooling contract is always renegotiated to a separating allocation. This implies that, along the equilibrium path, agents must report their type truthfully in stage 3 with the consequence that the principal learns precisely their type. In stage 3.1, the principal then renegotiates to an ex post efficient allocation (since incentive constraints are no longer relevant). Any renegotiation-proof contract must therefore include ex post efficient output levels.[13]

The renegotiation-proof optimal allocation is the solution to the following maximization problem.

$$\begin{aligned} \max_{w(\theta)} \quad & \mathrm{E}_{\theta}\left\{B(q^1(\theta)) - \sum_{i=1}^{2} w_i(\theta)\right\} \\ \text{s.t.} \quad & \text{for all } \theta, \tilde{\theta}_i,\; i, j = 1, 2, i \neq j : \\ & \mathrm{E}_{\theta_j}\left\{w_i(\theta) - \theta_i q^1(\theta)\right\} \geq 0 \\ & \mathrm{E}_{\theta_j}\left\{w_i(\theta_i, \theta_j) - \theta_i q^1(\theta_i, \theta_j)\right\} \geq \mathrm{E}_{\theta_j}\left\{w_i(\tilde{\theta}_i, \theta_j) - \theta_i q^1(\tilde{\theta}_i, \theta_j)\right\} \end{aligned}$$

[13] See Beaudry and Poitevin (1995b) for a formal proof of this argument.

where $q^1(\theta)$ represents the first-best optimal output vector. As before, the participation constraint of the inefficient agent and the incentive constraint of the efficient agent are binding. These can be used to solve for the optimal transfer payments.

$$\begin{aligned} U_i^{cr}\left(\bar{\theta}\right) &\equiv \nu\hat{w}_{i2}^{cr} + (1-\nu)\bar{w}^{cr} - \bar{\theta}\left(\nu\hat{q}^1 + (1-\nu)\bar{q}^1\right) = 0 \\ U_i^{cr}\left(\underline{\theta}\right) &\equiv \nu\underline{w}_i^{cr} + (1-\nu)\hat{w}_{i1}^{cr} - \underline{\theta}\left(\nu\underline{q}^1 + (1-\nu)\hat{q}^1\right) = \Delta\theta\left(\nu\hat{q}^1 + (1-\nu)\bar{q}^1\right) > 0 \end{aligned}$$

where the superscript "*cr*" stands for centralization-renegotiation. For the principal, the consequence of renegotiation is that larger informational rents are conceded to efficient agents to satisfy their interim incentive-compatibility constraints since output distortions cannot be used to limit rents. These rents amount to $\Delta\theta\left(\nu\hat{q}^1 + (1-\nu)\bar{q}^1\right)$, which are larger than second-best rents, $\Delta\theta\left(\nu\hat{q}^2 + (1-\nu)\bar{q}^2\right)$, where output was distorted to minimize them.

The presence of renegotiation supposes that any communication between the principal and the agents is costless. Such communication results in enhanced ex post efficiency of the organization, at the cost, however, of larger informational rents conceded to efficient agents. In an ex ante sense, the principal's payoff is reduced as the benefit of an efficient output vector is more than offset by the extra rents for the agents. It is in this sense that costless communication following the agents' reports can be detrimental to the ex ante efficiency of the centralized organization.

• Decentralization

The possibility for renegotiation also affects the design of the decentralized mechanism. Renegotiation may be introduced following the reports of agents 1 and 2, respectively. The decentralization game is then modified as follows.

1. Agent i learns its private information θ_i.

2. The principal offers a contract to agent 1. If it is accepted, the game continues. If not, the game ends and each player receives its reservation utility.

3. Agent 1 sends a message to the principal.

3.1 The principal offers a new contract to agent 1. If it is accepted, the game continues, and this new contract becomes the outstanding contract. If not, the initial contract remains in force.

4. Agent 1 offers a contract to agent 2. If it is accepted, the game continues. If it is rejected, the game ends and each player receives its reservation utility.

5. Agent 2 sends a message to agent 1.

5.1 Agent 1 offers a new contract to agent 2. If it is accepted, the game continues, and this new contract becomes the outstanding contract. If not, the initial contract remains in force.

6. Agent 1 makes a production decision and transfers take place.

As is the case under centralization, there is no loss of generality in considering renegotiation-proof contracts. In solving for this game, I therefore look for contracts that are not renegotiated in stages 3.1 and 5.1 along the equilibrium path.

The contract w_1^d is not renegotiation proof, since it requires agent 1 to report its type, and its remuneration schedule depends on this report. Renegotiation following a report $\bar{\theta}$ in stage 3.1 would lead to the elimination of the tax on output $\nu\Delta\theta q/(1-\nu)$. The only purpose of this tax is to introduce output distortions to limit agent 1's informational rents. It would be renegotiated away once agent 1 had reported its type.

Even though the contract w_1^d is not renegotiation proof, there exists another contract that is and that implements q^2. The principal needs to provide agent 1 with the appropriate incentives for having agent 2 producing efficiently and for producing q^2. Both goals can be achieved in a renegotiation-proof manner using a non-linear remuneration schedule for agent 1. Suppose agent 1 is paid according to a non-linear schedule that links transfers to delivered output. Agent 1 would then have no need to report its type in stage 3. This implies that no renegotiation can be successful at stage 3.1.[14]

[14] Beaudry and Poitevin (1995b) show why this is the case.

Similarly, the possibility for renegotiation in stage 5.1 changes the nature of the contract between agents 1 and 2. Suppose agent 1 offers to agent 2 a contract that prescribes to agent 2 send a report about its type to agent 1. Then, following the report, agent 1 can renegotiate to eliminate any distortion to output. This is obviously detrimental from an ex ante point of view, since agent 1 would have to pay agent 2 more rents. Agent 1 can alleviate the problem posed by the possibility of renegotiation by offering to agent 2 a non-linear price schedule that specifies a transfer as a function of the amount q that agent 2 delivers to agent 1. Such a contract avoids renegotiation since once output has been delivered, it cannot be modified, and hence renegotiated.[15] Only the transfer is left to be renegotiated, but this corresponds to a zero-sum game for which status quo is the only possible outcome.

Since agent 1 is responsible for remunerating agent 2, it has all the incentives for implementing q^2 in an efficient manner. In stage 4, agent 1 offers a contract to agent 2 that is designed to minimize the cost of implementing q^2. Agent 1 of type θ_1 offers agent 2 a contract $w_2^{dr}(q, \theta_1)$ with the following properties.

$$\begin{aligned}
&\frac{\partial w_2^{dr}}{\partial q}\left(\underline{q}^2, \underline{\theta}\right) = \frac{\partial w_2^{dr}}{\partial q}\left(\hat{q}^2, \bar{\theta}\right) = \underline{\theta} \\
&\frac{\partial w_2^{dr}}{\partial q}\left(\hat{q}^2, \underline{\theta}\right) = \frac{\partial w_2^{dr}}{\partial q}\left(\bar{q}^2, \bar{\theta}\right) = \bar{\theta} \\
&w_2^{dr}\left(\bar{q}^2, \bar{\theta}\right) - \bar{\theta}\bar{q}^2 = w_2^{dr}\left(\hat{q}^2, \underline{\theta}\right) - \bar{\theta}\hat{q}^2 = 0 \\
&w_2^{dr}\left(\hat{q}^2, \bar{\theta}\right) - \underline{\theta}\hat{q}^2 = \Delta\theta\bar{q}^2 > 0 \text{ and } w_2^{dr}\left(\underline{q}^2, \underline{\theta}\right) - \underline{\theta}\underline{q}^2 = \Delta\theta\hat{q}^2 > 0
\end{aligned}$$

where the superscript "*dr*" stands for decentralization-renegotiation. The function w_2^{dr} is also assumed to be concave in q so that second-order conditions are satisfied. This contract offered by agent 1 depends on its type θ_1. It therefore reveals to agent 2 agent 1's private information. This is inconsequential since the type of agent 1 does not enter into the payoff of agent 2. This contract is such that agent 2 maximizes its payoff $w_2^{dr}(q, \theta_1) - \theta_2 q$ at the

[15] There is an implicit assumption here that output is irreversible and inflexible. It cannot be increased or decreased once it has been produced. See Beaudry and Poitevin (1994) for a model in which renegotiation can occur as output is being produced. They show that renegotiation then has an effect.

output vector q^2. The level of rents is adjusted to ensure that the inefficient type of agent 2 earns zero, while the efficient type earns the minimal level of rents to satisfy its incentive constraint. The contract w_2^{dr} minimizes the cost to agent 1 of implementing q^2.

Since this contract does not require agent 2 to report its type, it is easy to see that it is renegotiation proof. The best agent 1 can do by renegotiating in stage 5.1 is to re-offer this same contract since it does not know yet the type of agent 2. Renegotiating after agent 2 has produced is too late because the amount q has already been delivered.

The principal has to offer a contract to agent 1 that gives incentives for implementing q^2, taking into account the fact that agent 2 is remunerated by agent 1. Consider a non-linear schedule w_1^{dr} specifying a wage for delivery of an output level. If agent 1 delivers $\underline{q}^2$, it is paid $\underline{w}_1^{dr}$; if it delivers $\hat{q}^2$, it is paid $\hat{w}_1^{dr}$; if it delivers $\bar{q}^2$, it is paid $\bar{w}_1^{dr}$. For any other output level, it is paid a large negative amount.

Once agent 1 has accepted the contract from the principal, it must decide which contract to offer agent 2. At this stage agent 1 knows its type. If it is type $\underline{\theta}$, it must offer agent 2 the contract $w_2^{dr}\left(q,\underline{\theta}\right)$ which implements the pair $(\underline{q}^2,\hat{q}^2)$. If it is type $\bar{\theta}$, it must offer agent 2 the contract $w_2^{dr}\left(q,\bar{\theta}\right)$ which implements the pair $(\hat{q}^2,\bar{q}^2)$. Agent 1 has the right incentives if the following incentive-compatibility constraint is satisfied for type $\underline{\theta}$.

$$\nu\left(\underline{w}_1^{dr}-2\underline{\theta}\underline{q}^2-\Delta\theta\hat{q}^2\right)+(1-\nu)\left(\hat{w}_1^{dr}-\left(\underline{\theta}+\bar{\theta}\right)\hat{q}^2\right)\geq$$
$$\nu\left(\hat{w}_1^{dr}-2\underline{\theta}\hat{q}^2-\Delta\theta\bar{q}^2\right)+(1-\nu)\left(\bar{w}_1^{dr}-\left(\underline{\theta}+\bar{\theta}\right)\bar{q}^2\right)$$

The principal minimizes the rents of agent 1 by satisfying the participation constraint of the inefficient type.

$$\nu\left(\hat{w}_1^{dr}-\left(\bar{\theta}+\underline{\theta}\right)\hat{q}^2-\Delta\theta\bar{q}^2\right)+(1-\nu)\left(\bar{w}_1^{dr}-2\bar{\theta}\bar{q}^2\right)\geq 0$$

In these two expressions, the remuneration of agent 2 has been added to agent 1's cost. For example, suppose agent 1 is efficient. With probability ν, agent 2 is efficient and must be paid $\underline{\theta}\underline{q}^2$, its cost of producing $\underline{q}^2$ plus informational rents, $\Delta\theta\hat{q}^2$.

The incentive constraint states that an efficient agent prefers to implement the pair $(\underline{q}^2, \hat{q}^2)$ rather than the pair $(\hat{q}^2, \bar{q}^2)$ taking into account the wage paid to agent 2. Agent 1 does not consider deviating to arbitrary output levels, since its wage is then a large negative amount. Note that agent 1 may be tempted to implement the pair $(\underline{q}^2, \bar{q}^2)$ instead of $(\underline{q}^2, \hat{q}^2)$ or $(\hat{q}^2, \bar{q}^2)$. This may be avoided by specifying a high enough value for $\hat{w}_1^{dr}$. This is always possible since there are two binding constraints that determine three wages.

When these two constraints are satisfied with equality, the incentive constraint of type $\bar{\theta}$ and the participation constraint of type $\underline{\theta}$ also are satisfied. Agent 1 then has all the incentives to offer the appropriate contract to agent 2, that is, to implement q^2.

The contracts w_1^{dr} and w_2^{dr} implement q^2 in a renegotiation-proof manner at a minimal cost for the principal. All players therefore earn the same expected payoff $U_i^2(\theta)$ as they do in the second-best environment without renegotiation.

When renegotiation is possible, some commitment can be restored by taking irreversible actions. This is exactly what a non-linear contract achieves. It eliminates communication on which renegotiation can be based. Communication, however, helps to coordinate on the desired output vector. With a non-linear contract, such coordination is achieved through the incentives it gives to the agents. The delivered output plays the same role as messages, since it can transmit the agents' private information. Transfers can then depend implicitly on the agents' types.

• Comparison of organizational structures

Renegotiation affects the implemented allocation under centralization but not under decentralization. Decentralization is therefore the preferred type of organization.

When the principal cannot commit not to renegotiate, decentralization is an efficient means of preventing such ex post opportunism. By delegating production to an informed agent and using non-linear remuneration schedules, some costly communication is eliminated. Furthermore, making agent 1 the residual claimant ensures that coordination requirements are fulfilled and that the second-best optimal allocation is implemented.

5 Collusion

Another assumption behind the Revelation Principle is that agents behave non-cooperatively when sending their reports to the principal (in the centralization game). This assumption is necessary in deriving the weak superiority of centralization over decentralization. I now assume that agents cannot commit not to communicate and coordinate their actions in the hope of extracting more rents from the principal. I do assume, however, that the principal can commit not to renegotiate.[16]

To make the analysis of collusion significant, Laffont and Martimort (1998) assume that q^2 must be implemented anonymously and in dominant strategies. These assumptions are necessary in our two-type environment for collusion to have any effect. They would not be necessary if agents had more types, since then the number of possible collusive coalitions would grow exponentially with the number of types. Collusion would then have an effect on the efficiency of the different structures without introducing these additional assumptions.

Anonymity implies that transfers depend only on the sum of types, not on the identity of the agents. This translates into $\hat{w}_1 = \hat{w}_2 = \hat{w}$. Dominant-strategy implementation implies that each agent must report truthfully regardless of the type of the other agent, that is, incentive constraints must hold *ex post.* Mookherjee and Reichelstein (1992) show that q^2 can be implemented in dominant strategies if agents are risk neutral and the output profile is decreasing in types. Both of these assumptions are satisfied in the current framework.

These assumptions imply that the following constraints are binding when q^2 is implemented.

$$\hat{w}_1 = \hat{w}_2 = \hat{w}$$

$$\underline{w} - \underline{\theta}\underline{q}^2 \geq \hat{w} - \underline{\theta}\hat{q}^2$$

$$\hat{w} - \underline{\theta}\hat{q}^2 \geq \bar{w} - \underline{\theta}\bar{q}^2$$

[16] The analysis of this section follows that of Laffont and Martimort (1998). For a more detailed survey of collusion, the reader is referred to Laffont and Rochet (1997).

$$\nu(\hat{w} - \bar{\theta}\hat{q}^2) + (1 - \nu)(\bar{w} - \bar{\theta}\bar{q}^2) \geq 0$$

The first constraint is the anonymity constraint. The next two constraints are the ex post incentive constraints of the efficient agent. The last constraint is the interim participation constraint of the inefficient agent. Because agents are risk neutral (and their payoffs depend only on expected transfers), there are degrees of freedom in setting transfers in the second-best problem. There were two binding constraints and four transfers. Anonymity and dominant-strategy implementation add two extra constraints that can be satisfied at no extra cost by the four transfers.

- CENTRALIZATION

Laffont and Martimort (1998) show that the centralized mechanism can implement q^2 in dominant strategies using anonymous transfers.

With this implementation, there is a benefit to collusion. Suppose agents have different types. The efficient agent's incentive constraint is strictly binding, that is, $\hat{w} - \underline{\theta}\hat{q}^2 = \bar{w} - \underline{\theta}\bar{q}^2$; that of the inefficient agent is not, that is, $\hat{w} - \bar{\theta}\hat{q}^2 > \underline{w} - \bar{\theta}\underline{q}^2$. Furthermore, it can be shown that the inefficient agent gets more when both agents are inefficient than when the other is efficient; that is,

$$\hat{w} - \bar{\theta}\hat{q}^2 = (1 - \nu)\Delta\theta(\bar{q}^2 - \hat{q}^2) < 0 < \nu\Delta\theta(\hat{q}^2 - \bar{q}^2) = \bar{w} - \bar{\theta}q^2.$$

This implies that the efficient agent, when reporting truthfully, creates a negative externality on the inefficient agent. If it reports truthfully, the inefficient agent gets $(1-\nu)\Delta\theta(\bar{q}^2 - \hat{q}^2) < 0$. If the efficient agent lies, the inefficient agent gets $\nu\Delta\theta(\hat{q}^2 - \bar{q}^2) > 0$. Since the efficient agent is indifferent between lying or telling the truth, the gain to colluding is

$$-(1 - \nu)\Delta\theta(\bar{q}^2 - \hat{q}^2) + \nu\Delta\theta(\hat{q}^2 - \bar{q}^2) = \Delta\theta(\hat{q}^2 - \bar{q}^2) > 0.$$

It is proportional to the difference in outputs between the two worst states. If agents can collude, they are likely to internalize this externality, thus destroying some of the ex ante incentives built in the second-best contract. The principal has to prevent such collusion by modifying the original contract.

Laffont and Martimort (1998) introduce the possibility of collusion in the centralization game as follows.

0. Agents can enter into a collusive agreement.

1. Agent i learns its private information θ_i.

2. The principal offers a contract to both agents. If both accept, the game continues. If not, the game ends and each player receives its reservation utility.

2.1 Agent i observes the type of agent j.

2.2 The two agents can collude to coordinate their reporting strategies. If collusion fails, agents report non-cooperatively to the principal.

3. If the collusion succeeds, reports are coordinated and sent to the principal. Collusive bribes are paid.

4. The principal makes a production decision and pays the agent.

The formulation for this game deserves some comments. First, agents are allowed to enter into an ex ante collusive agreement in which they promise to share all transfers equally. This is a technical assumption to ensure that transfers are anonymous, that is, that they depend only on aggregate types, not on individual types. In a two-type world, this assumption is necessary for collusion to have non-trivial effects. With more types or agents, this assumption would not be necessary since the number of collusion-proof constraints grows exponentially. Second, collusion also takes place after the principal has offered the contract and agents have observed their type. Third, collusion takes place under symmetric information, again to simplify the exposition. It is not necessary for the "qualitative" results to take place, as is shown in Laffont and Martimort (1998). Finally, it is assumed that agents must respect the collusive agreement.[17]

As with renegotiation, collusion-proof contracts are characterized, that is, contracts that are robust to the possibility for collusion by the two agents. The collusion-proof optimal allocation solves the following maximization problem.

$$\max_{q(\theta),w(\theta)} \quad \mathrm{E}_\theta \left\{ B(q(\theta)) - \sum_{i=1}^{2} w_i(\theta) \right\}$$

$$\text{s.t.} \quad \text{for all } \theta, \tilde{\theta}_i, \tilde{\theta}_j \ i, j = 1, 2, i \neq j :$$

$$\mathrm{E}_{\theta_j} \{ w_i(\theta) - \theta_i q(\theta) \} \geq 0$$

$$w_i(\theta_i, \theta_j) - \theta_i q(\theta_i, \theta_j) \geq w_i(\tilde{\theta}_i, \theta_j) - \theta_i q(\tilde{\theta}_i, \theta_j)$$

$$w_i(\theta_i, \theta_j) - \theta_i q(\theta_i, \theta_j) + w_j(\theta_i, \theta_j) - \theta_j q(\theta_i, \theta_j) \geq$$
$$w_i(\tilde{\theta}_i, \tilde{\theta}_j) - \theta_i q(\tilde{\theta}_i, \tilde{\theta}_j) + w_j(\tilde{\theta}_i, \tilde{\theta}_j) - \theta_j q(\tilde{\theta}_i, \tilde{\theta}_j)$$

The first set of constraints represents the interim participation constraints; the second set represents the ex post incentive constraints required for dominant-strategy implementation; and the third set represents the collusion-proof constraints that prohibit collective deviations from truth telling.

The challenge is to identify which constraints are binding at the optimal solution. Laffont and Martimort (1998) show that the following constraints are binding.

$$2(\underline{w} - \underline{\theta}\underline{q}) \geq 2(\hat{w} - \underline{\theta}\hat{q})$$

$$2\hat{w} - (\underline{\theta} + \bar{\theta})\hat{q} \geq 2\bar{w} - (\underline{\theta} + \bar{\theta})\bar{q}$$

$$\nu(\hat{w} - \bar{\theta}\hat{q}) + (1 - \nu)(\bar{w} - \bar{\theta}\bar{q}) \geq 0$$

There are two collective incentive constraints: one for the two efficient agents who prefer to tell the truth rather than pretend to be of different types;[18] one for agents of different

[17] Martimort (1997) relaxes this assumption in a dynamic context. Abiding with the collusive agreement is based on reciprocity.

[18] This constraint is also an individual incentive constraint for an efficient agent.

types who prefer to tell the truth rather than pretend they are both inefficient. The final constraint is the usual interim participation constraint of the inefficient type.

As shown above, for the second-best optimal allocation, the benefit to collusion is $\Delta\theta(\hat{q}^2 - \bar{q}^2)$, while the cost of lying for the efficient agent is nil, since its incentive constraint strictly binds. There are two ways of reducing collusion. The first is to decrease the benefit to collusion reducing the difference between $\hat{q}$ and $\bar{q}$; the other one is to increase the cost of lying for the efficient agent by relaxing its incentive constraint. At the optimum, the principal uses both: it reduces the gap in the two outputs and relaxes the efficient agent's incentive constraint (when the other agent is inefficient) by increasing the transfer $\hat{w}$.

Formally, the second binding constraint is used to relax the individual incentive constraint of the efficient type (when the other agent is inefficient). The efficient agent is no longer indifferent to telling the truth or lying, thus imposing a cost on colluding. Even though the benefit of collusion is still positive, it becomes costly for the efficient agent to lie, thus erasing any collective benefit lying may provide.

The solution has the following implicitly defined output schedule.

$$
\begin{aligned}
B'(\underline{q}^{cc}) &= 2\underline{\theta} \\
B'(\hat{q}^{cc}) &= \underline{\theta} + \bar{\theta} + \frac{(1+\nu)\Delta\theta}{2(1-\nu)} \\
B'(\bar{q}^{cc}) &= 2\bar{\theta} + \frac{\nu\Delta\theta}{1-\nu}
\end{aligned}
$$

where the superscript "*cc*" stands for centralization-collusion. Output is ex post efficient in the lowest-cost state, while it is distorted in the other states. Compared with the output vector q^2, the collusion-proof output profile is flattened; that is, $\hat{q}^{cc} - \bar{q}^{cc} < \hat{q}^2 - \bar{q}^2$.[19] A flatter output profile reduces the benefit to collusion, since output is not as sensitive to revealed information.

[19] A similar result was shown by Tirole (1986) in a model of collusion in a hierarchy.

The equilibrium expected utility for each type of agent is

$$
\begin{aligned}
U_i^{cc}\left(\bar{\theta}\right) &\equiv 0 \\
U_i^{cc}\left(\underline{\theta}\right) &\equiv \Delta\theta\left(\nu\hat{q}^{cc}+(1-\nu)\bar{q}^{cc}\right)+\frac{1-\nu}{2}\Delta\theta\left(\hat{q}^{cc}-\bar{q}^{cc}\right).
\end{aligned}
$$

The principal concedes the efficient agent extra rents to ensure that collusion is prevented. This corresponds to the last term in the above expression. It is exactly equal to the expected share of collusion benefits for the efficient agent.

For the principal, the consequence of collusion is that outputs are distorted further away from second-best optimal outputs and that extra rents are conceded to the efficient agent. It is then clear that the possibility of collusion reduces the principal's payoff in the centralized structure.

• DECENTRALIZATION

Consider the decentralization game described in Section 3.1. There are no benefits to collusion, because agent 1 reports to the principal before contracting with agent 2, that is, before having the opportunity to collude with agent 2. Agents cannot coordinate their reports to exploit the principal. Furthermore, since agent 1 is a residual claimant, it must support the full cost of rents conceded to agent 2. There is no room for exploiting the principal. The second-best optimal contract is then collusion proof.

The implementation of q^2 does not use anonymous transfers since an asymmetry between the agents has been introduced by the principal in the actual design of the decentralization game. This may facilitate the implementation of q^2 even under the threat of collusion. This result, however, is not an artefact of the precise decentralization game chosen. Laffont and Martimort (1998) show that if agent 1 contracts or colludes with agent 2 before communicating with the principal, the output vector q^2 can still be implemented. Furthermore, it can be implemented using anonymous transfers.

In a decentralized framework, the principal makes agent 1 the residual claimant. This effectively introduces a conflict of interest between the two agents that prohibits any successful collusion. The vector q^2 can consequently be implemented at no extra cost under the threat of collusion.

• Comparison of organizational structures

Collusion reduces the effectiveness of the centralized structure, while it does not affect that of the decentralized one. Decentralization is the preferred mode of organization when agents can collude. A principal should therefore attempt to delegate decision-making whenever possible to avoid collusion.

Under decentralization, collusion, by the design of the game, occurs under asymmetric information, while it occurs under symmetric information in the centralized setting. This makes collusive agreements harder to implement under decentralization. This asymmetry is not the decisive factor, however, in the superiority of the decentralized organization. Laffont and Martimort (1998) show that collusion affects the centralized structure even under asymmetric information, while it does not affect the decentralized structure. So the same result also would hold when agents collude under asymmetric information.

6 Limits on communication

The final modification I would like to consider is that of limits on communication. So far, it has been assumed that any information could costlessly be reported by the agents to the principal and could costlessly be used by the principal in the design of the contract. Information, however, is something that is hard to transmit precisely, and/or hard to describe precisely. Agents may not be able to report all the information they possess simply because the language of communication is not rich enough to transmit all the complexity of their knowledge; or the principal may not be able to include in the contract all contingencies because the information is hard to describe formally.

There are, therefore, two ways of incorporating limits on communication in a formal analysis of organizational design. First, one can limit communication but allow arbitrary contracts with an unlimited number of contingencies. Communication is formally limited by restricting the dimension of the agents' message space. This was first studied formally in a model of incentives by Green and Laffont (1982, 1986) and Melumad, Mookherjee, and Reichelstein (1992). Second, one can limit the number of contingencies that are included in the contract. This was first studied formally in a model of incentives by Melumad, Mookherjee, and Reichelstein (1997), where the agents' message space and the number of contingencies that can be incorporated into the contract have the same dimension, which is smaller than the space of relevant information.

For illustrative purposes, I present here the first approach, where agents communicate using a space of messages that is of a smaller dimension than their private information, but where they can write arbitrarily elaborated contracts. In a model with a continuum of types, Melumad, Mookherjee, and Reichelstein (1992, 1997) introduce limits to communication by assuming that the space of messages is a finite partition of the space of types. The same basic intuition can be captured in our simpler model: in a two-type world, this translates into assuming that agents cannot communicate at all.

• Centralization

Under centralization, the game is modified by removing stage 3 at which the agents were communicating their private information. The following game is played.

1. Agent i learns its private information θ_i.
2. The principal offers a contract to both agents. If both accept, the game continues. If not, the game ends and each player receives its reservation utility.
3. The principal makes a production decision and pays the agent.

The principal must make its production decision without any relevant information from the agents. A full pooling solution ensues. This solution must satisfy interim participation constraints for the two types. It is the solution to the following maximization problem.

$$\begin{aligned} \max_{q,w} \quad & B(q) - \sum_{i=1}^{2} w_i \\ \text{s.t.} \quad & w_i - \theta_i q \geq 0 \ \text{ for all } \theta_i, i = 1, 2 \end{aligned}$$

The binding constraint is that of the inefficient type. Production is then set at the ex post efficient level for the worst-case scenario; that is, $q^{cl} = \bar{q}^1$. The wage is $w^{cl} = \bar{\theta}\bar{q}^1$. The expected payoff to each agent is

$$\begin{aligned} U_i^{cl}\left(\bar{\theta}\right) &\equiv 0 \\ U_i^{cl}\left(\underline{\theta}\right) &\equiv \Delta\theta\bar{q}^1. \end{aligned}$$

Thus the efficient agent earns some rents, since it produces the same amount as the inefficient agent and is being paid the same wage while being more efficient. The principal earns $B(\bar{q}^1) - 2\bar{\theta}\bar{q}^1$.

Under centralization, limits on communication severely restrain the efficiency of the organization. The principal retains all production rights, but must produce without knowing anything about the agents' productivity. The pooling solution that ensues reflects a complete lack of coordination between the principal and the agents. In a more general model, the principal may be able to do slightly better. For example, in Melumad, Mookherjee, and Reichelstein (1992, 1997), there is a continuum of types but agents can send messages only from a finite partition of the type space. The principal can extract some information before making its production decision. The efficiency of the organization is still reduced by the limits imposed on communication.

• Decentralization

Under decentralization, again all communication is eliminated. The following game is played.

1. Agent i learns its private information θ_i.

2. The principal offers a contract to agent 1. If it is accepted, the game continues. If not, the game ends and each player receives its reservation utility.

3. Agent 1 offers a contract to agent 2. If it is accepted, the game continues. If it is rejected, the game ends and each player receives its reservation utility.

4. Agent 2 makes a production decision.

5. Agent 1 produces, and transfers take place.

Referring to the original game, stage 5 is replaced by agent 2's production decision, while stage 3, where agent 1 reports to the principal is eliminated.

Consider the contracts w_1^{dr} and w_2^{dr} that implement q^2 under the threat of renegotiation. All communication is eliminated by the use of non-linear contracts to avoid the costs associated with renegotiation. These same contracts are feasible when all communication is prohibited, and the vector q^2 can therefore be implemented. Limits to communication do not affect the efficiency of the decentralized structure.

There are two reasons why decentralization is an efficient means for avoiding costly limits to communication. First, the contract offer from agent 1 to agent 2 can incorporate some private information of agent 1. Agent 1 can offer a contract that induces agent 2 in taking into account agent 1's information in making its production decision. Second, production rights are delegated to agent 1, who can delegate them to agent 2. Delegation of production rights allows informed agents to use their information in making their production decision.

• Comparison of organizational structures

Again, decentralization is preferred to centralization by the principal. The superiority of decentralization stems from the fact that production and contracting rights are decentralized to informed agents. This makes possible the use of non-linear transfer schedules that do not require any formal communication. This is particularly useful when no communication can occur between players.

I have assumed that, even though no communication is allowed, contracts can still include many contingencies. For example, a contract based on output implicitly includes as many contingencies as there are possible output levels. The number of contingencies is therefore greater than the number of allowed messages (which is zero here). Although, this assumption is necessary here to derive the result, it would not be necessary in a more general setting. Melumad, Mookherjee, and Reichelstein (1997) show that restricting the number of messages and contingencies to be equal yields the same result of the superiority of decentralization over centralization as long as this number is greater than one. It is possible to have more than one message in their framework since they assume a continuum of types. The intuition behind their result is similar as the one above. Even though agent 1 is restricted in the number of contingencies that can be included in the contract offered to agent 2, the design of the contract itself can incorporate the information of agent 1 more finely than communication allows under centralization. This is sufficient for the superiority of decentralization.

7 Discussion

I would like to discuss the role of certain assumptions underlying the analysis. The model and assumptions were rigged to show the superiority of decentralization over centralization. This was done in order to get away from the Revelation Principle, which predicts the superiority of centralization. Doing so has shown that centralization is completely dominated by decentralization. No tradeoff between the two structures seems to exist when one moves away from the assumptions underlying the Revelation Principle. In reality, however, we do observe some decentralized as well as centralized economic organizations. There is an "empirical" tradeoff, and this tradeoff potentially can be explained by the theory.

• Bilateral private information

The superiority of decentralization is due to the fact that the principal delegates production rights to agents with superior information. Suppose now that the principal also possesses some relevant private information. Complete delegation to agents may then entail an efficiency loss owing to the fact that the agents would be ignorant of the principal's information when making their production decision. In cases where the principal's information is crucial for the efficiency of the organization, centralization may be the preferred organizational structure. Poitevin (1997) examines these issues in a model with an informed principal and an agent when the principal cannot commit not to renegotiate the original contract. Whether decentralization is optimal or not effectively depends on the importance of each player's private information. In general, production rights should be assigned to the player who has the most crucial private information.

- Risk aversion

The effectiveness of the decentralized mechanism in alleviating communication problems depends crucially on the ability of the principal to give agent 1 a high-powered incentive scheme, where agent 1's remuneration is sensitive to the information of the two agents. Delegation of production and contracting rights to agent 1 creates a moral-hazard problem, since agent 1's preferences are not perfectly aligned with those of the principal.[20] The principal eliminates this moral-hazard problem by giving agent 1 a high-powered incentive scheme. There is no control loss in delegating production and contracting rights to agent 1. The cost of delegation is completely eliminated. This resolution is possible when agent 1 is risk neutral. With risk aversion, there would be a tradeoff between delegating production and contracting rights to reduce communication costs and insurance to be provided to agent 1. It may well be the case that insurance can be better provided by the centralized mechanism, while the decentralized one alleviates communication problems. A tradeoff may emerge, depending on parameter values. A complete analysis of the role of risk aversion in the presence of communication costs remains to be done.[21]

[20]This problem is discussed at length in Melumad, Mookherjee, and Reichelstein (1995).

[21]Faure-Grimaud, Laffont, and Martimort (1999) study the efficiency of a decentralized organization when agent 1 is risk averse. They show that the moral-hazard problem cannot be completely eliminated. They do not, however, compare organizational structures.

• Dynamics

In this article, the negotiation of a new contract or of a collusive agreement takes a very simple form. If negotiation fails, the status quo stands. This is a very strong assumption. One may easily imagine that if a contract offer is rejected, negotiation can still continue. The appropriate way of modeling such possibility is to introduce an extensive form explicitly designed to take into account these ongoing negotiations. This is accomplished in Beaudry and Poitevin (1993, 1995a) in a model of renegotiation and in Martimort (1997) in a model of collusion. The drawbacks of explicitly modeling negotiations are that the results may depend on the precise dynamic extensive form one adopts and that the theory may lack predictive power if there is a multiplicity of equilibria. Using a simpler extensive form, as is done here, provides tractable solutions and sharp predictions (although this may be an implicit way of selecting among equilibria).

• Renegotiation and collusion

It would also be interesting to simultaneously introduce the possibility for renegotiation and collusion in the centralized mechanism. The prospect of renegotiation is alleviated by eliminating all output distortions and giving more informational rents to the agents. The prospect of collusion is alleviated by introducing more distortions to reduce collusion benefits. These respective solutions may be in conflict. It would be interesting to see how the two problems can be reconciled in the same model. This has not been analyzed yet.

• More agents

A natural question to ask is whether these results are robust to the introduction of more agents. The answer to this question depends on how the decentralized solution is affected by the presence of more agents. It seems that a sequence of non-linear contracts could still implement the second-best optimal production.[22]

Suppose there are three agents. Agent 1 would offer to agent 2 (who contracts with agent

[22] Mookherjee and Reichelstein (1997) show the equivalence of centralization and decentralization when the principal can monitor the contracting activities of the agents.

3) a non-linear contract that reflects its information and takes into account the fact that agent 2 must pay informational rents to agent 3. The principal would offer a contract to agent 1 that provides the right incentives. With three agents, there would be eight different states of nature, which leaves plenty of flexibility in setting the different transfers and giving incentives to agent 1.

If this conjecture is right, this implies that q^2 could still be implemented under the threat of renegotiation or limits to communication.

With collusion, however, the resolution is not as clear. The number of collusion-proof constraints grows large when the number of agents increases. Whether this affects centralization more or less than decentralization is an open question.

• ROBUSTNESS TO THE EXTENSIVE FORM

An interesting avenue of research would be to see if properties such as collusion-proofness and renegotiation-proofness can be made robust to the choice of the specific extensive form. So far, results often vary from one game to the other: for example, whether renegotiation or collusion occurs under asymmetric information or not, the identity of the player making the offer, whether informed or not, usually matters for the equilibrium.

Can the properties of collusion- and renegotiation-proofness be described by a set of constraints as the property of incentive compatibility can? A great contribution of the Revelation Principle is that it allows a simple characterization of the set of implementable allocations under asymmetric information by a set of constraints, namely, incentive-compatible constraints. A similar approach has not been developed yet when players can renegotiate or collude. The difficulty of doing so stems from the sensitivity of the results to the specifics of the bargaining framework. Laffont and Martimort (1997a) have proposed an approach in which a dummy uninformed third party is assumed to make collusion offers to all agents involved. They characterize all allocations that can be attained through collusion and use this characterization to define initial contracts that are collusion proof. This approach looks promising. It has not been applied yet to renegotiation problems, and it is surely a fruitful avenue to explore in future research.

8 Conclusion

In this paper I address the central question of whether the theory of incentives can explain decentralization or not. The Revelation Principle seems to suggest that it cannot. Recent work along these lines, however, shows that it is possible to explain decentralization when one relaxes some crucial assumptions underlying the Revelation Principle. Three avenues are explored here.

When the principal cannot commit not to renegotiate the initial agreement following communication by informed agents, decentralization of decision-making to these informed agents may be optimal. It limits the amount of communication on which detrimental renegotiation is based.

When agents can collude and coordinate their communication strategies to extract more rents from the principal, decentralization of decision-making may be optimal. If an agent becomes a profit center, it wants to limit the rents of the other agent. This introduces a conflict of interest between the two agents that can prevent any collusion.

When communication is costly, again decentralization of decision-making to informed agents may be optimal since it reduces communication costs.

In all these cases, decentralization may dominate centralization. The drawback of decentralization is that it may entail a loss of control on the part of the principal. This loss has to be assessed against the benefits of decentralization.

The models surveyed here help to endogenize the nature of communication costs and its implications for organization theory in a theory based on incentives. It is as if these models are providing foundations for earlier studies where communication costs are exogenously assumed and incentives are ignored. The requirement to reveal private information generates communication costs. Without this requirement, there is no need for communication and hence no communication costs. The consequence of satisfying incentive constraints is a loss in efficiency and rents conceded to agents. Communication can then be thought of in terms of inefficiencies and rents.

Finally, it is instructive to note that the analysis of organizational structure parallels the earlier debate about centralization and decentralization of economic systems. The conclusion of this debate is that decentralization is more efficient at incorporating agents' information through the pricing mechanism than is a system based on orders from a centralized agent. The main difference between this literature and organization theory is that the analysis of organizations usually involves small number of agents, implying that agents evolve in a strategic environment where they can earn rents from their private information. This is why a centralized organization may be optimal under some circumstances.

References

Arrow, K.J. (1975) "Vertical Integration and Communication," *Bell Journal of Economics* **6**, 173–183.

Beaudry, P. and M. Poitevin (1993) "Signalling and Renegotiation in Contractual Relationships," *Econometrica* **61**, 745–782.

——(1994) "The Commitment Value of Contracts under Dynamic Renegotiation," *Rand Journal of Economics* **25**, 501–517.

——(1995a) "Competitive Screening in Financial Markets when Borrowers can Recontract," *Review of Economic Studies* **62**, 401–423.

——(1995b) "Contract Renegotiation: A Simple Framework and Implications for Organization Theory," *Canadian Journal of Economics* **28**, 302–335.

Dewatripont, M. (1988) "Commitment through Renegotiation-Proof Contracts with Third Parties," *Review of Economic Studies* **45**, 377–390.

Faure-Grimaud, A., J.J. Laffont, and D. Martimort (1999) "The Endogenous Transaction Costs of Delegated Auditing," *European Economic Review* **43**, 1039–1048.

Gibbard, A. (1973) "Manipulation of Voting Schemes," *Econometrica* **41**, 587–601.

Green, J. and J.J. Laffont (1977) "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica* **45**, 427–438.

——(1982) "Limited Communication and Incentive Constraints," in Groves, T., R. Radner, and S. Reiter, eds., *Information, Incentives and Economic Mechanisms*, pages 308–329. University of Minnesota Press.

——(1986) "Incentive Theory with Data Compressing," in Heller, W., R. Starr, and R. Radner, eds., *Uncertainty, Information and Communication*, pages 239–253. University of Minnesota Press.

Grossman, S.J. and O.D. Hart (1986) "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy* **94**, 691–719.

Groves, T. and R. Radner (1972) "Allocation of Resources in a Team," *Journal of Economic Theory* **4**, 415–441.

Holmström, B. and R.B. Myerson (1983) "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica* **51**, 1799–1819.

Hurwicz, L. (1969) "Centralization and Decentralization in Economic Systems," *American Economic Review* **59**, 513–524.

Laffont, J.-J. and D. Martimort (1997a) "Collusion Under Asymmetric Information," *Econometrica* **65**, 875–911.

——(1997b) "The Firm as a Multicontract Organization," *Journal of Economics and Management Strategy* **6**, 201–234.

——(1998) "Collusion and Delegation," *Rand Journal of Economics* **29**, 280–305.

Laffont, J.-J. and J.J. Rochet (1997) "Collusion in Organizations," *Scandinavian Journal of Economics* **99**, 485–495.

Marschak, T. (1959) "Centralization and Decentralization in Economic Organizations," *Econometrica* **27**, 399–430.

Martimort, D. (1997) "A Theory of Bureaucratization Based on Reciprocity and Collusive Behavior," *Scandinavian Journal of Economics* **99**, 555–579.

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

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

——(1995) "Hierarchical Decentralization of Incentive Contracts," *Rand Journal of Economics* **26**, 654–672.

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

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

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

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

Poitevin, M. (1997) "Contract Renegotiation and Organizational Design," Mimeo, Université de Montréal.

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

Williamson, O.E. (1975) *Markets and Hierarchies*. Free Press.