ANNALS OF ECONOMICS AND FINANCE **1**, 231–263 (2000)

A Theory of Supervision with Endogenous Transaction Costs*

Antoine Faure-Grimaud

Institute of Management and FMG at the London School of Economics, and CEPR

Jean-Jacques Laffont

IDEI, GREMAQ, Institut Universitaire de France, Toulouse I, France

and

David Martimort

IDEI, GREMAQ, Toulouse, France

We propose a theory of supervision with endogenous transaction costs. A principal delegates part of his authority to a supervisor who can acquire soft information about an agent's productivity. If the supervisor were risk-neutral, the principal would simply make the better informed supervisor residual claimant for the hierarchy's profit. Under risk-aversion, the optimal contract trades-off the supervisor's incentives to reveal his information with an insurance motive. This contract can be identified with the one obtained in a simple hard information model of hierarchical collusion with exogenous transaction costs. Now, transaction costs are endogenous and depend on the collusion stake, the accuracy of the supervisory technology and the supervisor's degree of risk-aversion. We then discuss various implications of the model for the design and management of organizations.

Key Words: Supervision; Soft information; Collusion; Endogenous transaction costs.

JEL Classification Numbers: D82, G14, G32, L51.

1. INTRODUCTION

Since the seminal work of Barnard (1938) *"The Functions of the Executive,"* the firm has been seen as a cooperative entreprise in which each member *"has accepted a position of contact with others similarly associated*

* We thank Jean Tirole and Jeon Doh-Shin for helpful comments. All errors are ours.

1529-7373/2000

[..]" and *"from this contact, there must arise social interactions."* Barnard also pushes forcefully the view that inducements and incentives within the firm are both monetary and non-monetary. According to this perspective, organizations are not only formal in nature but also the locus where informal groups form and evolve. Even if in Barnard's theory, the formation of informal groups "oils the wheels" of the organization, other scholars in the industrial sociology literature quickly recognized that informal groups may also imped the achievement of firm efficiency.

For Roethlisberger and Dickson (1947), informal groups emerge within the firm to support systematic violations of what can be considered as a "fair day" output, or more generally to pursue disruptive strategies like cost overruns or refusal of adopting new technologies. Dalton (1959) and Crozier (1963) have both discussed the formation of the so-called vertical cliques within the firm and their roles in promoting objectives different from those of the top manager or owners of the firm. In his study of a chemical plant, *The Milo Fractioning Center*, Dalton argues that those vertical cliques between supervisors and low-level managers which appear both at the divisional and at the departmental levels are the *real centers of power*. Crozier has also stressed the importance of these supervisor-supervisees relationships for the internal efficiency of the firm. Echoing Dalton's distinction among several archetypical vertical cliques according to their ability to pursue their objectives,[1] he stresses the determinants of the cohesion of an informal group and in particular the amount of information shared within this coalition. He even calls for a *"fruitful study of the allocation of roles and formal rules within the organization"* which result from the collusive behavior of these subgroups.

The lessons from these sociological studies are clear: Informal groups matter and firms follow different behavioral rules depending on the efficiency of the side-contracting deals sustaining these coalitional behaviors. To get a rich theory of organizational design based on the linkage between formal and informal incentives as it was suggested by Barnard, we need therefore to build a theory of collusive behavior within the firm which derives endogenously, and does not assume a priori, the transaction costs associated with the side-contracts among collusive agents. This theory should also fill the gap between the formal approach of group behavior defended by Barnard and the informal approach suggested by the Human Relations School.[2]

[1] Dalton distinguishes the vertical *symbiotic* cliques which include a reciprocal exchange of favors between the supervisor and the supervisee and the *parasitic* cliques which are unequal in nature but such that partners know each other quite well, like family members or managers having graduated in the same school for instance.

[2] Roethlisberger and Dickson (1947).

This theory should link the efficiency of the internal side-contracting to various parameters of the organizational environment. Following the sociological studies of the Human Relations School, the ultimate goal of such a theory should be to give an economic content to the idea that organizational parameters affect the set of social networks which establish within the firm and have thereby a concrete impact on the firm's productivity. The present paper aims at providing some first steps towards such a theory.

Thereafter, we build a hierarchical model of a large firm in which there is separation of ownership, supervision and control of the productive assets. This can be seen as an highly centralized, functionally departmentalized and bureaucratic organization of the kind extensively described by Chandler (1962). A risk-neutral principal, top-manager or owner delegates some of his formal authority to a risk-averse supervisor.[3] This supervisor has the right to control the wage payments of workers standing at the bottom of the hierarchy as well as their tasks, their work conditions, their relationships with other workers, etc... The firm is thus seen as a nexus of formal contracts:[4] first, a grand-contract between the top management and the supervisor and second, a side-contract between this supervisor and his supervisee.

As there is separation between ownership and supervision, only the supervisor gets some (private) information correlated with the workers' performances and that information could be useful to improve the latter's incentive schemes.[5] Just as the private information learned by workers on their own efficiency in performing their jobs, these noisy signals observed by the supervisor are *soft information*, i.e., they can be fully manipulated. Hence, contracts within the firm aim at inducing revelation of all the decentralized information which exists along the hierarchy. The side-contract induces revelation of the agent's information to the supervisor. The principal designs the grand-contract so that it induces revelation of the supervisor's total information, namely his own private signal and the information on the worker he has learned through side-contracting.

A crucial point of our analysis is that providing incentives to report his signal to a risk-averse supervisor is costly for the principal. The logic here is extremely close to the standard moral hazard argument developed in Mirrlees (1976), Holmstrom (1979), Shavell (1979) and Grossman and

[3]Galambos (1995) discusses how changes in the size and the complexity of the firms' operations have been accompanied by an increase in the scope of responsabilities of CEOs who received more delegated powers from the owners and other stakeholders of the firms.

[4]See Cyert and March (1963), Jensen and Meckling (1976) and Laffont and Martimort (1997a) for discussions of the multilateral nature of contracting within the firm.

[5]It is by now a well-known result that correlated information helps solve agency problems (see Holmstrom (1979)'s informativeness results in a moral hazard setting and Crémer and McLean (1988) in adverse selection models).

Hart (1983). Because of asymmetric information between the supervisor and the agent he supervises, the informed supervisor faces some risks: his payoff from side-contracting is indeed a lottery depending on the agent's realized type. The grand-contract must therefore provide some insurance to this supervisor in order to induce him to exert authority over the agent. However, the exact details of the authority relationship between the supervisor and the workers, i.e., their side-contract, remain unknown to the principal. The side-contract which stipulates the functioning of the two lower tiers of the hierarchy becomes an unverifiable variable from the point of view of the principal. There is moral hazard in the choice of this variable when the supervisor pursues an objective different from that of the owner. When there is no possibility for making the supervisor residual claimant for the firm's profit, a conflict between insurance and incentives appears. This creates a new agency cost which adds up to the agency cost affecting the supervisor-supervisee relationship.

This paper provides a clear description of this new agency cost of delegation.[6] This new agency cost can be interpreted as the insurance risk-premium that the principal must bear when satisfying the *coalition incentive constraints* imposed by the signing of the side-contract between the risk-averse supervisor and his supervisee. To reduce this agency cost, the optimal grand-contract calls for some output distortions. Output allocations are made less sensitive to the manipulable information of the supervisor. Flatter incentives and more bureaucratic rules are implemented when the supervisor is more risk-averse. In the limiting case of an infinite degree of risk-aversion, no use of the supervisor's information can be made and the optimal mechanism calls for much bunching. Supervision becomes useless.

Similarly, when the supervisor's information becomes uncorrelated with the agent's type, supervision is not needed. More generally, even with delegated authority, a similar logic to that of Holmstrom (1979)'s theorem on informative signals applies: information obtained through supervision is useful to the principal as long as it is correlated with the agent's unknown productivity.

We then interpret these new agency costs in light of the theory of collusion in hierarchical structures put forward by Tirole (1986) and (1992) and used in Laffont and Tirole (1993, Chapter 11 and following), Laffont and Martimort (1995), Laffont and Meleu (1997), Kofman and Lawarrée (1993) and Felli (1997). These works view the side-contract between a risk-neutral supervisor protected by limited liability and his supervisee as purely im-

[6] Melumad, Mookherjee and Reichelstein (1995) and Baron and Besanko (1992) provide examples where delegation involves no cost for the hierarchy in settings with risk-neutral agents. McAfee and McMillan (1995) discuss agency costs of delegation when intermediate agents are instead protected by limited liability. We show thereafter that these two settings are polar cases of our analysis.

plicit, being enforced by a word of honor, trust, reciprocity,[7] or through repeated relationships.[8] This paradigm nevertheless short-cuts the modeling of the explicit constraints that side-contracting involves. In particular, the possible asymmetric information within an informal group stressed out by Crozier plays no role in the analysis.[9] These works assume that the efficiency of a side-contract is exogenously given. Side-contracting incurs some transaction costs. For instance, Tirole (1986) argues that these costs come from the fact that bribes are non-monetary in nature. Laffont and Martimort (1995) show how these transaction costs depend on the information available to the supervisor. Laffont and Meleu (1997) explain that exchanges of reciprocal favors are easier than asymmetric deals because the norm of reciprocity reduces transaction costs. Except in Martimort (1997) who obtains these transaction costs in a model where collusion is self-enforcing through a repeated relationship,[10] the collusion literature does not derive transaction costs of side-contracting from more fundamental features of the organizational environment like shared knowledge of information between the colluding partners, preferences, stakes of their relationship. Nevertheless, the hard information paradigm has proved to be extremely useful to study how the emergence of group behavioral norms within a hierarchy puts constraints on the internal efficiency of the firm. In particular, it has shown how administrative rules and routines become then optimal responses to the threat of collusion within the organization.

Our model of delegated contracting with formal contracting, soft information and endogenous transaction costs can be compared with Tirole's hierarchical model with informal contracting, hard information and exogenous transaction costs. In our framework, the transaction costs of side-contracting are endogenously derived. They are linked to the supervisor's degree of risk-aversion, the information structure within the coalition and the stakes of the delegated authority relationship.

As in Tirole (1986), the optimal design of incentives within the hierarchical structure becomes an important tool to curb coalitional behavior. However, the owners of the firm may also choose to play on other fundamen-

[7]See Gouldner (1961) for some discussion of the enforcement of these norms of reciprocity within the firm. Fehr, Gachter and Kirchsteiger (1997) argue that selfishness may lead to restrictions in the set of implementable allocations in a two tier principal-agent model and that preferences for reciprocity should be considered. Rotemberg (1994) and Prendergast and Topel (1996) analyze altruism and favoritism in three-tier hierarchies but do not take into account the optimal organizational response to the threat of these collusions.

[8]See Tirole (1992) and Martimort (1997) for some formal models of these self-enforcing collusive behaviors within public or private organizations.

[9]Except in Felli (1997) who models collusion under asymmetric information between a supervisor and his agent in a three-tier hierarchy.

[10]We comment more extensively on the comparison between the two approaches in Section 5.

tal features of the organization to affect the values of the transaction costs of side-contracting. Endogenous transaction costs help to understand how the principal can also attempt to influence the hidden gaming between the supervisor and his supervisee to better control their coalitional behavior. Laffont (1990) shows how incentive schemes can be used to influence the hidden gaming between colluding agents to constrain their collusive behavior. We go one step further by stressing that, on top of incentive schemes, organizational design itself is also useful in curbing these collusions.

From this explicit formulation of transaction costs, we derive several important principles for organizational design. These principles apply over a whole range of issues which include the distribution of heterogeneous agents within the hierarchy, the design of supervisory structures, the organizational choice between a M- and a U-form type of organization, or the consequences of vertical integration on internal incentives.

Section 2 presents the model and discusses our description of the firm as a nested delegated contracting game. Section 3 derives a Delegation-Proofness Principle which generalizes the Revelation Principle to the case of collusion between lower layers. We describe the *coalition incentive constraints* which constrain the allocation of resources whithin the firm. In Section 4, the optimal grand-contract is derived as well as and some comparative statics. Section 5 recalls the main assumptions and results of Tirole's (1986) model of supervision with hard information and exogenous transaction costs. Endogenous transaction costs which identify Tirole's model with ours are then obtained. In Section 6, using the characterization of these transaction costs we derive several principles of organizational design.

2. THE MODEL

2.1. Players and Information

We consider a three-tier model of a firm's bureaucracy where authority and responsability are formally allocated along the different nodes of a hierarchical tree. A principal (hereafter P), for instance the firm's owner, delegates contracting with a productive agent (A) to a supervisor (S).

This separation between ownership and supervision within the two upper tiers of the hierarchy is motivated by physical constraints: the principal is supposed to be unable to perform himself the supervisory task. This could well be the case because the activities of the firm are large in size or because supervision requires some specific skills and the specialization of labor imposes different costs of performing it. We focus on the polar case where the principal's cost of supervision is so high that this task has to be performed by a professional supervisor. Many real world situations support our modeling as, most of the time, incentive packages paid to members of the firm are not directly determined by its owners.

The agent produces a quantity q of output at a cost θq. The marginal cost θ is private information to the agent, drawn from a discrete distribution on $\Theta = \{\theta_1, \theta_2\}$ (we denote $\Delta\theta = \theta_2 - \theta_1 > 0$). The supervisor is uninformed about the agent's true type. Nonetheless, he receives a signal τ on his marginal cost. τ is drawn from a discrete distribution on $T = \{\tau_1, \tau_2\}$.

This signal is not observed by the principal, otherwise a supervisor would not be needed, but it is also learned by the agent. Hence, information sets are nested along the hierarchy: "nature" reveals to the agent both his type and the supervisor's information, only the latter is available to the supervisor while the principal observes none of these. This nested information structure is standard in the literature on collusion in hierarchies (see Tirole (1986) and McAfee and McMillan (1995) among others). This assumption can be justified by the fact that the supervisory activity consists in performing a certain number of checks on the agent's job and that the agent observes the result of these checks.[11]

The joint probabilities on (θ_i, τ_j) are defined as $p_{ij} = Prob(\theta = \theta_i, \tau = \tau_j)$ with $p_{ij} > 0$ for all i, j. The coefficient of correlation is $\rho = p_{11}p_{22} - p_{12}p_{21}$. We adopt the convention that $\rho \geq 0$, reflecting the supervisor's ability to collect information about the agent's cost. From the joint distribution above, one can define the conditional probabilities $p(\theta/\tau)$ and a positive correlation can also be interpreted as a monotone likelihood ratio property: $\frac{p(\theta_2/\tau_1)}{p(\theta_1/\tau_1)} \leq \frac{p(\theta_2/\tau_2)}{p(\theta_1/\tau_2)}$. A particular example of such an information structure that we use in Section 5 is the following: the agent has a low (resp. high) cost θ_1 (resp. θ_2) with probability ν (resp. $1-\nu$) and the conditional probabilities of the signal are: $p(\tau = \tau_1/\theta_1) = p(\tau = \tau_2/\theta_2) = \epsilon$. The probability ϵ ($\geq \frac{1}{2}$) can be thought of as the signal's precision. For ϵ close to one half, the signal conveys little information about the agent's type while if $\epsilon = 1$, the supervisor exactly observes the type.

An important feature of our modeling is that the supervisor's signal is supposed to be *soft information*. This means that this information is fully manipulable by the supervisor: whatever the true realisation of τ, there is no constraint whatsoever on what the supervisor can report.

This assumption contrasts with the existing literature on supervision. Following Tirole (1986)'s seminal paper, the supervisory technology has so far always been modeled as producing *hard information*, i.e., information which can be concealed but not manipulated. This approach gives a clear account of the discretionary power of a supervisor: concealing or not evi-

[11] This assumption simplifies the analysis. If the supervisor's information was not observed by the agent, the contractual game between him and the agent would occur in a setting of informed principal with private values (see Maskin and Tirole (1990)).

dence. Even though this short-cut has proved to be very useful for studying various kinds of problems an alternative approach is needed.

An immediate justification for modeling supervision with soft information is the complement this approach offers. Indeed, there are many real life examples like regulation, advocacy, banking, insurance etc.. where supervisors (or auditors) not only conceal some information but can also produce false reports. But more importantly, our main motivation is the need for theoretical consistency.

The whole adverse-selection literature has been developed in a context of soft information. In particular, in the case of direct mechanisms, an agent may report any possible message regarding his type. Actually, in the collusion literature mentioned above, soft and hard information always coexist: the agent himself is supposed to be unable to produce verifiable information about his type, otherwise supervisors will be useless. Hence, this literature is concerned with a special kind of private information: information which cannot be made verifiable by privately informed agents while external parties can do so. This peculiar feature calls for a "theory of verifiability:" What makes a piece of information verifiable? One possible route for a preliminary and partial answer is to explicitly consider the existence of communication costs in the organization: different agents have different costs of communicating different pieces of information. A supervisor, in addition to his competence, is also an agent with prohibitive costs of communicating false information. Unfortunately, there is no satisfying theory of communication costs on which these assumptions could be grounded. A related problem is that once we admit that supervisors can produce hard evidence, the notion of the privacy of this hard information becomes difficult to define. There is a scope for using yardstick mechanisms between agents who have access to this verifiable information. A final point is related to a second short-cut which has been extensively used in the hard information paradigm: the exogeneity of the transaction costs of side-contracting between agents and supervisors. How this exogeneity interacts with the constraints on communication defining hard information is an important question. We show below that, indeed, these transaction costs are not independent of the possibilities to manipulate information.

The present paper offers a more consistent approach of supervision in an adverse selection setting. By the same token, we also endogenize these transaction costs in an unambiguous way. From that point of view, the paper makes a methodological contribution as it provides some foundations for these costs, as well as a reduced form which becomes useful in discussing a wide variety of organizational problems.

2.2. Preferences and Contracts.

The agent is risk-neutral and his utility function is $U = t - \theta q$ where t is the monetary transfer from the supervisor to the agent. Under our assumption that the agent is informed on the supervisor's signal, nothing would be changed if the agent were risk-averse: contracts have to be accepted ex post, i.e., knowing both types of information. An agent accepts to produce as long as he gets his reservation utility, normalized w.l.o.g. to zero.

The principal delegates to a supervisor the right to contract with the agent. Indeed, this contractual choice does not entail any loss of generality: delegation is one way to implement the optimal contract that the principal would offer if he was able to communicate directly with the agent when the supervisor and the agent can collude.[12]

The supervisor receives a transfer s from the principal to perform the monitoring task and has authority on the agent's incentive scheme. We assume that the supervisor is risk-averse. For tractability, let his Von-Neuman Morgenstern utility function be $V(x) = 1 - exp(-rx)$ where r is the constant degree of risk-aversion.[13] $r = 0$ corresponds to the limiting case where the supervisor is risk-neutral. As already suggested, introducing a positive degree of risk-aversion will prove to be extremely useful in identifying the agency cost of a hierarchical structure. Taking conditional expectations, the supervisor's utility writes as $E_\theta\big(V(s-t)|\tau\big)$.
Note that if the supervisor expects a level of wealth w_1 with probability p and w_2 with probability $1-p$, the certainty equivalent of his utility is simply:

$$w_1 - \frac{1}{r} ln\big(p + (1-p)e^{-r(w_2 - w_1)}\big)$$

The production of q units of output yields a revenue $R(q)$ to the principal. His profit is then $\Pi = R(q) - s$.

2.3. Timing

The timing of the game is as follows:

• The agent learns his productivity parameter θ and the supervisor's signal is τ. The supervisor learns only τ.

---

[12] Given that asymmetric information in the collusive ring is the only (endogenous) constraint on the efficiency of collusion, the methodology developed by Laffont and Martimort (1997b) could be used to show that the optimal contract offered by the principal when this collusion is organized by a biased third party which maximizes the supervisor's ex ante utility coincides with the one developed hereafter. Consequently, in such a context, delegation is the best organizational response of the principal/owner to the possibility of biased collusion.

[13] Itoh (1993) develops a model of collusion between two risk-averse agents. However, collusion takes place under symmetric information. Prendergast and Topel (1996) develop also a model of favoritism in organizations with agents and supervisor having CARA utility functions.

• The principal offers a contract (denoted thereafter "grand-contract") to the supervisor which stipulates a wage as a function of the firm's output. He also delegates to the latter the right to contract with the agent.[14] The supervisor accepts or refuses this contract under asymmetric information on the agent's type.

• The supervisor offers a contract (denoted thereafter "side-contract")[15] to the agent which stipulates a wage as a function of the firm's output. The agent accepts or refuses this contract.

• Production takes place and transfers within the grand- contract and the side-contract realize.

The acceptance of the grand-contract by the supervisor being made before the learning of the agent's type, the supervisor's interim participation constraints must be satisfied by this grand-contract.[16] Since the agent is informed of the supervisor's signal at the time of accepting the side-contract, the side-contract must also satisfy the agent's ex post participation constraints.

2.4. Benchmark

As a benchmark, let us consider the case where the principal directly receives the signal τ on the agent's private information. This can be viewed as a stylized model of a small firm in which the supervisory task can be performed by the principal himself. Absent any supervisor, the firm's owner pays directly the agent's wage. Alternatively, if we stick to the interpretation of our model as a picture of a large firm in which supervision is needed, everything happens as if the supervisor would costlessly reveal his information to the principal before the latter contracts with the agent.

Using the Revelation Principle, there is no loss of generality in looking for the optimal contract within the class of direct truthful revelation mechanisms of the form $\{t(\hat{\theta}, \tau); q(\hat{\theta}, \tau)\}$ where $\hat{\theta}$ is the agent's report on his efficiency parameter and τ is the signal observed by the principal.

For ease of notations, we denote thereafter $t_{ij} = t(\theta_i, \tau_j)$, $q_{ij} = q(\theta_i, \tau_j)$ and we define $u_{ij} = t_{ij} - \theta_i q_{ij}$ the rent of the agent in state (θ_i, τ_j). As it is standard in two-type adverse selection models, the following constraints are of particular importance:[17]

[14] We assume that the supervisor and the agent cannot write a contract together before the principal's offer. This timing captures a setting in which the principal has a maximal ability to commit by designing the contours of the organization.

[15] We assume that this contract is deterministic, because the enforceability of stochastic contracts is too demanding.

[16] The timing of our game is thus similar to other models in the literature on hierarchical delegation (see Baron and Besanko (1992), Melumad, Mookherjee and Reichelstein (1995) and Laffont and Martimort (1998) among others).

[17] When the following constraints are binding, as it will be the case at the optimum of the principal's problem, it is easy to show that the other constraints are strictly satisfied.

- Incentive compatibility constraints for an efficient agent:

$$u_{1j} \geq u_{2j} + \Delta\theta q_{2j}, \quad \text{for} \quad j \in \{1,2\}, \tag{1}$$

- Participation constraints of an inefficient agent:

$$u_{2j} \geq 0, \quad \text{for} \quad j \in \{1,2\}. \tag{2}$$

When he has observed a signal τ_j, the principal updates his beliefs on the agent's type. Conditional probabilities become $p(\theta_1|\tau_1) = \frac{p_{11}}{p_{11}+p_{21}}$ and $p(\theta_1|\tau_2) = \frac{p_{12}}{p_{12}+p_{22}}$. Accordingly, the optimal contract solves:

$$\max_{\{q_{1j}, q_{2j}, u_{1j}, u_{2j}\}} p(\theta_1|\tau)(R(q_{1j}) - \theta_1 q_{1j} - u_{1j}) + p(\theta_2|\tau)(R(q_{2j}) - \theta_2 q_{2j} - u_{2j})$$

subject to (1) and (2).

Solving this problem yields the *conditional optimum* defined as:

$$R'(q_{1j}^*) = \theta_1 \tag{3}$$

$$R'(q_{2j}^*) = \theta_2 + \frac{p_{1j}}{p_{2j}}\Delta\theta. \tag{4}$$

To reduce the cost of the incentive compatibility constraint (1) and make less valuable for an efficient agent to mimic an inefficient one, the principal reduces the output produced by an inefficient agent. A positive rent is left to the efficient agent ($u_{1j} = \Delta\theta q_{2j}^*$) while the participation constraint (2) of a high cost agent is binding ensuring that his rent u_{2j} is equal to zero. Offering a contract where the quantity produced by a high cost agent is distorted entails some benefits when the agent turns out to be good (less rents) but some costs when he is in fact a θ_2 type (less than optimal production).

Importantly, the monotone likelihood ratio property implies that the inefficient agent's output is more distorted following the observation of τ_1 than when τ_2 is observed,

$$q_{21}^* < q_{22}^*.$$

Indeed, the firm is more likely to be efficient when τ_1 is observed than when τ_2 is observed. Lowering the informational rents of a good agent becomes relatively more important than ensuring that the production of a high cost one is not too far from the optimal level.

3. THE COST OF DELEGATED AUTHORITY

3.1. The Delegation-Proofness Principle

We now consider the case where the supervisor's information about the agent's type is not observable by the principal. The principal proposes an incentive scheme to the supervisor to induce him to exert authority. This contract specifies some quantities to be produced and a wage paid to the supervisor as a function of the quantity which is produced. We call the contract between the principal and the supervisor the grand-contract (GC). Then, the supervisor will contract with the agent and we refer to this contract as a side-contract (SC). We start by characterizing the optimal side-contract for any grand-contract.

For any GC, a SC is composed of two objects: the first one is an "internal" transfer paid by the supervisor out of the funds he receives from the principal. The second one is a report function from the supervisor to the principal. Equivalently, a SC is a menu of internal transfer-quantity pairs: to any report made by the supervisor in the GC corresponds a quantity.

Given that the supervisor's signal is observed by both the supervisor and the agent, we define a SC for any τ as $SC_\tau = \{t(\hat{\theta},\tau); m(\hat{\theta},\tau)\}$ where $\hat{\theta}$ is now the agent's report to the supervisor and $m(\hat{\theta},\tau)$ denotes the message sent by the supervisor to the principal. Observe immediately that our presentation focuses on direct mechanisms for the side-contracts: the agent sends a message about his *type*. Indeed, from the Revelation Principle applied at the stage of side-contracting, there is no loss of generality in looking for the optimal side-contract within the class of *direct* mechanisms. Let us turn our attention now to the report function $m(\hat{\theta},\tau)$. For instance, $m(\theta_2,\tau_1)$ is the message that the supervisor will report in the GC when the agent has reported (truthfully or not) θ_2 to the supervisor and that τ_1 is observed. The message reported by the supervisor in the GC can contain more than the supervisor's own type. Allowing for this possibility accounts for the fact that the supervisor may also report information not only on his own type but also on what he has learned from his relationship with the agent. If, for instance, $m(\theta_2,\tau_1) = (\theta_1,\tau_1)$, then the supervisor is manipulating the report because he is not transmitting the value of θ reported by the agent. Of course, both variables could be misreported. There is no reason a priori why we should restrict the supervisor's message space M_s to $\Theta \times T$, i.e., to restrict the GC to be a direct mechanism. Nonetheless, the Revelation Principle can easily be extended to this delegation framework. We sketch the argument here and we refer the reader to the Appendix for a formal proof.

The risk-averse supervisor informed on τ but not on the agent's efficiency parameter proposes a side-contract which is solution to the following problem (denoted (S) hereafter):

$$\begin{gathered}\max_{\{m(\theta_i,\tau_j),u_{ij}\}} p(\theta_1|\tau_j)V(s(m(\theta_1,\tau_j)) - \theta_1 q(m(\theta_1,\tau_j)) - u_{1j}) \\ +p(\theta_2|\tau_j)V(s(m(\theta_2,\tau_j)) - \theta_2 q(m(\theta_2,\tau_j)) - u_{2j}) \\ \text{subject to (1) and (2).}\end{gathered}$$

The internal transfers t are such that both constraints (1) and (2) are binding.[18] Consequently, the manipulation function $m(\theta_i,\tau_j)$ is solution to:

$$\begin{aligned}&\max_{\{m(\theta_i,\tau_j)\}} p(\theta_1|\tau_j)V\left(s(m(\theta_1,\tau_j)) - \theta_1 q(m(\theta_1,\tau_j)) - \Delta\theta q(m(\theta_2,\tau_j))\right) \\ &+p(\theta_2|\tau_j)V\left(s(m(\theta_2,\tau_j)) - \theta_2 q(m(\theta_2,\tau_j))\right)\end{aligned}$$

Or, equivalently:

$$m^*(\theta_1,\tau_j) \in arg\max_{\tilde{m}} \{s(\tilde{m}) - \theta_1 q(\tilde{m})\} \tag{5}$$

$$\begin{aligned}&m^*(\theta_2,\tau_j) \in arg\max_{\tilde{m}} \Big\{V\big(s(\tilde{m}) - \theta_2 q(\tilde{m})\big) \\ &+\frac{p(\theta_1|\tau_j)}{p(\theta_2|\tau_j)}\, V\bigg(s(m^*(\theta_1,\tau_j)) - \theta_1 q(m^*(\theta_1,\tau_j)) - \Delta\theta q(\tilde{m})\bigg)\Big\}.\end{aligned} \tag{6}$$

The solution to this problem defines a mapping $m^*(\cdot)$ from $\Theta \times T$ into M_s. There is no loss of generality in restricting the analysis to *delegation-proof grand-contracts* such that $m^*(\cdot) = Id$ is an optimal strategy for the supervisor. The logic of the argument is by now standard and is a simple generalization of that behind the Revelation Principle: take any grand-contract GC with message space M_s and such that the optimal side-contract of the supervisor with the agent stipulates a response $m^*(\cdot)$. One can construct a new grand-contract $\tilde{GC} = GC \circ m^*(\cdot)$ with message space $\Theta \times T$ and such that the optimal side-contract that is then offered by the supervisor entails now no manipulation, i.e., $\tilde{m}^*(\cdot) = Id$. This delegation-proofness principle allows a direct and more tractable description of the set of implementable grand-contracts.

To illustrate the role played by the report function $m(\cdot)$ in the side-contract, let us compute what the supervisor reports when the GC implementing the conditional second best is offered, assuming that the transfers are first paid to the supervisor. For instance, looking at the solution characterized by equations (1) to (4), it is easy to check that the supervisor will

[18] Again, the incentive compatibility constraint of an inefficient agent and the participation constraint of an efficient one turn out to be strictly satisfied at the optimum.

misreport τ when the agent has a low cost: $s(\theta_1, \tau_1) - \theta_1 q_{11} = \Delta\theta q_{21} < s(\theta_1, \tau_2) - \theta_1 q_{12} = \Delta\theta q_{22}$. Whatever the true value of τ, the supervisor will report $\tau = \tau_2$ when the agent is good: $m^*(\theta_1, \tau_j) = (\theta_1, \tau_2)$ for all τ_j. Actually, the problem may be more severe as the supervisor may also lie along the θ dimension. But already, we can conclude that the benchmark allocation is not delegation-proof. To understand the intuition behind our results hereafter, it may be useful to develop an analogy with the concept of "hidden gaming" in hierarchies (Laffont (1990)): here, the supervisor is playing a hidden game with the agent as he does not adopt a truthful report function but instead a manipulative one. To some extent, this side-contract can be thought of as a moral hazard variable: the principal has to design a contract ensuring that the supervisor chooses the right "game," i.e., the truthful report function. This analogy will prove to be useful in the sequel because the crux of our argument lies in the interplay of this incentive problem with the risk-aversion of the supervisor, as in more standard moral hazard models.

When the supervisor's utility function is CARA, we can rewrite equations (5) and (6) which, together with the fact that $m^*(\cdot) = Id$ at the optimum, yields the description of the following delegation-proof constraints.

PROPOSITION 1. *A grand-contract C is delegation-proof if and only if $M_s = \Theta \times T$ and the following coalition incentive constraints are satisfied:*

$$(\theta_1, \tau_j) \in \arg\max_{\tilde{m}} \{s(\tilde{m}) - \theta_1 q(\tilde{m})\}, \tag{7}$$

$$\begin{aligned}(\theta_2, \tau_j) \in \arg\max_{\tilde{m}} \Bigg\{ & - \Delta\theta q(\tilde{m}) \\ - & \frac{1}{r} \ln \left(p(\theta_1|\tau_j) + p(\theta_2|\tau_j) e^{-r\left(s(\tilde{m}) - \theta_2 q(\tilde{m}) - (s(\theta_1,\tau_j) - \theta_1 q(\theta_1,\tau_j)) + \Delta\theta q(\tilde{m})\right)} \right) \Bigg\}\end{aligned} \tag{8}$$

The implementable grand-contracts must not only be delegation-proof but must also satisfy the supervisor's interim participation constraints. To better describe all these constraints, it is useful to define $v_{ij} = s_{ij} - \theta_i q_{ij}$ as the aggregate payoff, or ex post rent, of the supervisor-agent coalition. With this notation we can easily write the incentive constraints ensuring that the supervisor chooses a truthful report function.

From two revealed-preference arguments using (7) when $\tau = \tau_1$ and $\tau = \tau_2$, we have immediately that:

$$v_{11} = v_{12}. \tag{9}$$

One cannot discriminate between two coalitions with an efficient agent in terms of their aggregate payoffs whatever the supervisor's signal.

For a given τ, the supervisor will not manipulate the report on θ when the agent is efficient if and only if:

$$v_{11} \geq v_{21} + \Delta\theta q_{21}, \tag{10}$$

$$v_{12} \geq v_{22} + \Delta\theta q_{22}. \tag{11}$$

Moreover, when the agent turns out to be inefficient (θ_2), (8) also prevents a supervisor having observed a τ_1 signal from reporting (θ_2, τ_2). This constraint rewrites as:

$$\begin{aligned} & v_{11} - \Delta\theta q_{21} - \tfrac{1}{r}\ln\left(\tfrac{p_{11}}{p_{11}+p_{21}} + \tfrac{p_{21}}{p_{11}+p_{21}} e^{-r(v_{21}-v_{11}+\Delta\theta q_{21})}\right) \\ \geq\ & v_{11} - \Delta\theta q_{22} - \tfrac{1}{r}\ln\left(\tfrac{p_{11}}{p_{11}+p_{21}} + \tfrac{p_{21}}{p_{11}+p_{21}} e^{-r(v_{22}-v_{11}+\Delta\theta q_{22})}\right). \end{aligned} \tag{12}$$

We have isolated above the downward coalition incentive constraints which will turn to be the only relevant ones. It is shown in the Appendix that the other coalition incentive constraints are satisfied at the optimum of the principal's problem.

Still using the notations above and expressing the supervisor's utility with certainty equivalents, his interim participation constraints rewrite respectively as:

$$v_{11} - \Delta\theta q_{21} - \frac{1}{r}\ln\left(\frac{p_{11}}{p_{11}+p_{21}} + \frac{p_{21}}{p_{11}+p_{21}} e^{-r(v_{21}-v_{11}+\Delta\theta q_{21})}\right) \geq 0, \tag{13}$$

$$v_{12} - \Delta\theta q_{22} - \frac{1}{r}\ln\left(\frac{p_{12}}{p_{12}+p_{22}} + \frac{p_{22}}{p_{12}+p_{22}} e^{-r(v_{22}-v_{12}+\Delta\theta q_{22})}\right) \geq 0. \tag{14}$$

3.2. The Optimal Grand-Contract

The principal optimizes the firm's expected profit subject to coalition-incentive and interim participation constraints. The optimal grand-contract solves the following problem (denoted thereafter by (P)):

$$\max \sum_{i,j} p_{ij}(R(q_{ij}) - \theta_i q_{ij} - v_{ij})$$

subject to constraints (9) to (14).

PROPOSITION 2. *The optimal delegation-proof grand-contract entails:*

- *Constraints (11) to (14) and (9) are all binding. All other constraints are strictly satisfied.*

- *A decreasing schedule of outputs* $q_{11}(r) = q_{12}(r) > q_{22}(r) > q_{21}(r)$ *is implemented where* $q_{ij}(r)$ *are implicitly defined by:*

$$R'(q_{11}(r)) = R'(q_{12}(r)) = \theta_1, \tag{15}$$

$$R'(q_{21}(r)) = \theta_2 + \Delta\theta \frac{p_{11}e^{-r\Delta\theta(q_{22}(r)-q_{21}(r))}}{p_{21} + p_{11}(1 - e^{-r\Delta\theta(q_{22}(r)-q_{21}(r))})}, \tag{16}$$

$$R'(q_{22}(r)) = \theta_2 + \frac{\Delta\theta}{p_{22}}\left(p_{11} + p_{12} - \frac{p_{21}p_{11}e^{-r\Delta\theta(q_{22}(r)-q_{21}(r))}}{p_{21} + p_{11}(1 - e^{-r\Delta\theta(q_{22}(r)-q_{21}(r))})}\right) \tag{17}$$

- *Moreover,* $q_{22}(r) < q_{22}^*$ *and* $q_{21}(r) > q_{21}^*$.

To better understand the nature of these distortions, assume first that the supervisor is risk neutral: $r = 0$. From (15) to (17), it appears that the conditional second best derived in the benchmark case is still the optimal contract. Nonetheless, the monetary transfers are now different. Instead of giving an ex post utility to the supervisor always equal to zero as in Section 2.4, now his ex post utility varies. A supervisor who has observed τ_2 faces a riskless lottery. Indeed, his ex post utility is either $v_{22} = 0$ or $v_{12} - \Delta\theta q_{22} = 0$. A τ_1 supervisor is instead offered a risky lottery. He receives a reward $v_{11} - \Delta\theta q_{21} = \Delta\theta(q_{22} - q_{21}) > 0$ when $\theta = \theta_1$ but is inflicted a punishement when the agent turns out to be bad: $v_{21} = -\frac{p_{11}}{p_{21}}\Delta\theta(q_{22} - q_{21}) < 0$. To restore incentive compatibility the principal has indeed to link the fate of the supervisor to the "quality" of his supervisory report: to reward him when the two pieces of information coincide (τ_1 with $\theta = \theta_1$) and to punish him when they do not.[19] Consequently, with a risk-neutral supervisor, the principal can restore incentive compatibility *at no cost*. Alternatively, one can think that the principal simply sells the firm to the supervisor. Without risk-aversion, the supervisor is willing to pay what the right to contract with the agent is worth to. Unfortunately, this is no longer true when the supervisor is risk-averse. Monetary transfers with the same expected values for the principal differ in the amount of risk they impose on the supervisor. A risk-averse supervisor accepts the lottery proposed by the principal only if he is paid a wage corresponding to the risk-premium associated with this lottery. The principal can no longer sell the firm at a price equal to his own valuation. Given the binding constraints, the total risk borne by the supervisor depends on the spread between q_{22} and q_{21}: the larger this spread, the higher the risk faced by the supervisor. As a result, distorting outputs becomes also a valuable tool to reduce the cost of incentives. In particular, q_{22} is distorted downwards and q_{21} is distorted upwards so that

[19] This result is reminiscent of the Crémer and McLean (1988) mechanisms where the correlation between different agents' information is used to costlessly induce information revelation.

the payoff lottery faced by a supervisor who has observed τ_1 becomes less risky. The risk-premium required by the supervisor is then reduced. The next subsection turns to a more careful analysis of these distortions but clearly their magnitude trades off the efficiency loss in production with the insurance motive.

Proposition 2 shows that even though two economic organizations share the same technical characteristics -same production and supervision technologies- their optimal production plans as well as their possibilities to delegate authority can significantly differ. The reason is the conflict between incentives and the exploitation of gains from trade coming from insurance motives. The optimal contract has to provide insurance to a risk-averse supervisor and to induce him to use his delegated authority in the most efficient manner from the point of view of the firm's profit. The next section discusses how the optimal response of the organization changes with the central parameters of the model.

4. ENDOGENOUS TRANSACTION COSTS

At this point, two lessons emerge from our analysis. Even though the information produced by a supervisor is not hard evidence and can be manipulated, supervision helps to improve incentives in organizations. However, this benefit comes at a cost which depends on the supervisor's degree of risk-aversion. The second lesson reinforces previous findings in the literature by showing how rather simple mechanisms which are relatively less sensitive to the supervisor's signal are chosen when authority must be delegated to lower tiers of the hierarchy. The separation of ownership and supervision implies that contracts are tilted towards more administrative rules leaving little discretion to the supervisor.

We now move one step further into the analysis of optimal hierarchical structures and compare the results of our analysis with those of Tirole (1986)'s model of supervision with hard information and exogenous transaction costs. Beyond the aim of proposing an unifying approach to supervision, our analysis provides a theoretical justification to the (so far) exogenous transaction costs of side contracting in organization, as well as a reduced form for these. This reduced form allows us to address various issues about the internal design and management of organizations.

4.1. Supervision with Hard Information.

We start by briefly presenting an adaptation of Tirole (1986)'s model to our informational setting.

• The supervisor is risk-neutral for positive wealth and infinitely risk-averse below zero wealth. An alternative interpretation is that he is protected by limited liability.

• The supervisor observes a signal $\tau \in \{\tau_1, \tau_2\}$ on the agent's type. The signal τ_1 is "good news" on the fact that the agent is efficient. After having observed τ_1, the supervisor updates his beliefs on the agent's type, giving more weight on the firm being efficient. The monotonic likelihood ratio property $\frac{p(\theta_2|\tau_1)}{p(\theta_1|\tau_1)} \leq \frac{p(\theta_2|\tau_2)}{p(\theta_1|\tau_2)}$ holds.[20]

The crucial assumption regarding the supervisor's signal is that it is hard information. Following Tirole, the supervisor has the possibility not to reveal good news about the agent. Nonetheless, if the supervisor receives a signal which is "bad news" about the agent's type he cannot forge evidence and claim that he learned "good news". Alternatively, one can say that it is common knowledege that the supervisor's signal is at least τ_2 and good news refer to the fact that there is an improvement $\tau_2 - \tau_1$. Depending on his signal, the message correspondences available to the supervisor are thus $M(\tau_1) = \{\tau_1, \tau_2\}$, $M(\tau_2) = \{\tau_2\}$.

• The principal offers a grand-contract (quantities and transfers) to both the agent and the supervisor.[21] Nonetheless, the supervisor and the agent can collude through an enforceable side-contract. However, in the background, this side-contract is a reduced form for some informal relational contract among the two partners.

• The supervisor and the efficient agent collude under symmetric information, i.e., before the agent learns θ but after the supervisor and the agent have learned τ. The collusive agreement specifies the bribes the supervisor will receive when he accepts not to report τ_1 and the agent is efficient. It is assumed that the supervisor can then extract the rent differential of the efficient agent between what he gets when τ_2 is revealed and what he gets when τ_1 is instead announced, namely $u_{12} - u_{11} = \Delta\theta(q_{22} - q_{21})$.[22] However, because of frictions in side-contracting, the benefit from colluding with the agent writes as $k(u_{12} - u_{11})$. The fact that $k < 1$ captures the existence of some transaction costs.

[20] Stricto sensu, the monitoring technology available to the supervisor in Tirole's model is different from that in this paper. The supervisor may observe or not the true type of the agent. Since an inefficient agent receives zero rent from the optimal contract, collusion is a concern only when the supervisor observes that the firm is efficient. Hence, in Tirole's model, collusion arises when the supervisor gets information which brings "good news." We have chosen to work with a noisy information structure even in the hard information version of our model.

[21] This is a difference with the approach above where the principal does not communicate with the agent.

[22] The assumption on bargaining power is also made in Tirole. Moreover, as in his analysis, there is no collusion stake with an inefficient agent since the latter's rent is equal to zero whatever τ.

Let s_{ij} be the transfer from the principal to the supervisor when the messages received by the principal are (θ_i, τ_j). The collusion-proofness constraint which needs to be satisfied by the grand-contract to induce revelation of τ_1 by the supervisor writes therefore as:

$$s_{11} - s_{12} \geq k(u_{12} - u_{11}) = k\Delta\theta(q_{22} - q_{21}). \tag{18}$$

Moreover, limited liability constraints also impose:

$$\text{For all } i, j \qquad s_{ij} \geq 0. \tag{19}$$

The optimal grand-contract with hard information supervision obtains therefore as a solution to the following program (denoted thereafter by (H)):

$$\max \sum_{i,j} p_{ij}(R(q_{ij}) - \theta_i q_{ij} - u_{ij} - s_{ij})$$

$$\text{subject to (1), (2), (23), (24).}$$

Solving (H) yields the following outputs[23]:

$$R'(q_{11}) = R'(q_{12}) = \theta_1 \tag{20}$$

$$R'(q_{21}) = \theta_2 + \Delta\theta(1-k)\frac{p_{11}}{p_{12}}, \tag{21}$$

$$R'(q_{22}) = \theta_2 + \Delta\theta\left(\frac{p_{12} + p_{11}k}{p_{22}}\right). \tag{22}$$

First, observe that the conditional optimum q_{21}^*, q_{22}^* still obtains when $k = 0$. The transaction costs of collusion are then so large that no collusive agreement can be enforced. Consequently, we are back to the case of a benevolent supervisor.

To reduce the cost of the collusion-proofness constraint, the principal distorts q_{22} downwards and q_{21} upwards as shown respectively in (21) and (22). Those distorsions are all the more important that k is close to one, i.e., that collusion is very efficient. When transaction costs of side-contracting are small, colluding parties can manipulate information at no cost and deterring collusion requires larger distortions.

4.2. Stake-Dependent Transaction Costs

We now give a clear comparison between our model of hierarchical formal side-contracting under delegation and the model of hierarchical informal

[23] We assume $k \leq \frac{\rho}{p_{11}(p_{12}+p_{22})}$. For larger k, $q_{21} = q_{22}$

side-contracing. The soft and hard information models of supervision can indeed be immediately identified and compared:

• First, in the case of soft information, even if a large number of coalition incentive constraints may be potentially important due to the unrestricted possibilities for manipulations, only a few of them turn out to be relevant. This reduction in the set of relevant incentive constraints fits in with the exogenous communication constraint imposed in the hard information model. The relevant coalition incentive constraints are the same.

• Second, a *pointwise* identification of the two sets of first order conditions characterizing outputs in the soft and in the hard information paradigms yields immediately the following expression for k:

$$k(\Delta q(r), \Delta\theta, r) = 1 - \frac{p_{21}e^{-r\Delta\theta\Delta q(r)}}{p_{21} + p_{11}(1 - e^{-r\Delta\theta\Delta q(r)})} \tag{23}$$

where $\Delta q(r) = q_{22}(r) - q_{21}(r)$ with $q_{22}(r)$ and $q_{21}(r)$ being the optimal outputs which solve (16) and (17). The output distortions of the soft and hard information models are the same when k depends in fact on the collusion stake $\Delta\theta\Delta q(r)$, the degree of risk-aversion r and the probability distribution.

In the case of infinite risk-aversion, if one assumes that the collusion stake Δq is held constant, $k(\Delta q, \Delta\theta, r)$ converges towards 1. Collusion becomes increasingly efficient with a greater risk-aversion. In the case of risk-neutrality, we have instead $k(\Delta q(r), \Delta\theta, r) = 0$ and collusion has no bite on the firm's profit. Risk-aversion puts a limit on the principal's ability to play the agent against the supervisor by using the correlation between their private information. Similarly, a large exogenous k implies that the supervisor and the agent can easily coordinate their reports.

More generally, $k(\Delta q(r), \Delta\theta, r)$ belongs to $[0, 1[$ as suggested by Tirole (1986). However, from the inspection of the last equation, $k(\cdot)$ is *not independent* of the stake $\Delta q(r)$. This has potentially far reaching consequences as this shows that one cannot simultaneously assume that k is independent of the collusion stake and strictly lower than 1.

The above analysis nevertheless suggests that it may be possible to fully identify the results of Section 3 with those obtained when transaction costs are *stake-dependent* in Tirole's model. In this case, the principal should play on the stake $\Delta\theta\Delta q$ not only because it helps to reduce the cost of the collusion-proofness constraint (18) but also because it affects indirectly the transaction costs of side-contracting. Formally (18) should now be replaced by:

$$s_{11} - s_{12} \geq \tilde{k}(q_{22} - q_{21}, \Delta\theta, r)\Delta\theta(q_{22} - q_{21}) \tag{24}$$

where $\tilde{k}(\cdot) < 1$ and $\tilde{k}(0, \Delta\theta, r) = 0$.

Optimizing (H) with (24) replacing (18), taking into account the indirect effect of the stakes on transaction costs, and comparing the solutions with (16) and (17) yields immediately:

PROPOSITION 3. *The models of soft and hard supervision can be identified provided that transaction costs are stake-dependent. Moreover, endogenous transaction costs are given by:*

$$1 - \tilde{k}(\Delta q, \Delta\theta, r) = \frac{p_{21}}{p_{11} r \Delta\theta \Delta q} ln \left(1 + \frac{p_{11}}{p_{21}}(1 - e^{-r\Delta\theta\Delta q})\right) \tag{25}$$

for any stake Δq. When Δq is small enough, the following approximation holds:

$$\tilde{k}(\Delta q, \Delta\theta, r) \approx \frac{r\Delta\theta}{2}\left(1 + \frac{p_{11}}{p_{21}}\right)\Delta q. \tag{26}$$

It is immediate to observe that $\tilde{k}(0, \Delta\theta, r) = 0$ and $\tilde{k}(\Delta q, \Delta\theta, r)$ goes to one as the stake $\Delta\theta\Delta q$ becomes infinitely large. The assumption of constant transaction costs is only valid for very large stakes. For small stakes, $\tilde{k}(\cdot)$ becomes instead almost linear and collusion becomes easier to enforce when stakes are greater. However, from (25), the efficiency of side-contracting exhibits decreasing returns as the stake becomes larger.

To get a better intuition about the stake-dependence of transaction costs, it is useful to go back to the solution to (P). We showed there that $v_{22} = 0$ and $v_{11} = \Delta\theta q_{22}$. Hence, after some computations one can rewrite the (binding) coalition incentive constraint ensuring that the coalition (θ_2, τ_1) reports the truth rather than (θ_2, τ_2) as:

$$v_{21} = -\frac{1}{r} ln \left(1 + \frac{p_{11}}{p_{21}}(1 - e^{-r\Delta\theta\Delta q})\right). \tag{27}$$

Developing in the neighborhood of 0 for $\Delta\theta$ and Δq small enough, we find that

$$v_{21} = -\frac{p_{11}}{p_{21}}\Delta\theta\Delta q + \frac{r}{2}\frac{p_{11}}{p_{21}}\left(1 + \frac{p_{11}}{p_{21}}\right)\Delta\theta^2\Delta q^2$$

where the first term represents the coalition's payoff under risk-neutrality and the second one is the risk-premium that the principal must pay to the risk-averse supervisor to have him reveal information. As v_{21} is paid with probability p_{21}, the expected extra cost due to risk-aversion becomes then:

$$\frac{r}{2} p_{11}\left(1 + \frac{p_{11}}{p_{21}}\right)\Delta\theta^2\Delta q^2.$$

From Proposition 3, it is also the same expected implementation cost of collusion-proofness as that obtained in a Tirole's model with endogenous transaction costs given by (26) since the principal pays then an expected wage:

$$p_{11}s_{11} = p_{11}\frac{r}{2}\left(1 + \frac{p_{11}}{p_{21}}\right)\Delta\theta^2\Delta q^2.$$

This interpretation pushes further forward the view that the agency costs of delegation are linked to the insurance problem faced by intermediate layers. The cost of inducing collusion-proofness can be interpreted as the risk-premium borne by the firm's owner. The nexus of contracts within the firm suggests to view it as an allocation of risky portfolios in which each agent gets a collection of supervisory and productive tasks associated with different communication channels. This is this view of the firm that we use extensively in the next section.

5. ENDOGENOUS TRANSACTION COSTS AND ORGANIZATIONAL DESIGN

With formula (26) in the background, we can obtain new insights about a number of important issues in organizational design. Faced with the problem of better designing the internal structure of the firm, a principal must figure out how this design affects transaction costs of side-contracting. The internal design of the organization becomes an important tool for reducing the costs of side-contracting because it influences the hidden gaming that agents are playing.

One should stress here the strong parallel of our work with Holmstrom and Milgrom (1991) multi-task agency model. In this latter paper, the unverifiable variables that the principal is willing to induce are the agent's efforts along his different tasks. In our paper, these unverifiable variables are instead the side-contracts which take place along the different communication channels of the firm.

Before investigating those issues it is useful to rewrite the endogenous transaction costs using the example of the particular probability distribution for the states of nature we gave in Section 2. Transaction costs depend now on the size of the collusive stake $\Delta\theta\Delta q$, the precision of the supervisor's signal and his degree of risk-aversion r, namely:

$$\tilde{k}(\Delta q, \Delta\theta, \epsilon, r) = 1 - \frac{(1-\nu)(1-\epsilon)}{\nu\epsilon r\Delta\theta\Delta q}\log\left(1 + \frac{\nu\epsilon}{(1-\nu)(1-\epsilon)}(1 - e^{-r\Delta\theta\Delta q})\right) \tag{28}$$

and

$$\tilde{k}(\Delta q, \Delta\theta, \epsilon, r) \approx \frac{r}{2}\left(1 + \frac{\nu\epsilon}{(1-\nu)(1-\epsilon)}\right)\Delta\theta\Delta q \tag{29}$$

in the limiting case of small stakes. Immediate inspection of these formulae provides:

PROPOSITION 4. *The cost of side-contracting $(1 - \tilde{k})$ within the firm decreases with the supervisor's degree of risk-aversion, r, the precision of his information, ϵ, and the size of the collusive stake, $\Delta\theta\Delta q$.*

The only new insight at this point concerns the role of the information structure on the efficiency of side-contracting. Confirming Crozier's intuition, a more precise signal for the supervisor makes side-contracting easier. A more precise information for the intermediate layer of the hierarchy weakens the ability of the principal to control the coalition of the lower layers. Better information channels within the firm at nodes which are not completely controlled by the principal exacerbates the loss of control along the hierarchy.[24]

These results are similar to those found in Martimort (1997). There, endogenous transaction costs are derived within a model which explicitly relaxes the assumption of perfect enforceability of side-contracting. Collusion between a risk-neutral supervisor and the agent is only enforced through a repeated relationship. When the principal cannot use history dependent contracts to detect coalitional behavior, the long-run equilibrium of the organization can be captured by Tirole's static model provided that transaction costs are correctly endogeneized. As in the present analysis, better information channels between the supervisor and the agent foster their collusion. In a repeated game framework when it is relatively costly for the principal to use monetary rewards to fight collusion, i.e., when the principal and the colluding agents have rather similar preferences for the future, collusion becomes rather efficient and k is close to one. Here, the cost of monetary rewards for the principal is instead associated with the supervisor's degree of risk-aversion but a similar logic applies.

5.1. The Role of Uncertainty

(29) shows that a larger spread of the agent's productivity distribution increases the efficiency of side-contracting. Indeed, the amount of risk borne by the supervisor in exerting authority over the agent increases and

[24] Note however that a better information structure has also allocative consequences in the absence of any collusion through its effect on the distribution of rents. Formally q^*_{22} and q^*_{21} change with ϵ even without collusion.

the risk-premium paid by the principal must be raised. As a result, organizations are more likely to be prone to bureaucratic failures in a more uncertain environment. Productive tasks which are relatively uncertain should thus be controlled according to more administrative rules. We state this result as our first principle of organizational design.

PRINCIPLE 1. *Organizations present more bureaucratic features in more uncertain environments.*

5.2. Job Design and Preferences

As we have already noted, our analysis carries over immediately to the case where the agent is also risk-averse. Let us envision now the problem of a principal willing to allocate productive and supervisory tasks between two agents having different degrees of risk-aversion. A second principle for organizational design is the following:

PRINCIPLE 2. *The least risk-averse agent should be given the supervisory task. The most risk-averse agent should be given the productive task.*

By giving the supervisory task to the least risk-averse agent, the principal increases transaction costs of side-contracting without changing the incentive constraints of the productive agent. Principle 2 argues that heterogeneous agents should be ranked along the hierarchy. The most "able" to supervise employees are those to whom it is less costly to provide incentives for exerting authority. Calvo and Wellisz (1979) also showed how most able agents should be given higher positions in the hierarchical tree of the firm. In their efficiency wage model, the technology for detecting shirking by supervisees is given just as the monitoring technology in the present analysis. Wages at all levels of the hierarchy are nevertheless decided in a centralized manner. But the most striking difference with our analysis is the inexistence of any incentive to form coalitions among members of lower layers of the hierarchy. Even when one accounts for this possibility, our analysis suggests that the well-known lesson from the management literature that "most able agents" should be given higher responsabilities still holds once "ability" is understood as their propensity to accept a risky exercice of authority.

Along the same lines, differences in risk-aversion may also explain striking organizational differences between organizations, regardless of any technological consideration.

Principle 1 bis *: Suppose two organizations differ by the degree of risk-aversion of their members. Then, the organization with the more risk-*

averse members will offer a more bureaucratic structure with less powerful incentive schemes.

This theory of bureaucracy and job design within the firm is reminiscent of the results of Kihlstrom and Laffont (1979). In a general equilibrium model, these authors investigate firms' formation. The most risk-averse agents choose to become workers and receive the prevailing riskless wage. The least risk-averse agents choose instead to operate firms and become entrepreneurs. Adopting a general equilibrium perspective here would allow us to ask how agents differentiated by their degrees of risk-aversion will allocate themselves *within* a set of existing firms between supervisory and productive jobs. As agents have to be indifferent between these two jobs for any equilibrium with non-trivial hierarchies to exist, this extension of our set-up should include non-zero reservation utilities both for the supervisor and the agent. Thanks to the absence of wealth effects in our model, these endogenous participation constraints affect the levels of the transfers in the optimal grand- and side-contracts, not the allocative distortions. Transaction costs are therefore independent of the market outside opportunities. Following the logic of Kihlstrom and Laffont (1979), firms will endogenously be formed with less risk-averse agents occupying higher hiearchical levels.

5.3. Span of Control

Since Simon (1957), most of the theory of hierarchies assume an exogenous span of control for each supervisory layer.[25] Endogenous transaction costs provide some elements to understand the cost of different supervisory structures associated with various spans of control.

Assume that the organization involves n productive agents, all of them having individual efficiency parameters drawn independently from the same distribution. These agents are supervised by a single risk-averse supervisor endowed with n monitoring technologies providing independent signals correlated with each agent's type. Moreover, there are no productive externalities among agents so that the conditional optimum is a n-replicate of that described in Section 2.4. The risk-averse supervisor faces in fact n independent risks in each of his bilateral relationships with the agents.

Our vision of the firm as a set of risky portfolios shows immediately that the risk-premium associated with the sum of n independent risks is the sum of the risk-premia associated with each of them when those risks are small. Extending this logic in our context, a principal willing to prevent collusion between the supervisor and the n-agents pays the same amount as if he was facing n supervisor-agent pairs. To put it differently, the transaction

[25]See Williamson (1967), Calvo and Wellisz (1978) and (1979) and Qian (1994) in managerial models and Keren and Levhari (1983) in models of the firm as a communication network.

costs of side-contracting between each agent and the unique supervisor does not depend on the presence of other side-contracting opportunities for this supervisor.

The cost of supervising n-unrelated agents increases linearly with n. There are constant returns to scale in supervision and the design of the exact supervisory structure remains undetermined if the prevailing market wage of supervision (the fixed cost of hiring a supervisor) is normalized at zero.

When agents face positively correlated productive shocks, the risk-premium of the sum of these risks is greater than the sum of the corresponding risk-premia. Hence, still extending the logic of our previous arguments, transaction costs of side-contracting are lower in this positively correlated environments when a single supervisor controls all agents.[26] In negatively correlated environments, the reverse logic holds. Transaction costs of side-contracting are larger when the supervisor has a larger span of control. Negatively correlated shocks allow the supervisor to get some insurance against the shocks which affect each worker. Again, we state this observation as a principle:

PRINCIPLE 3. *Supervision of agents with positively (resp. negatively) correlated shocks should be split (resp. merged).*

A well-known observation is that firms having chosen an M-form organization along product lines better perform than their U-form counterparts which have instead chosen to be organized along functions like finance, market designs, production, sales. Chandler (1962) describes the numerous incentive impediments that large firms like Ford or Dupont which were organized under the latter form were facing at the turn of the century. Bureaucratic rules, delays in project completion, inefficiencies in dealing with new environments, unclear measures of performances were common features of those organizations and impeded their efficiency.

The usual justification for the choice of the M-form is that creating independent profit centers helps the owners of the firm to compare the unit's performances with that of other similar firms on the market.[27] Endogenous transaction costs provide a rather different explanation: in a U-form, agents under the authority of a given intermediate manager are more likely to face correlated shocks on their cost parameters, finance opportunities

[26] In discussing this extension of our framework, we implicitely assume that the supervisor does not use correlation among the agents' types to compute his optimal side-contract along the lines of Section 3. In other words, the side-contract is implemented with dominant strategies and each side-contracting deal realizes independently of the other.

[27] See Tirole (1988, Chapter 1).

or market demands. This suggests that the U-form faces huge internal incentive problems due to the collusion within functional divisions. On the contrary, in a M-form, the buying and the selling units within a given profit center may face completely uncorrelated shocks. Their collusion with the manager of the profit center is relatively inefficient. The M-form reduces the internal incentive problem within large hierarchies.

As the competitive environment of those firms become more complex, delegated authority is necessary and under delegated authority the choice of an M-form organization indeed reduces bureaucratic rules and restores incentives.

5.4. Vertical Integration

Williamson (1985, Chapter 6) argues that vertical integration improves communication channel but that it is also associated with lower powered incentives than those between separated units.[28] Our model with endogenous transaction costs yields some insights about the bureaucratization of large integrated firms. Let us consider the vertical integration by an upstream buyer of a downstream seller. The upstream firm suffers from separation between ownership and control. The principal, owner of the upstream firm, delegates to a supervisor the control of contracting for the provision of the good bought from a vertically separated seller. Suppose now that this downstream firm is vertically integrated, and that integration is associated with an improvement in the precision of supervisory information. From (29), transaction costs of side-contracting become lower under integration since $k(\cdot)$ increases with ϵ. As suggested by Williamson (1985), vertical integration is then associated with lower powered incentives and more bureaucratic rules.

PRINCIPLE 4. *Vertical integration decreases the cost of internal side-contracting.*

6. CONCLUSION

This paper has offered a theory of supervision which links both the formal and the informal dimensions of supervisor-supervisee relationships in an unified framework. This framework has rich implications derived from the endogeneity of the transaction costs associated with these vertical relationships.

---

[28] Grossman and Hart (1986) argue that integration is unlikely to change the information structure but focus on owner-managed firms. Aghion and Tirole (1997) show instead how ownership affects incentives to acquire information.

The lessons of this paper do not only apply to the internal theory of the firm but also to other organizations with delegated authority like for instance, auditing structures, regulatory hierarchies and the political process. We plan to investigate these issues in future research.

APPENDIX

Proof of the Delegation-Proofness-Principle.

- Denote $m^*(\cdot)$ from $\Theta \times T$ into M_s the best manipulation of messages to the principal by the supervisor, i.e., the solution to (S).

- Consider the new grand-contract $GC^* = GC \circ m^*$ where $s^*(\cdot)$ and $q^*(\cdot)$ from $\Theta \times T$ into $\mathbb{R}$ such that $s^*(\cdot) = s(m^*(\cdot))$ and $q^*(\cdot) = q(m^*(\cdot))$. Then, for the offer GC^*, the best strategy $\tilde{m}^*(\cdot)$ of the coalition is to truthfully report any pair (θ, τ). Suppose indeed it is not the case then $m^* \circ \tilde{m}^*(\cdot)$ would be a better strategy than $m^*(\cdot)$ in the first place when the coalition is offered the grand-contract GC^*. A contradiction. ▮

Proof of Propositions 2 and 3. It is useful to make the following change of variables $z_{ij} = e^{-rv_{ij}}$. The objective function of the principal becomes strictly concave in z_{ij} and constraints (9) to (14) form now a system of linear constraints.[1] Hence, an optimum exists defined by Lagrangean technics. For the moment, assume that the only relevant constraints are (9), (11), (12) and (14). We will check later on that the other constraints are satisfied at the optimum as well as the monotonicity conditions $q_{2j} \geq q_{1j}$. Regarding the monotonicity constraint $q_{22} \geq q_{21}$, adding (12) with the coalition incentive constraint ensuring that (θ_2, τ_2) does not mimic (θ_2, τ_1), namely

$$-\Delta\theta q_{22} - \frac{1}{r} ln\left(\frac{p_{12} + p_{22}e^{-r(v_{22}-v_{12}+\Delta\theta q_{22})}}{p_{12}+p_{22}}\right) \geq$$
$$-\Delta\theta q_{22} - \frac{1}{r} ln\left(\frac{p_{12} + p_{22}e^{-r(v_{21}-v_{12}+\Delta\theta q_{22})}}{p_{12}+p_{22}}\right)$$

yields

$$\frac{p_{11} + p_{21}e^{-r(v_{22}-v_{12}+\Delta\theta q_{22})}}{p_{12} + p_{22}e^{-r(v_{22}-v_{12}+\Delta\theta q_{22})}} \geq \frac{p_{11} + p_{21}e^{-r(v_{21}-v_{11}+\Delta\theta q_{21})}}{p_{12} + p_{22}e^{-r(v_{21}-v_{11}+\Delta\theta q_{21})}}.$$

Since $f(x) = \frac{p_{11}+p_{21}x}{p_{12}+p_{22}x}$ is decreasing in x when $p_{11}p_{22} - p_{12}p_{21} > 0$, this amounts to $v_{22} + \Delta\theta q_{22} \geq v_{21} + \Delta\theta q_{21}$. Inserting into the coalition incentive

[1] See Grossman and Hart (1983) for a similar trick in a moral hazard context.

constraint ensuring that (θ_2, τ_2) does not mimic (θ_2, τ_1) yields:

$$r\Delta\theta(q_{22} - q_{21}) \geq ln\left(\frac{p_{11} + p_{21}e^{-r(v_{21}-v_{11}+\Delta\theta q_{21})}}{p_{11} + p_{21}e^{-r(v_{22}-v_{12}+\Delta\theta q_{22})}}\right) \geq 0.$$

Rewriting constraints (9), (11), (12) and (14) with this change of variables gives:
(9) becomes

$$z_{11} = z_{12}, \tag{A.1}$$

(11) becomes

$$z_{11} \leq z_{22}e^{-r\Delta\theta q_{22}}, \tag{A.2}$$

(12) becomes

$$p_{21}(z_{22} - z_{21}) - p_{11}z_{11}(e^{r\Delta\theta q_{21}} - e^{r\Delta\theta q_{22}}) \geq 0 \tag{A.3}$$

(14) becomes

$$p_{12}z_{12}e^{r\Delta\theta q_{22}} + p_{22}z_{22} \leq p_{12} + p_{22}. \tag{A.4}$$

For a given schedule of outputs to be implemented, the principal's problem is

$$\min_{v_{ij}} \sum_{i,j} p_{ij}v_{ij} \Leftrightarrow \max_{z_{ij}} \frac{1}{r}\sum_{i,j} p_{ij}ln(z_{ij})$$

subject to (35) to (38).

Let us introduce $s \in [0,1]$ the possible slack in (36): $z_{11} = sz_{22}e^{-r\Delta\theta q_{22}}$. For a given s, we easily check that the maximization of the above problem implies that (36), (37) and (38) are binding. From that, we get

$$\begin{aligned} z_{22}(s) &= \frac{p_{12} + p_{22}}{p_{12}s + p_{22}}, \\ z_{12}(s) &= z_{11}(s) = \left(\frac{p_{12} + p_{22}}{p_{12}s + p_{22}}\right) se^{-r\Delta\theta q_{22}} \end{aligned}$$

and

$$z_{21}(s) = \left(\frac{p_{12} + p_{22}}{p_{12}s + p_{22}}\right)\left(1 + \frac{p_{11}}{p_{21}}s(1 - e^{-r\Delta\theta(q_{22}-q_{21})})\right).$$

Define $W(s) = \frac{1}{r}\sum_{ij} p_{ij} \log z_{ij}(s)$. It is immediate to derive

$$W'(s) = \frac{-p_{12}}{p_{12}s + p_{22}} + \frac{p_{12}+p_{11}}{s} + \frac{p_{21}p_{11}(1-e^{-r\Delta\theta(q_{22}-q_{21})})}{p_{12}+p_{11}s(1-e^{-r\Delta\theta(q_{22}-q_{21})})}.$$

Given the monotonicity condition, $q_{22} \geq q_{21}$, the last term is positive. The first two terms reduce to

$$\frac{\rho s + (p_{12}+p_{11})p_{22}(1-s)}{s(p_{12}s+p_{22})} \geq 0$$

for all $s \in [0,1]$. Hence $W(\cdot)$ is increasing in s and $s = 1$ maximizes the principal's objective. (35) is also binding. We thus obtain $v_{11} = v_{12} = \Delta\theta q_{22}$, $v_{22} = 0$ and $v_{21} = -\frac{1}{r}ln(1 + \frac{p_{11}}{p_{21}}(1 - e^{-r\Delta\theta(q_{22}-q_{21})}))$.

• Inserting these values of $z_{ij}(1)$ into the principal's objective and optimizing with respect to q_{ij} yields (15) to (17).

• We prove that $q_{22}(r) > q_{21}(r)$ for all r. Consider the solutions to equations (16) and (17). For $r = 0$, $q_{22}(0) = q^*_{22}$ and $q_{21}(0) = q^*_{21}$ since $\rho > 0$.

Moreover, differentiating (16) and (17) w.r.t. r yields:

$$R''(q_{22})\frac{dq_{22}}{dr} = \frac{p_{21}p_{11}(p_{21}+p_{11})\Delta\theta^2 e^{-r\Delta\theta(q_{22}-q_{21})}}{p_{22}\left(p_{21}+p_{11}(1-e^{-r\Delta\theta(q_{22}-q_{21})})\right)^2}\left(r\left(\frac{dq_{22}}{dr}-\frac{dq_{21}}{dr}\right)+q_{22}-q_{21}\right), \tag{A.5}$$

and

$$R''(q_{21})\frac{dq_{21}}{dr} = -\frac{p_{22}}{p_{21}}R''(q_{22})\frac{dq_{22}}{dr}.$$

Hence $\frac{dq_{21}}{dr}$ and $\frac{dq_{22}}{dr}$ have opposite signs. Suppose that $\frac{dq_{22}}{dr} > 0$ and $\frac{dq_{21}}{dr} < 0$, then since $q_{21}(0) < q_{22}(0)$ it is true that $q_{21}(r) < q_{22}(r)$. Inserting into (39) and taking into account that $R''(\cdot) < 0$, we get $\frac{dq_{22}}{dr} < 0$. A contradiction. Hence $\frac{dq_{22}}{dr} \leq 0$ and $\frac{dq_{21}}{dr} \geq 0$.

Suppose that there exists $r^* > 0$ such that $q_{22}(r^*) = q_{21}(r^*)$. Inserting into (16) and (17), we obtain $q_{22}(r^*) = q^*_{22}$ and $q_{21}(r^*) = q^*_{21} < q^*_{21}$, a contradiction. hence, $q_{22}(r) > q_{21}(r)$ for all r and there is no bunching along the τ-dimension.

• Since $q_{21}(r) < q_{22}(r) < q^*_{22} < q_{11}(r) = q_{12}(r)$, other monotonicity conditions are satisfied.

• As $q_{22} \geq q_{21}$, $e^{-r\Delta\theta(q_{22}-q_{21})}$ is less than 1. Consequently, $q_{22}(r)$ is bounded below by q_2^* and $q_{21}(r)$ is bounded above by q_2^*. As $q_{22}(r)$ (resp. $q_{21}(r)$) is decreasing (resp. increasing) it converges towards $q_{22}(\infty)$ (resp. $q_{21}(\infty)$) as r goes to infinity. Suppose that $q_{22}(\infty) \neq q_{21}(\infty)$. Then going to the limit into (16) and (17) we obtain

$$R'(q_{21}(\infty)) = \theta_2 < R'(q_{22}(\infty)) = \theta_2 + \left(\frac{p_{11}+p_{12}}{p_{22}}\right)\Delta\theta.$$

A contradiction since $R''(\cdot) < 0$. Hence, necessarily $q_{21}(\infty) = q_{22}(\infty) = q_2^{sb}$.

• Finally, we have to check that the neglected coalition incentive constraints are satisfied. (10) is immediate as (9) and (11) are binding and $q_{21} \leq q_{22}$ and $v_{21} < v_{22}$.

Let us check that a coalition (θ_2, τ_2) does not want to mimic a (θ_2, τ_1) coalition. This constraint writes as:

$$\begin{aligned} &p_{12}V(v_{12} - \Delta\theta q_{22}) + p_{22}V(v_{22}) \geq p_{12}V(v_{12} - \Delta\theta q_{21}) + p_{22}V(v_{21}) \\ &\Leftrightarrow 0 \geq -\rho V(\Delta\theta(q_{22} - q_{21})) \end{aligned}$$

which is always satisfy with positive correlation.

We also need that a coalition (θ_2, τ_j) does not want to mimic a (θ_1, τ_j'):

$$\begin{aligned} &p_{1j}V(v_{1j} - \Delta\theta q_{2j}) + p_{2j}V(v_{2j}) \geq p_{1j}V(v_{1j} - \Delta\theta q_{1j'}) + p_{2j}V(v_{1j} - \Delta\theta q_{1j'}) \\ &\Leftrightarrow 0 \geq (p_{1j} + p_{1j})V(\Delta\theta(q_{22} - q_{1j'})) \end{aligned}$$

which is always the case as $q_{1j'}$ is the first best output level. ▮

REFERENCES

Aghion, P. and J. Tirole, 1997, Formal and real authority in organizations. *Journal of Political Economy* **105**, 1-29.

Barnard, C., 1938, *The Functions of the Executive.* Harvard University Press, Cambridge.

Baron, D. and D. Besanko, 1992, Information, control, and organizational structure. *Journal of Economics and Management Strategy* **1**, 237-275.

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

Calvo, G. and S. Wellisz, 1979, Hierarchy, ability, and income distribution. *Journal of Political Economy* **87**, 991-1010.

Chandler, A., 1962, *Strategy and Structure.* MIT Press, Cambridge.

Crémer, J. and R. McLean, 1988, Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica* **56**, 1247-1258.

Crozier, M., 1963, *Le Phénomène Bureaucratique.* Editions du Seuil, Paris.

Cyert, R. and J. March, 1963, *A Behavioral Theory of the Firm*. Englewood Cliffs, N.J.: Prentice-Hall.

Dalton, M., 1959, *Men Who Manage*. Wiley and Sons, New-York.

Felli, L., 1997, Preventing collusion through delegation. Mimeo STICERD, London.

Fehr, E., S. Gachter, and G. Kirchsteiger, 1997, Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica* **65**, 833-860.

Galambos, L., 1995, The authority and responsibility of the chief executive officer: Shifting patterns in large U.S. enterprises in the twentieth century. *Journal of Industrial Change* 187-203.

Gouldner, A., 1961, The norm of reciprocity. *American Sociological Review* **25**, 161-179.

Grossman, S. and O. Hart, 1983, An analysis of the principal-agent's problem. *Econometrica* **51**, 7-45.

Grossman, S. and O. Hart, 1986, The costs and benefits of ownership: A theory of lateral and vertical integration. *Journal of Political Economy* **94**, 691-719.

Holmstrom, B., 1979, Moral hazard and observability. *Bell Journal of Economics* **10**, 74-91.

Holmstrom, B. and P. Milgrom, 1991, Multi-task principal agent analysis: Incentive contracts, asset ownership, and job design. *Journal of Law, Economic, and Organization* **7**, 24-52.

Itoh, H., 1993, Coalitions, incentives, and risk sharing. *Journal of Economic Theory* **60**, 410-427.

Jensen, M. and W. Meckling, 1976, Theory of the firm: Managerial behavior, agency costs, and capital structure. *Journal of Financial Economics* **3**, 305-360.

Keren, M. and D. Levhari, 1983, The internal organization of the firm and the shape of average costs. *Bell Journal of Economics* **14**, 474-486.

Kihlstrom, R. and J. J. Laffont, 1979, A general equilibrium entrepreneurship theory of the firm based on risk-aversion. *Journal of Political Economy* **87**, 719-748.

Kofman, F. and J. Lawarrée, 1993, Collusion in hierarchical agency. *Econometrica* **61**, 629-656.

Laffont, J. J., 1990, Analysis of hidden gaming in a three level hierarchy. *Journal of Law, Economic, and Organization* **6**, 301-324.

Laffont, J. J. and D. Martimort, 1995, Separation of regulators against collusive behavior. *Mimeo* IDEI Toulouse.

Laffont, J. J. and D. Martimort, 1997a, The firm as a multicontract organization. *Journal of Economics and Management Strategy* **6**, 201-234.

Laffont, J. J. and D. Martimort, 1997b, Collusion under asymmetric information. *Econometrica* **65**, 875-911.

Laffont, J. J. and D. Martimort, 1998, Collusion and delegation. Forthcoming *Rand Journal of Economics*.

Laffont, J. J. and M. Meleu, 1997, Reciprocal supervision, collusion, and organizational design. *Scandinavian Journal of Economics* **99**, 519-540.

Laffont, J. J. and J. Tirole, 1993, *A Theory of Incentives in Procurement and Regulation*. MIT Press, Cambridge.

Martimort, D., 1997, The life cycle of regulatory agencies: Dynamic capture and transaction costs. *Mimeo* Toulouse.

Maskin, E. and J. Tirole, 1990, The principal-agent relationship with an informed principal 1: Private values. *Econometrica* **58**, 379-410.

McAfee, P. and J. McMillan, 1995, Organizational diseconomies of scope. *Journal of Economics and Management Strategy* **4**, 399-426.

Melumad, N., D. Mookherjee, and S. Reichelstein, 1995, Hierarchical decentralization of incentive contracts. *Rand Journal of Economics* **26**, 654-692.

Mirrlees, J., 1976, The optimal structure of incentives and authority within an organization. *Bell Journal of Economics* **7**, 105-131.

Prendergast, C. and R. Topel, 1996, Favoritism in organizations. *Journal of Political Economy* **104**, 959-978.

Qian, Y., 1994, Incentives and loss of control in optimal hierarchy. *Review of Economic Studies* **61**, 527-544.

Roethlisberger, F. and W. Dickson, 1947, *Management and the Worker*. Harvard University Press, Cambridge.

Rotemberg, J., 1994, Human relations in the workplace. *Journal of Political Economy* **102**, 184-777.

Shavell, S., 1979, Risk sharing and incentives in the principal and agent relationship. *Bell Journal of Economics* **10**, 55-73.

Simon, H., 1957, *Models of Man*, Wiley, New-York.

Tirole, J., 1986, Hierarchies and bureaucracies: On the role of collusion in organizations. *Journal of Law, Economic, and Organization* **2**, 181-214.

Tirole, J., 1988, *The Theory of Industrial Organization*, MIT Press.

Tirole, J., 1992, Collusion and the theory of organizations. In *Advances in Economic Theory*. Edited by J. J. Laffont. **Vol 2**, 151-206. Cambridge University Press, Cambridge.

Williamson, O., 1967, Hierarchical control and the optimal firm size. *Journal of Political Economy* **75**, 123-138.

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