

Chapter 2

THE THEORY OF THE FIRM

BENGT R. HOLMSTROM

Yale University

JEAN TIROLE*

Massachusetts Institute of Technology

Contents

1. Introduction	63
2. The limits of integration	65
2.1. Incomplete contracts	66
2.2. Information flows and incomplete contracts	74
2.3. Reputation	76
2.4. Concluding remark	78
3. Capital structure	78
3.1. The incentive argument	79
3.2. The signalling argument	82
3.3. The control argument	85
3.4. Concluding remark	86
4. The separation of ownership and control	86
4.1. Internal discipline	87
4.2. Labor market discipline	94
4.3. Product market discipline	95
4.4. Capital market discipline	97
4.5. Implications for firm behavior	101
4.6. The hazards of a black-box view of the firm	104
4.7. Concluding remark	105
5. Internal hierarchies	106
5.1. Hierarchies as information systems	107
5.2. Supervision in hierarchies	110

*Support from NSF and the Sloan Foundation is gratefully acknowledged. We would like to thank Jeff Borland, Joel Demski, Bob Gibbons, Oliver Hart, Paul Milgrom, John Moore and Richard Schmalensee for very helpful comments on an earlier draft.

5.3.	Hierarchies as incentive structures	114
5.4.	Hierarchies as internal labor markets	117
5.5.	The hierarchy as a nexus of contracts	120
5.6.	The hierarchy as an authority structure	123
5.7.	Organizational forms	125
6.	Conclusion	126
	References	127

1. Introduction

The theory of the firm has long posed a problem for economists. While substantial progress has been made on the description and analysis of market performance, firm behavior and organization have remained poorly understood. Typically, the firm has been treated in no more detail than the consumer; indeed, the standard textbook analysis of production corresponds closely to the analysis of consumption. In the light of scale differences, equal treatment is plainly peculiar. The volume of trade within firms is probably of the same order as market trade. Large firms are substantial subeconomies of their own with thousands of participants. This alone warrants more attention to non-market modes of transaction.

The nature of decision-making within firms is of a different kind than individual choice in markets. Firm members act as agents for their superiors rather than themselves. In the aggregate, firm behavior is the result of a complex joint decision process within a network of agency relationships. One can justly ask what forces ensure that the process will maximize profits as postulated in the neoclassical theory. Thus, the question of firm organization is not an independent appendix to value theory. It could well have ramifications for market analysis.

Yet another reason for studying firms – perhaps the most important one – is that firms have, as ever-developing institutions, played a central role in the growth and prosperity of a country's economy. In tandem with technological innovations, innovations in firm organization (as well as other institutions) have enhanced welfare greatly. It would seem essential to understand the underlying forces behind such institutional dynamics, both for a proper appreciation of how institutions have conditioned economic development and for policy decisions that relate to institutional change. To analyze institutional legislation purely from a market perspective, as has commonly been the case (cf. anti-trust analysis), is narrow at best.

It is our purpose to discuss analytical models of the firm that go beyond the black-box conception of a production function. Today economists are groping for a deeper understanding based on a contractual view. The firm is seen as a contract between a multitude of parties. The main hypothesis is that contractual designs, both implicit and explicit, are created to minimize transaction costs between specialized factors of production. This follows Coase's original hypothesis that institutions serve the purpose of facilitating exchange and can best be understood as optimal accommodations to contractual constraints rather than production constraints.

The premise that institutions are optimal solutions to various exchange programs warrants a comment. The approach assumes rationality of a high order.

How an efficient arrangement will be found is rarely if ever detailed. Yet, it is easy to envision problems with locating organizational improvements, because of substantial externalities in experimenting with new organizational forms. Few things are as easy to imitate as organizational designs. Information is a public good and patents that would prevent imitation have to our knowledge never been awarded.¹ The fact that organizational innovations often look like fads (witness today's take-over rush) is evidence in point. These doubts notwithstanding, the Coasian postulate lends substantial discipline to the methods of organizational analysis. It is an empirical matter to find out how closely the predictions line up with evidence and, if necessary, to elaborate later on the detailed processes of organizational change and the possible problems that informational externalities present.

A prime source of transaction costs is information. For technological reasons it pays to have people become specialized as specialization vastly expands the production potential. But along with specialization comes the problem of coordinating the actions of a differentially informed set of experts. This is costly for two reasons. Processing information takes time and effort even when parties share organizational goals. More typically, individuals have differing objectives and informational expertise may permit them to pursue their own objectives to the detriment of the organization as a whole. The organization must succeed in capturing the returns from informational expertise by alleviating the exchange hazards that inevitably accompany asymmetric information.

Consequently, much of recent analytical work on organizations has centered on an improved understanding of how one goes about contracting when people know different pieces of information of relevance for the organization as a whole. With the advent of information economics in the early 1970s, the door was opened for these studies. Our survey is chiefly directed towards reporting on the progress of these research efforts.

Oliver Williamson, Chapter 3 in this Handbook and elsewhere, has discussed at length the transaction cost point of view and some of its ramifications. Our efforts are complementary. Analytical models that attempt to articulate contractual problems are useful insofar that they succeed in offering a firmer test of our intuition and logic. They are not meant as competing alternatives to less formal theorizing, but rather as supportive exercises. For those looking for a broader view of the firm we recommend reading Williamson's chapter as well as the related chapter by Martin Perry (Chapter 4) on vertical integration.

With all young and immature fields of inquiry, a survey is made difficult by the limited vision and generality of the initial research efforts. This is particularly true when it comes to modelling the firm. The theory of the firm addresses a wide

¹Problems with appropriating the returns from organizational innovations are somewhat alleviated by consulting firms.

range of questions. At the highest level of aggregation, one is interested in the firm's behavior towards markets. From there one goes down all the way to individual labor contracts and the organization of work in the smallest units of production. Obviously, no single model or theory will capture all elements of the puzzle. Nor is it clear where one most appropriately begins the analysis. As a consequence, modelling efforts have been all over the map, often with more attention paid to the methodological side than to the economic side of the analysis.

Trying to organize these fragments of a theory into a coherent economic framework is difficult. Indeed, an easier task would have been to present the material either chronologically or from a methodological point of view. There has been a distinct line of development in modelling approaches. But, we have tried to face the challenge of looking at present models from the perspective of issues rather than methodology. We hope this will reveal gaps in the overall structure of research on the firm and thereby direct future efforts. Our discussion will not be instructive for those seeking to learn about methods and techniques. For a more methodological perspective on much the same material the reader may find the survey paper by Hart and Holmstrom (1987) useful. [See also Chapter 24 in this Handbook by David Baron on optimal regulation regimes, which contains a detailed discussion of related modelling techniques, as well as Caillaud et al. (1988).]

The chapter is organized around four issues. The first concerns the limits and nature of firms. What determines a firm's boundaries and what explains its existence? The second issue is the financing of firms. What determines a firm's capital structure? The third issue concerns the role of management. How does separation of ownership and control affect a firm's objectives? The last issue is the internal organization of the firm. How is the firm's hierarchy structured and what are the rules of decision-making and the nature of rewards within that hierarchy?

Needless to say, these four issues are interrelated and strains arise when one tries to deal with them separately. Moreover, many models, being so abstract and methodologically oriented, say a little about all the issues rather than a lot about just one. The reader will encounter, if not the same arguments, at least very similar ones in separate places of the chapter.

2. The limits of integration

What is the purpose of firms and what determines their scale and scope? These are two basic questions that a theory of the firm must address. Yet, satisfactory answers have proved very difficult to come by. The challenge is to offer a genuine trade-off between the benefits and costs of integration. One needs to explain both

why firms exist as well as why all transactions are not organized within a single firm. While it is relatively easy to envision reasons for integration, it is substantially harder to articulate costs of increased size.

Williamson (1975, 1985) has phrased the problem sharply. He asks why could one not merge two firms into one and, by *selective intervention*, accomplish more in the integrated case than in the decentralized case. In other words, let the two firms continue as before and interfere (from the top) only when it is obviously profitable. The fact that there are limits to firm size must imply that selective intervention is not always feasible. Trying to figure out why this is so provides a useful focus for theorizing about the nature of the firm.

Traditional theories of firm size – beginning with Viner's (1932) classical analysis of long-run average cost curves – are technology based. Scale economies explain concentrated production while minimum average costs determine the optimal size. More substance can be added by specifying particular cost structures. Baumol et al. (1982) offer a considerably extended version of scale choice in their analysis of contestable markets (see Chapter 1 by John Panzar's in this Handbook).² Lucas (1978) and Kihlstrom and Laffont (1979) focus on cost-minimizing allocations of scarce managerial inputs (talent and risk tolerance, respectively), identifying firms with managers. In Geanakopolos and Milgrom (1985) firm size is determined by pairing the benefits of coordination with the costs of communication and acquiring information. There are also dynamic models – beginning with Lucas (1967) – which center on adjustment costs with the objective of explaining finite growth rates rather than absolute limits to firm size. A natural source of adjustment costs is imperfect knowledge either about the technology [Jovanovic (1982), Lippman and Rumelt (1982) and Hopenhayn (1986)] or about worker–job matches [Prescott and Visscher (1980)].

These technological models offer interesting insights into the role of firms but none is able to address the costs of selective intervention. They all fail to provide a genuine trade-off between integration and non-integration. This suggests that firm limits are determined by contracting costs. For the rest of this section we will focus on the contractual avenue, with particular attention paid to incomplete contracting.

2.1. Incomplete contracts

We begin with a brief description of Alchian and Demsetz's (1972) theory of the firm, which is contractual, but nevertheless fails to draw clear organizational

²Vassilakis (1985) is another notable contribution. His model provides a theory of vertical integration, derived from the tension between competition for rents and the desire to exploit scale economies.

boundaries, at least as originally envisioned. The same type of failure is common to most early contracting work. The problem can be remedied within the framework of incomplete contracting, which also suggests a reinterpretation of the Alchian–Demsetz theory.

Alchian and Demsetz's theory centers on the incentive problems of joint production. Suppose it takes two workers to perform a given task and assume initially that the workers form a partnership. The design problem amounts to choosing a reward structure for each of the partners. How should the partners divide the proceeds from the joint output? If the inputs can be observed and contracted upon, the answer is simple. Pay one the cost of his input and let the other receive the residual. Then it will be in each partner's interest to set input levels in a way that is socially efficient.

But what if inputs cannot be verified so that rewards must be based on joint output alone? This leads to a free-rider problem. There is no way of dividing the joint output in such a way that each worker receives his social marginal product in equilibrium. To see this, suppose the technology is given as $y = f(a_1, a_2)$, where a_1 and a_2 are the effort levels of the two workers, measured in effort cost units. The efficient choice of effort would occur where the partial derivatives $f_1 = f_2 = 1$. Now, let $s_1(y)$ and $s_2(y) = y - s_1(y)$ be the rules by which the joint output is divided between the two partners. Assume for simplicity that these rules are differentiable. In a non-cooperative equilibrium, workers would choose input levels so that $s'_1 f_1 = s'_2 f_2 = 1$. For this equilibrium to coincide with the efficient choice of inputs, it must be that $s'_1 = s'_2 = 1$. But that cannot be, because $s'_2 = 1 - s'_1$.

The problem is that cheating cannot be detected. Based on joint output alone, either of the two workers could be responsible for a suboptimal outcome. A natural solution would be to introduce some monitoring and this is what Alchian and Demsetz propose. They argue for an organizational change in which a monitor is brought in to measure inputs and mete out appropriate rewards. Of course there may be a problem with the monitor not having the incentive to monitor – after all, there is still joint production, this time with three workers. To solve this dilemma, it is suggested that the monitor is given the residual rights to output. He pays the input factors fixed amounts (contingent upon proper input supply) and keeps the difference. In the tradition of identifying ownership with the rights to the residual income stream, the monitor in this story also becomes the owner of the firm.

The limited extent of partnerships and cooperatives in our economy lends some support to the owner-monitor model, since free-riding could be a big problem in these organizations. The importance of monitoring is also evident quite generally. Firms invest in elaborate control systems joined with complex reward structures, implicit and explicit. Without monitoring, the problem of paying individual rewards so as to equate marginal and social products, even approximately, would be overwhelming.

Yet, the simple story of the owner-monitor has its problems. First, those who do the monitoring in firms are rarely the residual claimants. Except for small entrepreneurial firms, owners hardly monitor management, at least not for the purpose of separating individual contributions. In fact, it has frequently been suggested that one of the major problems with corporate organization is the limited interest it creates for a diverse ownership to monitor management (see Section 4). Second, horizontal mergers are hard to understand from a monitoring perspective. One would have to argue that there are scale economies in monitoring, which seems implausible.³ Third, monitoring is not the distinguishing feature of corporations. Partnerships and cooperatives certainly have supervision as well. One might argue, in line with what was said above, that the distinctive feature of corporations is the separation of ownership and active participation in firm decisions. This point is elaborated on in Holmstrom (1982a), where it is shown that separation (budget-breaking) may be desirable from an incentive perspective.

The main problem, however, is that the monitoring story (as told) does not offer an explanation of firm boundaries. Nothing would preclude the monitor from being an employee of a separate firm with a service contract that specifies his reward as the residual output. Similarly with the workers. They could be monitored and paid as independent agents rather than employees.

The problem with organizational anonymity can be traced to the nature of contracts considered. Most contractual models have the property that contract execution is independent of the institutional setting in which it is placed. (Alchian and Demsetz make a point of erasing the distinction between employment and service relationships.) Contracts which are *comprehensive* in the sense that they will never need to be revised or complemented, are of this kind. If parties can agree today to a contract that needs no changes in the future, then it does not matter what affiliations the parties have and hence where the contract is embedded organizationally. Governance structures become important only insofar as the evolution of the contract varies with the organizational setting.⁴

Williamson (1975, 1985) has argued for a long time that comprehensive contracting analyses are misguided and that an incomplete contract perspective is essential for explaining the relative merits of governance implied by different organizational forms. He emphasizes the problems caused by incomplete contracting in relationships where parties make irreversible investments. His standard paradigm is one in which partners are locked into a relationship ex post

³One paper that does develop the theme of monitoring economies is Diamond (1984). He argues that banks as creditors may perform the task of monitoring more effectively than a diverse ownership. However, the logic is quite different from Alchian and Demsetz's theory.

⁴We use the term comprehensive rather than complete in order to avoid a mistaken association with Arrow-Debreu contracts. Contracts that are not comprehensive are called incomplete, despite a potential confusion with the traditional meaning of incomplete contracts. The term is so widely used that it is likely to resist change.

because of investments that have substantially higher value within the relationship than outside of it. To the extent that one cannot specify *ex ante* how the surplus should be divided between the two, i.e. if one cannot write a comprehensive contract, the division will depend on *ex post* bargaining positions. Bargaining positions in turn will depend on the organizational context. Where relationship-specific investments are large, Williamson argues against the use of market exchange, because parties will fear that they will be unable to appropriate the returns from their investments in an *ex post* non-competitive bargaining environment. Bringing the transaction within a firm will offer safeguards against opportunistic behavior. [See also Klein et al. (1978).]

Grossman and Hart (1986) have sharpened the argument by suggesting that the crucial difference between governance structures resides with their implied residual decision rights. Residual decision rights are those rights to control that have not been explicitly contracted for beforehand. In Grossman and Hart's framework, the allocation of residual decision rights is identified with the ownership of assets. Thus, ownership defines the default options in an incomplete contract.⁵ A transaction within the firm (concerning the firm's assets) is controlled by the owner of the firm (or the manager, if he has been delegated the authority) in those situations where the contract does not specify a unique course of action. In contrast, a market transaction must be resolved through negotiation between relevant asset owners if the contract is incomplete. These two modes of transaction will imply a different division of the surplus from the relationship *ex post* and therefore lead to different levels of investment in relationship specific capital *ex ante*. Let us illustrate this with a simple example.

Example

Consider a buyer and a seller who have signed a contract for exchanging a unit of a good tomorrow at a specified price. They are aware of the possibility that a technological innovation may make the present design redundant, but they cannot foresee the nature of the innovation and hence cannot make the contract contingent on a change in design [e.g. assume there is always a costless but non-improving change in design that the court cannot distinguish from a real change so that it would be fruitless to index the contract merely on a design change; see Hart and Moore (1988) for an analysis of contingent contracting].

⁵Grossman and Hart's definition of ownership is essentially the same as the legal notion, though the law, of course, recognizes a variety of different ownership and institutional modes. For instance, the employment relationship has its own set of defaults that distinguishes it from a sale of services by somebody outside the firm. Simon's (1951) seminal paper on the employment relationship makes a similar observation. More recently, the implications of different legal blueprints on the organization and operation of firms has been elaborated on by Masten (1986b).

Denote the buyer's benefit from a design change by v and the seller's cost of implementing the design change by c . These figures are net of benefits and costs from the present design. The values of v and c are uncertain today; tomorrow their actual values will be realized. Both the buyer and the seller will be able to observe the realized values of v and c , but in order to preclude the possibility of contracting on the realization, assume that the values cannot be verified by a third party.

For concreteness assume there are only two possible values for v : 20 and 40; and two for c : 10 and 30. The buyer can influence the outcome of v by making an unobserved investment today. Let $x = \text{Prob}(v = 40)$ represent the buyer's investment decision and assume the cost is $10x^2$. Similarly, the seller makes a relationship specific investment $y = \text{Prob}(c = 10)$ at a cost $10y^2$.

Ownership determines who has the right to veto a design change. There are three cases of interest. In *non-integration* both sides can block a change. In *buyer-integration* the buyer can implement a change by fiat and in *seller-integration* the reverse is true. In addition one needs to specify what happens to cost and benefit streams under the different regimes. In Grossman and Hart's original analysis benefits and costs were inalienable. In our context it would mean that the seller bears the costs y and c and the buyer bears the cost x and receives the benefit v , irrespective of ownership structure. We will assume instead that v and c (and later x and y as well) get transferred with ownership. In reality most financial streams get transferred. However, one needs to explain why these streams cannot be transferred by contract rather than ownership change. Our argument is that separating the return streams of the productive assets from the decision rights of these assets is not feasible, because the return streams cannot be verified. Put differently, the owner of the asset can use the asset to generate returns for his own benefit, which cannot for reasons of verifiability be appropriated by the owner of the return stream. (For example, a contract that specifies that the buyer pays the costs of the seller if a design change is implemented is subject to misuse by the seller – he can load costs onto the buyer which are unrelated to the design change.) Thus, incomplete contracting explains the joining of decision rights concerning asset use with the title to residual return streams.⁶

Let us first analyze the non-integrated case. There are four possible outcomes for the pair (v, c) . In three of them $v > c$. Assuming that bargaining is costless, these three situations will lead to the implementation of the new design since

⁶In general, of course, parts of the return stream as well as the decision rights can and will be contracted for. Incentive contracts are examples of the first kind, while leasing contracts and delegation of authority are examples of the second kind. Note though that even in these cases there is typically a connection between the right to decide and the financial responsibility for the outcome of the decision.

both sides can observe v and c and implementation is efficient. Only if $v = 20$ and $c = 30$ will the new design not be implemented.

Assume that both sides have equal bargaining power so that it is reasonable to predict an equal division of the surplus from implementation of the new design. For example, if $v = 40$ and $c = 10$, the price of the change will be negotiated so that both sides gain an additional 15 (buyer pays seller 25 for the change). With this rule for dividing the surplus the marginal return to the buyer from investing x is $5y + 5$, where y is the forecasted level of investment of the seller; and symmetrically for the seller. The Nash equilibrium in investment choices will then be $x = 1/3$ and $y = 1/3$, considering the marginal costs of investment: $20x$ and $20y$, respectively. The social surplus, net of investment costs, is $50/9 = 5.6$ for the non-integrated form of organization.

Consider next, buyer integration. The buyer's net return in the second period is $v - c$ by our earlier arguments. The seller will merely cover his labor costs and hence earn zero returns in the second period. Consequently, he will have no incentive to invest in the relationship ($y = 0$). The cost of implementing the new design will therefore equal 30 for certain. The buyer's returns from investing x are $10x - 10x^2$ (if the value of the new design is 40 it will be implemented and the buyer will receive the total gain of 10). Thus, the buyer will choose $x = 1/2$. The social surplus, net of investment costs, is in this case 2.5.

The third case of seller integration is symmetric to the previous one and therefore yields the same social surplus.⁷

Since the buyer and the seller can divide the social surplus in any desired way by a transfer payment in the first period it is reasonable to assume that they will agree on implementing the socially efficient organizational form. We conclude that with the particular parameter values chosen here, buying and selling would be conducted under separate ownership.

The example demonstrates that with incomplete contracts the allocation of residual decision rights via ownership can affect investments in relationship-specific capital and thereby efficiency. In particular, this mode of analysis offers a reason why selective intervention is not possible and therefore why integration may not be desirable. The prerequisite is that initial investments are not contractible and comprehensive contracts are infeasible; it is not possible to sign a contract today that will be effective in all contingencies tomorrow.

In the example, specific conclusions about the desirability of integration obviously depend on parameter values. For instance, suppose the high cost is 11

⁷The fact that the two forms of integration are identical is an artifact of symmetry. In general, they will be different. This is interesting since the literature on integration has commonly taken for granted that it does not matter who takes over whom. There are only two modes: integration and non-integration.

instead of 30. Then buyer integration is best because reducing costs becomes less important than increasing value. Similarly, changes in the costs of relationship-specific investments would affect the optimal design. Such comparative static exercises are rather naive in this overly simple setting, but nevertheless point to possibilities in deriving testable hypotheses.

More interestingly, we note that the organizational design is quite sensitive to the nature of assets involved. In particular, the role of human capital as an inalienable asset is important. The ownership of human capital cannot legally be transferred and hence places particular constraints on contracting. Going back to the example, the assumption was made that investment costs were borne by the investing persons irrespective of ownership structure – in other words, non-transferable human assets were used, and their services could not be compensated for by incentive contracts because of enforcement problems. Incentives could only be affected by a change in ownership of physical assets. However, suppose instead that the investments are financial outlays, necessarily borne by the owner (for reasons explained above). Now the seller-employee under buyer integration would have no objections to incurring those costs, because the money would not be out of his pocket. Consequently, buyer integration (or seller integration) would lead to first-best and be superior to non-integration.

Notice that only this version of the example matches Williamson's vision of the benefits of integration. Here integration does reduce opportunistic tendencies, while in the earlier version it was just the reverse. Apparently the value of integration is quite sensitive to the nature of assets being used for investment as well as to the limitations in contracting that relate to return streams. Our two variations fit the common claim that human capital investment and use is best encouraged by independent ownership, while coordination of capital investments is better accomplished by joint ownership.

Although the example was inspired by Williamson's central theme that ex post contracting hazards distort ex ante investments and that changes in ownership affect outcomes via a change in bargaining positions, we want to stress that this scenario is not the only one in which ownership plays a role. It could also be the case that bargaining costs are affected directly by a change in ownership.⁸ For instance, suppose that information about v and c remains private to the buyer and seller, respectively. Neither can observe the other's parameter value. Also, suppose as before that the nature of the design innovation cannot be envisioned in period one so a mechanism for communicating the private information cannot

⁸Milgrom and Roberts (1987) have emphasized the role of bargaining costs more generally. They note that incomplete contracting need not lead to inefficiencies if bargaining is costless and there is no unobservable specific investment. [See also Fudenberg et al. (1986), Malcolmson and Spinnewyn (1988), Masten (1986a) and Rey and Salanié (1986) on gains to long-term contracting.] Crawford (1986) is an early contribution to the role of bargaining under incomplete contracting. Crawford shows that ex ante underinvestment is not always implied by non-integration. Tirole (1986a) provides reasonable conditions under which underinvestment will occur, for instance, in the context of procurement contracting.

be set up today. Then, assuming that bargaining under asymmetric information is costly (any of a number of models of bargaining deliver this; either through costly delays or through incomplete trading),⁹ we would typically conclude that ownership would matter for the outcome of the bargaining process.

The simplest case is the following. Departing from our earlier parameterization, assume that v is always greater than c . Then integration will always lead to an immediate implementation of the design change. By contrast, non-integration will lead to costly negotiations on how to split the surplus, which carries only social costs and no benefits. Buyer integration is clearly superior, because it will prevent needless delays in decision-making.¹⁰ Presumably, the value of authority is frequently one of resolving conflicting private interests in an expedient fashion. One would expect that authority relationships are more prevalent, the more costly are the delays (as is the case in an army in a wartime economy, or in a complex hierarchy when conflicts between two individuals hinders the proper functioning of the organization).

Let us finally return to the Alchian–Demsetz theory with which we began this section. We have emphasized that changes in ownership may imply inevitable transfers of return streams, because of incomplete contracting. Therefore ownership may be the only means by which proper financial incentives can be provided. Consider joint production in this light. Suppose that one worker's marginal product is more easily (though imperfectly) assessed than the other's. This is not possible in the original Alchian–Demsetz model, because the joint product was assumed observable, which implies that knowing one of the marginal products tells the other. But if we accept that the joint product is not always observed or contractable (actual returns will be revealed only in the long run, say), then the distinction makes sense and ownership will matter. Ownership should go with the input factor whose marginal contribution is hardest to assess (relative to the value of that factor). Reinterpreted this way, the Alchian–Demsetz theory can be read as suggesting that the monitor is the owner because his product is important but diffuse (cf. discussion in Subsections 5.2 and 5.3).

We do not subscribe to this revised version of the monitoring story either, because it is rare to see owners monitor the firm's operations. We believe it is more likely that the contribution of capital is hardest to measure, because capital is easy to misappropriate. Consequently, capital should hold title to the residual return stream. This idea deserves further elaboration. Our main point here is that the allocation of return streams via ownership can be a significant component in understanding which factor becomes the owner. This is overlooked in the model provided in Grossman and Hart (1986). Indeed, the authors stress the importance

⁹See, for instance, Fudenberg and Tirole (1983), Sobel and Takahashi (1983), Cramton (1984) and Admati and Perry (1987).

¹⁰Milgrom and Roberts (1987) make the same observation. The idea goes back at least to Weitzman (1974) who noted that centralized decision-making can be much more effective than decentralization in delivering an urgently needed service, which is known to be socially desirable.

of not confusing return streams with ownership of physical assets. This contrasts with the older property rights literature, which identified ownership expressly with the right to a residual return stream [see De Alessi (1983) for a recent survey].

Our view is that these two definitions really should be subsumed in one: ownership provides residual rights to all kinds of assets, physical as well as financial. The right to a residual return stream is after all nothing more than a right to decide how to spend the firm's money, which has not been explicitly contracted for.

2.2. Information flows and incomplete contracts

Williamson (1985, p. 134) has taken issue with the notion that firms are primarily distinguished by their implied residual decision rights. He wants to place corresponding emphasis on the fact that organizational changes imply concomitant changes in information flows. Certain information that is available at one cost before integration may no longer be available at the same cost after integration. Assuming for the moment that this is true, it is a short step to conclude that organizational design can influence performance, since information is used both in decision-making and in the construction of incentive schemes.

It remains to argue why the set of feasible information systems would depend on organizational structure. Grossman and Hart (1986) expressly take the view that this is not the case. Differences in information flows are endogenously chosen, not exogenously conditioned by the choice of organizational form.

Consider a concrete example. Two publicly traded firms merge. Typically, one of the stocks will be withdrawn from trading as a consequence. This elimination of a variable that is crucial for managerial incentives would seem to validate Williamson's position. But one must really ask why the firm could not continue to trade both stocks. In fact, it can and sometimes it will. When General Motors bought EDS, a new GM-E stock carrying no voting right was created, the value of which was tied to EDS performance.¹¹ Presumably this was done to maintain an independent outside monitor of EDS. As it happens, this arrangement has run into difficulties in a way that suggests why it may be infeasible or ineffective to trade stocks on pieces of a company. EDS and the rest of GM have had a hard time agreeing on transfer prices. Apparently they are trying to resolve contractual disputes arising from an incomplete contract. Indeed, brief thought would suggest that as long as GM has substantial control rights in the transfer of goods – and by definition it will as soon as unspecified contingencies arise – the GM-E stock will to some extent be manipulable by GM. This itself does not render the stock valueless. The stock can be protected by covenants and in the

¹¹We are grateful to Mark Wolfson for bringing this example to our attention.

GM-EDS case it was. However, as soon as covenants are necessary and/or the presence of GM-E stock causes distortions in transfers, the costs of replicating the old pre-merger stock information are higher than before or – more likely – the information simply cannot be replicated.

The loss of a stock measure is but one instance of a change in information flow associated with a transfer of authority. Centralized procurement provides another example. For instance, at GM, managerial compensation at the division level is based to a significant extent on division as well as on corporate profits. Centralized procurement of materials and parts (which is meant to exploit returns to scale in procurement and increase GM's bargaining power with suppliers) has generated little enthusiasm [see Corey (1978)]. This is partly because the division manager loses control over the cost of inputs that represent a non-negligible fraction of his budget. The measure of his performance becomes garbled by centralized procurement.

We conclude that organizational changes affect the cost of information flows. Interestingly, the argument for how this comes about must apparently rely on an incomplete contract somewhere in the chain of logic. Thus, the information paradigm and the incomplete contract paradigm are not competitors at all. Incomplete contracting provides the proper framework in which to discuss implications on information flows due to ownership change.

Milgrom (1988) offers a somewhat related discussion. In looking for limits on firm size, he is led to the idea of "influence costs". By influence costs he means undesired employee activities that are intended to change a superior's actions to the sole benefit of the employee. (The emphasis is on "undesired", since subordinates are continuously invited to influence decision-making in constructive ways.) He argues that non-market organizations are particularly susceptible to influence costs, because of their authority structure and the quasi-rents associated with jobs within the hierarchy. Stringent rules and bureaucratic inflexibility can be understood as rational responses, intended to limit the scope of authority and the returns to influence activity.

An important case of costly influence activity arises, because employees are concerned with their careers. They do things with an eye on how their actions will change their superior's (or other evaluators', like the market's) perceptions about their qualifications for future (more desirable) jobs. Such a career concern model is presented in Holmstrom (1982b). The analysis shows that a labor market, which rationally learns (makes inferences) about ability from past performance, can induce employees to exert excess effort early in their careers as well as force managers to make wrong investment decisions [see also Narayanan (1985) and Stein (1987)]. In order to reduce these costs, it may be desirable to change managerial incentives through explicit contracts [Holmstrom and Ricart i Costa (1986)], limit the manager's exposure in the labor market [Gibbons (1985)] or make other organizational adjustments. The fact that young professionals tend to join established firms before going on their own may be explained, in part, as a

way of limiting market exposure that would otherwise lead to distorting influence activities.

Thus, influence activities are pursued both in markets and in hierarchies; indeed, in all situations where individuals care about their careers. As information flows unavoidably change with the organizational form (because of contractual incompleteness), the returns from influence activities differ across organizations. This may provide a basis for a partial theory of organizational choice.

2.3. Reputation

Another theory of the firm that takes as its starting point the inability or the cost to sign comprehensive contracts has been offered by MacCaulay (1963), Williamson (1975, pp. 107–108) and most explicitly by Kreps (1984) and Cremer (1986a). In this theory the soul of the firm is its reputation. Reputation is an intangible asset that is beneficial for transacting in environments where one frequently encounters unforeseen contingencies. Reputation offers an implicit promise for a fair or reasonable adjudication process when events occur that are uncovered by contract. The more faith the firm's trading partners have in the firm's ability and willingness to fill in contractual voids in a reasonable (efficient) manner, the lower the costs of transacting. Thus, establishing and nurturing a good reputation is of much strategic significance.

Kreps argues that "corporate culture" is a main vehicle in this process. It serves two purposes: it conditions and synchronizes the employees' behavior in accordance with desirable reputational objectives and it sends a message to its transacting partners, which informs about expectations of the trading relationship. Thus, the firm's corporate culture acts as the language for telling "how things are done and how they are meant to be done".

As an example, IBM's policy not to lay off employees (in the absence of misdemeanor) is part of its corporate culture. It is not a guarantee that comes in the form of a written or even oral contract. It is a principle that has been established by the historic record. This distinction is crucial. If it were a written contract it would not be as flexible. One can imagine that under some yet to be seen event there will be a need to back out of the pattern and lay workers off. With a contract this would trigger expensive negotiations and perhaps lead to a distribution of surplus that if foreseen would interfere with a smooth and efficient employment policy today. With only a principle and an implicit promise, the adjudication process can be less straining and give a division of surplus that is more conducive to efficient trading today. This of course assumes that IBM can be relied upon. It is crucial for IBM to portray an image of reliability by not laying off anybody except under extreme circumstances. In consequence, today's

workers are partly protected by the threat of IBM losing the value of its investment in reputation.¹²

The management of reputation capital is affected by the allocation of decision rights. It is important to note that only those with residual decision rights can establish a reputation. The other parties will simply follow prescribed conditions in the contract, which signal nothing about future intentions.¹³ Thus, parties with significant interest invested in acquiring a reputation should typically be given residual decision rights, assuming that the potential loss of reputation will assure a more efficient and fair adjudication process in the event of the unforeseen.¹⁴ For instance, in transactions between firms and single individuals, one would expect the firm to have the authority to fill contractual gaps if the firm is more visible in the market and transacts more frequently. [Another reason is that the firm has the relevant information; Simon (1951).]

A central ingredient in a reputational theory of the firm is the mechanism for transferring reputation capital from one generation of managers to the next. Both Cremer (1986a) and Kreps (1984) offer overlapping-generations models in which transfers are feasible. They show that there are supergame equilibria in which reputations will be maintained. In Kreps's model, managers own the firm and thereby the title to future income streams. These can be sold to future managers, who buy themselves into a favorable supergame equilibrium and continue to play it. In Cremer's model the reputation asset is not sold explicitly. It is simply the case that new managers enter into the hierarchy over time and become recipients of as well as contributors to the favorable equilibrium returns.

One problem with the reputation story, taken as the defining characteristic of firms, is that it leaves unexplained why firms could not simply be labels or associations that carry the requisite reputation capital. At present the theory does not make a distinction between the firm as a label and the firm as a collection of physical and human capital assets.

Another dimension that deserves elaboration is the joint responsibility for reputation in a firm with many employees. After all, reputations are in the end attached to individuals and their actions. The incentives of individuals not to milk the firm's reputation has not been clarified; it must be the case that somehow the incentives of the stock-holding layer trickles down through the rest of the hierarchy. The internal organization models studied in Section 5 may have something to say about this.

¹²A recent article in the *Wall Street Journal* (8 April 1987) provides a corroborative account of IBM's corporate culture.

¹³Note also that reputation can only be built if explicit contracting is costly or incomplete. Else there would be no cost to defaulting on an implicit promise; one could costlessly continue with an explicit contract after the default.

¹⁴A concern for reputation need not always be good. Managers overly concerned about their reputation may not always be trusted with authority. [See Holmstrom and Ricart i Costa (1986).]

A more technical point is that reputation is viewed as a bootstrap phenomenon; its formalization relies on supergames, which permit many outcomes. Reputation may, but need not, arise in equilibrium. An alternative theory of reputation was offered by Kreps and Wilson (1982) and Milgrom and Roberts (1982). The reputation of a firm in such a model could refer, for instance, to the outsiders' beliefs that the firm's managers may be intrinsically honest or that their cost of renegeing on an implicit contract may be sufficiently high to discourage unfriendly behavior [Hart and Holmstrom (1987)]. The way intrinsic honesty is transmitted would be technological in the second case, and sociological in the first (intrinsically honest managers would only choose successors with the same attitude towards business).

2.4. Concluding remark

In discussing the limits of firm organization we have heavily advertised the incomplete contracting paradigm and the attendant idea of allocating residual decision rights via ownership. It is the only approach we have seen that succeeds in resolving the selective intervention puzzle raised by Williamson. Obviously, the incomplete contracting framework is not in itself a theory of the firm or its limits. It is merely a tool – an extremely useful one – for articulating, in a clear and consistent way, specific hypotheses about the determinants of firm size. It is in this spirit that we offered the elaborations in Subsections 2.2 and 2.3.

The questions of selective intervention and the nature of the firm are of interest in their own right. One can hardly be satisfied with a theory of the firm that does not understand its main purpose. However, one should not overlook the equally important consideration that, eventually, there is a need to tie the logical arguments for firm limits to the empirical facts about size distribution. Industries show systematic differences in this regard. Also, the finding that firms grow roughly in accordance with Gibrat's Law [growth rates are independent of size; see Mansfield (1962) and Evans (1986)] must be attended to. It remains to be seen how successful the incomplete contracting paradigm will be in addressing itself to real data.

3. Capital structure

Work on the capital structure of the firm was paralyzed for two decades upon discovery of Modigliani and Miller's (1958, 1963) famous irrelevance propositions. Their finding that the value of the firm in a frictionless and tax-free capital market was in fact independent of the mix of equity and debt, as well as changes in dividend policy, stunned the profession. Yet, their reasoning was simple. If the firm's value could be changed by altering the financial mix, this would imply a pure arbitrage opportunity. An entrepreneur could come in and purchase the

firm, repackage the return stream to capitalize on the higher value and yet assure himself of the same risk by arranging privately an identically leveraged position.

The elegant logic of arbitrage proved extremely useful in the development of finance in general, but troubling in the context of capital structure. As Ross (1977) notes, if capital structure does not matter, how can one explain the substantial amount of time and resources corporate treasurers and investment bankers spend on decisions concerning financing? And although the empirical evidence on debt-equity patterns is quite inconclusive [see papers in the volume by Friedman (1985)], it is hard to escape the casual impression that regularities do exist both cross-sectionally and over time. Capital structure does not appear to be a matter of indifference, either on the input side or the output side of the decision.

Efforts to introduce a role for financial decision making have focused on challenging the major premise in the MM-logic, namely that the firm's return stream (or more generally the market perception of the return stream) is unaffected by capital structure. Indeed, the basic logic says that no matter how one divides up a *given* return stream, either over time or across states, the total value stays the same, provided that the capital market offers linear pricing of the pieces (which free arbitrage will imply).

But it is quite possible that the return stream itself may be altered by the financial decision. (Social) bankruptcy costs and non-neutral tax treatment provide one line of reasoning which was pursued early on as an amendment to the MM-theory. Taxes favor debt financing, while equity reduces expected bankruptcy costs. However, this trade-off is not compelling, because debt-equity ratios have been a concern much longer than taxes have existed.

We will discuss three more recent theories of capital structure that also turn on the idea that perceived or real return streams are affected by the firm's financing decision. One argument is based on incentive reasoning. The capital structure is part of an incentive scheme for management; if it is changed the incentives for management – and hence the return stream – are changed. A second argument rests on signalling. If the firm (or its management) is better informed about the return stream, then capital structure may signal information and alter market perceptions about future returns. Finally, a third line takes note of the fact that changes in capital structure involve changes in control rights, which in a world of imperfect information and incomplete markets have ramifications for decision-making.

3.1. The incentive argument

Jensen and Meckling (1976) originated the incentive argument. They developed a theory of the firm, with specific emphasis on capital structure, based on the notion that firms are run by self-interested agents. The separation of ownership

and control gives rise to agency costs. Articulating what these agency costs are, gives the theory its operational content.

According to Jensen and Meckling, there are agency costs associated with both equity financing and debt financing. When "outside" equity is issued (equity not held by those in control), it invites slack. If 50 percent of the firm is owned by outsiders, manager-entrepreneurs realize that each wasted dollar will cost them only fifty cents. Cost-reducing activities will not be pursued to the point where social marginal benefits equal social marginal costs; instead, they will be chosen to equalize private benefits and costs, with resulting excess slack. Of course, the less of a claim on the firm that managers have, the weaker will be the incentives to reduce slack. Thus, from a "shirking" point of view, the firm should be fully owned by management with no outside equity at all. To the extent that there is a need for outside capital it should all be in the form of debt.

Having managers own 100 percent of the firm is not efficient for other reasons. First, managers may want to diversify their portfolio for risk-spreading reasons. Second, financially constrained managers need to raise debt to finance a large holding in the firm. But debt financing incurs agency costs as well. Jensen and Meckling elaborate on the traditional theme that debt and equity holders will not share the same investment objectives. Typically, a highly leveraged firm controlled by the equity holder will pursue riskier investment strategies than debt holders would like (because of bankruptcy).

Pitting the agency cost of equity against the agency cost of debt produces the desired trade-off. The optimal capital structure minimizes total agency costs. The debt-equity ratio is set so that marginal agency cost of each category is equalized. Of course, measurement problems are enormous and Jensen and Meckling offer little guidance for quantification. One qualitative prediction they note is that firms with significant shirking problems – ones in which managers can easily lower the mean return by "theft, special treatment of favored customers, ease of consumption of leisure on the job, etc.", for example restaurants – will have little outside equity. On the other hand, firms which can alter significantly the riskiness of the return – for example conglomerates – will, according to Jensen and Meckling, rely relatively more on equity financing.

Obviously, the above account of agency costs is terse. Jensen and Meckling elaborate on alternative safeguards that can limit both types of agency costs. These include monitoring as well as explicit contracting. We will take up some of these arguments in the next section in connection with managerial incentives.

Grossman and Hart (1982) work out a formal agency model with a slightly different emphasis. In their model a professional manager with little or no stake in the firm (presumably because of limited wealth) controls the allocation of funds raised, either through equity or debt, from the capital market. The manager's allocation decision is very simple. He has to decide how much to invest in a project with uncertain returns and how much to spend on himself. Funds

diverted to private consumption should be interpreted as a stream of benefits (perks, status, etc.) that come from investments (or distortions in investment) that are not valued by shareholders.

The manager does not want to spend all the money on himself, because if the firm goes bankrupt and he is fired, he will no longer be able to enjoy the stream of benefits that he has set up for himself. The trade-off is between a higher stream of private benefits versus a higher risk of bankruptcy and a consequent loss of all perks (it is assumed that more funds invested in the real project will lower bankruptcy risk). Since the actual model has only one period, the allocation decision must of course precede the realization of the investment return or else the manager would always take out the residual, leaving nothing for equity owners.

The key point of the paper is that since the manager has to bear bankruptcy costs, debt financing can be used as an incentive device. Debt acts as a bond which the manager posts to assure equity holders that their funds will not be completely misappropriated. The choice of debt is influenced by its incentive effect as well as the risk that the manager will have to carry. Too much debt will imply excessive risk, while too little will encourage fund diversion.

If there is no uncertainty, then the firm must be financed by debt alone, because the manager can pocket all excess returns.¹⁵ A less trivial conclusion is that increased project risk will increase the market value of equity and reduce the market value of debt. Unfortunately, the model analysis is so complicated that much more cannot be said economically.¹⁶

The major shortcoming of these and other incentive arguments is that they beg the question: Why should capital structure be used as an incentive instrument, when the manager could be offered explicit incentives that do not interfere with the choice of financing mode? For an unexplained reason both Jensen and Meckling, as well as Grossman and Hart, assume that the only way to influence the manager is via changes in capital structure. But this is true only if the manager's compensation contract remains fixed. If the contract can be varied,

¹⁵A related argument is in Diamond (1984) and Gale and Hellwig (1985). These papers establish that, if the entrepreneur is the only one who can observe the outcome of the return and therefore appropriate all residual income, then the only feasible investment contract is a standard debt contract. The distinguishing feature of the Grossman–Hart model is that diversion of funds occurs before returns are in.

¹⁶The idea of debt as a bonding device can be exploited in other directions. Jensen (1986) has recently suggested a Free Cash Flow theory of the firm's capital structure, which argues that debt financing reduces managerial incentives to misallocate funds, because it commits management to return cash to the capital market. Thus, leverage lowers agency costs and raises the value of the firm in cases where mismanagement of free cash is a serious concern. Jensen points to the oil industry, which received windfall profits in the wake of the oil crises, as an example. He argues that the restructuring that followed was partly due to a free cash flow problem further aggravated by the paucity of profitable oil exploration projects. In general, declining industries that are being (or should be) milked are likely to face this type of incentive problem.

then one could presumably provide the same incentives under rather different capital structures. Thus, the challenge for future work is this: to explain why changes in capital structure cannot be undone by corresponding changes in incentive schemes. Without a satisfactory answer to this question, the incentive arguments can only be consistent with the MM-propositions they were designed to dispel. This criticism applies equally to the signalling models we turn to next.

3.2. The signalling argument

Several models have been developed that suggest the debt-equity ratio signals information about the return distribution. We begin with a simple model by Leland and Pyle (1977).

Leland and Pyle consider an entrepreneur who has identified a valuable project with an uncertain return. The entrepreneur is better informed about the distribution of returns. For concreteness, let the return be $x = \mu + \theta$, and assume that $E(\theta) = 0$ and only the entrepreneur knows μ . The structure of the technology and the information is common knowledge.

Because the manager is risk averse and/or because he has limited wealth, he would like to share the project with investors. His problem is to convince investors about the project's true value, μ . Talking does not help. However, a credible communication device is available. The entrepreneur can vary his own stake in the project and use that as a signal of the project's quality.

The formal analysis involves solving for a rational expectations (signalling) equilibrium in which the entrepreneur's *share* of the equity investment (i.e. the ratio of inside equity to outside equity) fully reveals his beliefs about the mean return of the project, μ .¹⁷ Firm debt is determined as the residual amount necessary to finance the project (this could alternatively be private debt to the entrepreneur). For simplicity it is assumed that such debt is riskless. As is typical for signalling models, there is a continuum of equilibria; Leland and Pyle give a selection argument for singling out a particular one. As one would expect, this equilibrium has the property that a higher entrepreneurial share signals a higher project value, μ .¹⁸

The debt-equity ratio is uniquely determined in this equilibrium. It is shown that the value of debt (its face value because it is riskless by assumption) will fall

¹⁷The paper does not consider the possibility that the entrepreneur first takes a large position in order to signal a favorable investment return and then resells the shares. This would complicate the analysis.

¹⁸An interesting feature of the equilibrium is that all projects with positive net present value (accounting for the relevant risk) will be undertaken. This is explained by the fact that a risk averse person is risk neutral for small enough gambles and hence willing to invest a bit in any project with positive net return. Consequently, investment decisions are efficient. The social cost of asymmetric information manifests itself in an inefficient distribution of risk. The entrepreneur will have to carry more risk than he would like to in a world of symmetric information.

with increased risk. Also, an unconditional regression between the value of debt and the value of the firm would reveal a positive correlation; more debt will raise the value of the firm. However, as they are quick to point out, this is not a causal relationship, but rather a statistical property of equilibrium, which comes about because a higher amount of debt goes hand in hand with a higher share of equity held by the entrepreneur. The ratio of debt to equity should not matter in a regression conditioned on the entrepreneur's share.¹⁹ The MM-proposition reappears in a conditional regression.

Myers and Majluf (1984) have analyzed a model closely related to that of Leland and Pyle. The main distinguishing feature is that the firm seeking capital is already established. Its shares are publically traded and its operations are controlled by a manager.

The basic point of the paper is to argue that, because of adverse selection, there are severe problems in raising outside equity. Suppose the market is less informed about the value of shares than the manager of the firm and assume for the moment that there is no new investment to undertake. Then no new equity (from new shareholders) can be raised, if the manager is acting in the interest of old shareholders. He will be willing to issue new shares only if the shares are overvalued, but of course no one would want to buy under those circumstances. Just as in the famous lemon's market of Akerlof (1970), adverse selection will preclude any trade (except in the lowest value state).

Now, suppose capital is needed for an investment. Extending the argument above, Myers and Majluf show that debt financing is preferred to equity financing even when debt is not riskless. Most of the paper, however, focuses on the case where debt is not a feasible option (for reasons outside the model) and new projects have to be financed by issuing equity. This is of course unrealistic, but it leads to an interesting insight. The logic of adverse selection implies that the stock price will always decline in response to a new issue – a result that has empirical support. This may appear paradoxical. How could it be worthwhile to take an action that lowers stock price? The explanation is that a new project is undertaken only if the firm was overvalued *given the manager's private information*. The manager's action is in the best interest of the present shareholders. At the same time it reveals the bad news that the old price was too high in light of his information.²⁰ Another way of reaching the same conclusion is to note that if the share price were to increase with a new issue then it would always pay to raise equity irrespective of the project's value (assuming the proceeds could be rein-

¹⁹However, in this model there are no additional error terms to make such a regression meaningful.

²⁰It is assumed that the market is aware that there is a potentially valuable investment and that no debt is available. Also, the manager has private information about the value of the investment, which varies sufficiently for the decision to be sensitive to this information. If the investment were so good that it would always be undertaken, then issuing equity would not signal any information and the price would remain unaffected.

vested in the market rather than in the project if its net present value is negative). An uncontingent increase in the share price is, of course, inconsistent with market equilibrium.

A major weakness with the signalling approach to capital structure is that the qualitative conclusions are quite sensitive to what is being signalled. If it is the mean of the return distribution, then equity financing is bad news, as discussed above. On the other hand, if the manager's private information pertains to the riskiness of the project (but not the mean) then debt financing would be bad news. Debt would indicate that the variance is high rather than low. It is difficult to build discipline into a theory which depends on something as inherently unobservable as the nature of the information that the manager possesses.

A weakness with the Myers–Majluf model is the treatment of the manager's preferences. One would assume that the manager is driven by his own financial interests, induced by an incentive scheme of some kind, but this dimension is omitted. Ross (1977), who pioneered the signalling approach with Leland and Pyle, was sensitive to this question and went on to study the ramifications of having an endogenously determined managerial incentive scheme [Ross (1978)]. A key observation is that the manager's incentive scheme will signal information jointly with the choice of the firm's capital structure. In fact, the relevant information is really the manager's choice within the set of "securities" that the incentive scheme permits (as the debt–equity ratio is varied). Generally, there are several different pairs of incentive scheme/capital structure that will lead to precisely the same signalling information and the same value of the firm. Thus, very different financial packages could be consistent with the same outcome in reduced form. (This conclusion is partly due to the assumption that managers are risk neutral.)

Ross also shows that (theoretically at least) very rich signals may be communicated through complex managerial incentive schemes. The idea is that, by structuring lotteries that are favorable only in one of the manager's information states, he, as a risk neutral person, can be induced to reveal his precise knowledge. This observation pushes the signalling idea to an extreme conclusion: by constantly changing managerial incentives and capital structure the market can be provided with perfect information. This is obviously unrealistic, but one is then left wondering what determines permanence in incentives and debt–equity ratios. While some form of signalling through debt–equity ratios seems plausible, its strength and relevance is quite open to further research.²¹

²¹A very interesting aspect of signalling arises when there is more than one "audience" who is interested in the signal. This has recently been studied by Gertner, Gibbons and Scharfstein (1987). For instance, signalling that the firm has a high value is valuable for the capital market, but it may lead to more difficult and costly labor negotiations.

3.3. The control argument

The finance literature has traditionally ignored the fact that a share does not merely confer a right to a residual return stream. It also gives a vote. Likewise, loan contracts confer some contingent control rights either implicitly through bankruptcy threats, or explicitly through covenants. As we discussed in Section 2, the distribution of control rights is important for incentives if contracting is incomplete, which certainly is the empirically relevant case. Thus, interest in the distribution of control could well be a key part of the capital structure puzzle. This point has recently been pursued in a paper by Aghion and Bolton (1986).

The ultimate objective of this line of reasoning is to explain why equity and debt are chosen as financing instruments in the first place. The presumption of course is that financing with equity and debt is optimal in some economic environments. It should be noted that optimality in this context refers to more than the nature of the return streams. One also needs to explain why the typical debt contract is linked to a bankruptcy mechanism and the equity contract to a right to run the firm as long as it remains solvent. In other words, one needs to construct a model in which the efficient form of financing is found by maximizing over return streams as well as control rights, with the result that debt and equity – both in terms of their financial and their control characteristics – emerge as optimal.

Aghion and Bolton provide a model which goes some way towards meeting these ambitious objectives. Their primary focus is on explaining features of the bankruptcy mechanism in a debt contract. In a multi-period world they show that it may be optimal to shift control rights to the lender contingent upon unfavorable, publicly observed return information.²² The argument requires a reason for differences in objectives between the lender and the equity holder (without reference to differences driven by the return characteristics of equity and debt, since these could be contractually altered). In one version of the model the difference in objectives comes from different prior beliefs; the lender is pessimistic about future returns contingent upon low intermediate profit reports, while the investor is not. In another version objectives are different because of moral hazard. The difference in objectives explains why the two parties do not simply coinvest in the project using equity, though the return characteristics in Aghion and Bolton's solution need not coincide with those of standard debt and equity contracts.

Two features of their analysis are notable. Their model clarifies the distinction between preferred stock (or non-voting shares) and debt. This would not be

²²Interpreting this control shift as a bankruptcy mechanism overlooks of course many of the intricacies of actual bankruptcy laws. In particular, the firm could file for protection under Chapter 11, permitting management to reorganize the firm.

evident in a model which focused on return streams alone. Also, in their model bankruptcy does not necessarily imply liquidation. In some events in which the lender gets control liquidation occurs, and in other events the lender merely uses his decision rights to reorganize the firm. This accords with reality and is in stark contrast with earlier economic analyses of bankruptcy.

3.4. Concluding remark

The debt-equity ratio has been an enigma in the theory of finance for a long time. As we have discussed, there are models that suggest a role for capital structure based on signalling and screening arguments. These models are not very powerful predictively and consequently have been subjected to little empirical testing. They also have theoretical weaknesses, as we have indicated. The problem is probably that we have not looked deeply enough at the question of capital structure. Rather than taking debt and equity as given instruments, we may get a better understanding of both their role and their determinants by asking why particular instruments are used in the first place. The paper by Aghion and Bolton is a start [see also Grossman and Hart (1987) and Harris and Raviv (1987)]. It seems clear that the most fruitful direction of research at this stage is to pursue further the notion that different instruments imply different control rights. They protect different sources of capital in different ways. In the language of Jensen and Meckling (1976), an optimal capital structure is one that minimizes agency costs, some of which arise from separation of ownership and control, some of which stem from conflicts of interest between different sources of capital.

4. The separation of ownership and control

In reality, firms are mostly controlled by managers. The typical owner will have very little if any influence on the course that the firm takes. Even though there is a formal channel of influence and monitoring through the board of directors, anecdotal evidence suggests that boards rarely take a very active role in running the firm. Also, the choice of directors is often influenced more by management than shareholders. [On these matters, see Mace (1971).]

This raises the question: What keeps management from pursuing its own goals and, if it does, how will the firm actually behave? Some, like Galbraith (1967), are convinced that managerial capitalism (management in effective control of decision-making with few constraints from owners) is a distinct peril for our economy and that the objectives of the firm are far removed from those of a profit-maximizing price taker. That specter may be overly grim. As Alchian (1968) has noted, it is a marvel that millions of people willingly hand billions of their money

over to managers against very limited explicit assurances their investments will be handled responsibly. This could not be going on each day without some strong forces that keep management in check.

We will describe some of the reasons why management may behave despite potential incentive problems. These include the use of explicit incentive schemes as well as the indirect policing forces of the labor market, the capital market and the product market. We will also consider the implications that a managerial theory of the firm has on the objectives that the firm pursues. In this connection we will touch on the more traditional discussions of the objective function of the firm in incomplete markets.

4.1. Internal discipline

Increasing attention is paid to the design of executive compensation plans. Of particular concern are their incentive properties. A good plan should support the strategic objectives of the firm as well as motivate the manager to excel. Contingent compensation constitutes a substantial fraction of a top manager's remuneration. It is not uncommon that over half of the yearly income of an executive derives from stock or option plans and bonus schemes.²³

Principal–agent models offer a more theoretical paradigm within which managerial incentive problems can be studied. In the principal–agent abstraction, owners are viewed as a homogeneous group, a syndicate to use Wilson's (1969) terminology, which can be represented by the preferences of a single person, the principal. The top manager is the agent. The rest of the firm is represented by a stochastic technology, which the manager operates. The manager's compensation scheme is designed by the principal to maximize firm value subject to the constraint that the manager's opportunity cost is covered; or equivalently, the scheme maximizes the manager's expected utility subject to a minimum welfare level for the principal. Either way, the design will be Pareto optimal relative to incentive constraints.

The presumption that the relationship between stockholder and management can be adequately described in a principal–agent paradigm is not innocuous. In the next section we will argue that a stochastic technology of the kind typically used in principal–agent models is generally inadequate for describing the rest of the firm, even if viewed as black-box. More importantly, perhaps, legal scholars, notably Clark (1985), have criticized the agency notion for overlooking the fiduciary nature of management. Both officers and directors are fiduciaries rather

²³The popular press has often questioned the incentive role of stock and option plans, citing evidence that there really is no connection between pay and performance. The data do not support such claims. See Murphy (1984) for a study which indicates that pay and performance are related when all forms of contingent compensation are accounted for properly.

than agents with respect to the corporation and its shareholders. This distinction is important in an incomplete contract framework. For instance, the board of directors – not shareholders – has the right to intervene in the firm's operation. Were the directors agents, the shareholders would retain the ultimate right to control and could, if they wished, impose their preferred policy on the directors and the company.

The independence of directors raises several issues. First, they must be given incentives to exert supervisory effort. Second, they must not collude with the manager and permit him to divert funds for joint benefits. There is substantial evidence [Mace (1971)] that directors have close ties to management and are therefore unlikely to be too critical about inadequate performance. The main option that shareholders have is to sue directors or the management for violating their fiduciary duties (such lawsuits have been more successful recently). Another incentive is that directors are frequently large shareholders of the company (or represent a firm that is a large shareholder). Also, like management, directors may have a reputation to protect. But unlike management, directors are rarely (though sometimes) paid contingent fees for their services. The role of directors as a control layer between the shareholders and management is an important issue that has not been studied theoretically as far as we know.²⁴

These considerations notwithstanding, the principal–agent paradigm is a first step towards modelling how control is exercised in a company and how agency costs are kept within manageable limits.

In order for any managerial incentive problem to arise, it is of course essential that preferences do not coincide. It would seem easy to come up with reasons why a manager would not want to pursue the objectives of owners, say value maximization. The manager may want to divert company funds for private consumption; he may want to expand the business for reasons of prestige; he may cater to the tastes of other stakeholders like employees in order to enjoy an easier life within the organization; he may prefer leisure to work; and so on. Yet, to build a disciplined theory, one cannot formulate models with too much flexibility in the choice of preferences for the manager. One needs to derive his behavior from a narrower set of basic assumptions. For this reason a lot of extant agency models have been based on the notion that the agent is averse both to risk and to work. Aversion to work gives a primitive and obvious reason for incentive problems, but it may not be the most realistic assumption. Managers seem to be quite industrious by inclination ("workaholics"). An alternative, and possibly

²⁴To our knowledge, there also has been little empirical work on the control exercised by the board of directors. An exception is Hermalin and Weisbach (1987) who find that: (i) firms with poor performance tend to add outsiders to the board; (ii) new CEOs put more outsiders on the board; and (iii) large shareholdings of top management are a strong predictor of the proportion of insiders on the board.

more attractive hypothesis, which we will come back to later, is that the manager is driven by concerns for his career and its implied lifetime income stream.

Differences in preferences alone are not sufficient to explain why there is a serious incentive problem with management. One also needs to explain why incentive alignment carries costs. An obvious reason is asymmetric information. Managers are experts who know more about the relevant aspects of decision-making. They also supply unobserved inputs like effort, which cannot be accurately inferred from output. It is the presence of private information that prevents inexpensive contractual solutions and provides a potential opportunity for the manager to pursue his own objectives rather than the owner's.

Let us elaborate on this theme with some examples, which will illustrate the kinds of models that have been analyzed. Suppose the technology is of the form $x = x(a, \theta)$, where x is output, a is the manager's effort and θ is a stochastic term. Assume that the manager is risk and work averse, so effort is costly. Furthermore, assume that both sides agree on the probability distribution of the stochastic term. The manager's effort cannot be observed, nor can it be inferred with certainty from the jointly observed and contractible variable, x . This means that θ cannot be observed either, or else the effort could be inferred from the knowledge of θ and x , assuming x is increasing in a .

An incentive scheme is a sharing rule $s(x)$. The owner's design problem can be viewed as one of instructing the manager to take a particular action, a , and finding a sharing rule that will make the manager obey that instruction. A Pareto optimal design $\{a, s(x)\}$ maximizes the owner's welfare subject to the constraints that the manager gets a minimum level of expected utility and the design pair is incentive compatible, i.e. $s(x)$ induces the manager to choose a .

This is an example of a *moral hazard* problem. Its characteristic feature is that there is symmetric information at the time of contracting. The economic trade-off in the model is between risk sharing and incentive provision. An optimal design will have to compromise between these two conflicting objectives, offering some incentives without exposing the manager to excessive risk. The significance of risk sharing is underscored by noting that if the manager is risk neutral (which implies unlimited access to funds) there is a costless solution: let the manager rent or purchase the technology from the owner.

A common variation of moral hazard is obtained by assuming that the manager observes θ before taking the action a (but after contracting). This enriches the manager's strategic options. His strategy is now a contingent decision rule $a(\theta)$ rather than a single choice a . More options for the manager is bad from the point of view that he is more difficult to control (more incentive constraints), but good from the point of view that information about the technology before an action is taken expands the production set. The net value of information could have either sign.

When there is asymmetric information at the time of contracting, the situation is labeled *adverse selection*. For instance, assume the manager observes θ before he begins negotiating a contract. This case is different than the one just discussed, because the manager's information changes his bargaining position. Now the owner does not know what the manager's reservation utility is and the manager will be able to use this to extract informational rents. One implication is that even with a risk neutral manager, the problem has a non-trivial solution. The rental solution that works for moral hazard does not work here, because the proper rental price is not common knowledge. A more complicated solution, e.g. a royalty scheme, which uses the outcome x as a signal about the value of the technology can reduce managerial rents and be preferred by the owner [Sappington (1984)].

Adverse selection is studied in detail in Chapter 24 by David Baron in this Handbook. Here I will constrain myself to discuss some features of moral hazard solutions, most of which are relevant also for adverse selection.²⁵

In reduced form all moral hazard models have the manager choose a distribution over contractible as well as payoff relevant variables. For instance, in the example introduced above, the manager, by his choice of effort, picks a distribution over output x , induced by the distribution of θ . Note that this is true whether he chooses his effort before or after θ is realized. The feasible set of distributions available if he chooses effort after observing θ is larger, but conceptually the two cases are equivalent. To indicate the dependence of the distribution on effort, one may write the manager's distribution choice as $F(x|a)$ [or $F(x|a(\theta))$]. As an example, if $x(a, \theta) = a + \theta$ and θ is distributed normally with zero mean, then $F(x|a)$ is normal with mean a . The simplest possible case is one in which the manager has only a choice between two distributions, $H(x)$ and $L(x)$. Say, he can work hard or be lazy. What can we say about the optimal contract in that case (assuming that it is desirable to have the manager work hard)?

The solution is quite intuitive. Relative to a first-best contract which provides optimal risk sharing, the manager is paid more the more strongly the outcome x conforms with the view that he worked hard. Conversely, he is paid less if the signal x indicates that he did not work hard. Technically, the optimal sharing rule is a function of the likelihood ratio of the two distributions, i.e. the ratio of the density functions $h(x)/l(x)$. This statistical connection is notable in that no inferences really are made; the principal knows the manager's action given the incentive scheme.²⁶

²⁵Moral hazard models have been analyzed extensively. See, for instance Spence and Zeckhauser (1971), Mirrlees (1974, 1976), Stiglitz (1975), Harris and Raviv (1979), Holmstrom (1979, 1982a), Shavell (1979) and Grossman and Hart (1983). Similar models were earlier studied by Wilson (1969) and Ross (1973). For surveys of agency theory, see MacDonald (1984) and Arrow (1985).

²⁶For a more detailed discussion, see Hart and Holmstrom (1987).

The statistical intuition is in fact the central feature of the basic model. It has both good and bad implications. The most problematic feature is that the shape of the optimal scheme is extremely sensitive to the distributional assumptions, because shape is determined by the likelihood ratio, which varies with the minute informational details of the model. The model can be made consistent with almost any shape of the sharing rule by altering the information technology suitably. The mapping from distribution choices to sharing rules is intuitive, but not useful for explaining regularities about shape. Linear schemes, for instance, are used across a wide range of technologies, so it is clear that they cannot possibly derive from statistical properties of the environment. Yet the simple model could explain linearity only on the basis of the information content of the output signal.

Holmstrom and Milgrom (1987) argue that the failure to explain shape is due to the fact that simple agency models do not capture an important piece of reality: real world schemes need to be robust. It is not enough that a scheme performs optimally in a limited environment. It must also perform reasonably as circumstances change, since constant updating of schemes is not feasible. The schemes that are optimal in simple agency models are fine-tuned to a specific environment. They tend to be complex, because they exploit, unrealistically, every bit of information provided by the output signal.

The great virtue of linear schemes is probably their robustness. They perform well across a wide range of circumstances. They also prevent arbitrage, which often would be possible with non-linear schemes. Robustness is hard to capture in a Bayesian model. Holmstrom and Milgrom show, however, that one can construct Bayesian models in which linear schemes arise out of a richer set of distributional options for the agent than is typically assumed. In particular, they consider a model in which the agent can choose his effort over time, conditioning his choice on how well he has done up to that time. Technically, the agent, who has an exponential utility function over consumption, controls the drift rate of a Brownian motion over a fixed time period. In this environment linear rules are optimal, because they provide the agent with the same incentive pressure irrespective of how he has done in the past. The agent will choose the same level of effort throughout the period and the optimal linear scheme can be solved from a static model in which the agent picks the mean of a normal distribution (constraining the principal to linear rules). Paradoxically, a complex model is needed to provide a simple and computationally tractable solution.²⁷

While simple moral hazard models say little of predictive value about shape because of a strong statistical connection, the statistical intuition is very powerful

²⁷A linearity result in the adverse selection context is found in Laffont and Tirole (1986). [See also McAfee and McMillan (1987), Laffont and Tirole (1987a), Picard (1987) and Chapter 24 by David Baron in this Handbook.]

in predicting what information sharing rules should depend on. The main result states that optimal sharing rules should be based on sufficient statistics about the manager's actions [Holmstrom (1979) and Shavell (1979)]. For instance, suppose that there is a signal y that the parties can contract on in addition to output x . Then y should be included in the contract if and only if x is not a sufficient statistic for the pair $\{x, y\}$ with respect to the manager's action a . The reason is that the likelihood ratio mentioned earlier, which determines the sharing rule, depends both on x and y precisely when x is not a sufficient statistic.

The most interesting implication of the sufficient statistic result relates to relative performance evaluation [Baiman and Demski (1980) and Holmstrom (1982a)]. Managerial performance should to some extent be measured against the competition as well as against general economic circumstances. Performance in a bad year ought to be valued more highly than the same performance in a good year. The rationale for relative comparisons is that they filter out uncontrollable risk. In some cases, the filter is simple. Suppose, for instance, that managerial technologies take the form $x_i = a_i + \theta + \varepsilon_i$, where i is an index for manager i , a_i is his effort, θ is an economy wide shock and ε_i is an idiosyncratic noise term. Then, a weighted sum $\sum \tau_j x_j$, where τ_j is the precision (the inverse of the variance) of the noise term ε_j will be a sufficient statistic if distributions are normal. The optimal scheme for manager i can be based on the difference between his output and this sufficient statistic.

Relative performance evaluation is common in managerial compensation. Promotions are presumably based on relative merit. (The literature on tournaments discusses this type of incentive; see Subsection 5.3). The newest innovations in executive compensation plans also move towards the use of explicit relative measures. Schemes which explicitly compare management performance with competitors are becoming more popular. The fact that competitors constitute the comparison set is in agreement with the notion that closely related technologies are more informative (cf. the weights in the sufficient statistic above). Also, indexed stock options have been introduced. In these the exercise price is contingent on industry or economy wide circumstances.

Antle and Smith (1986) offer more systematic evidence. They study the extent to which executive compensation reflects relative performance, either explicitly or implicitly. Their tests pick up statistically significant evidence that relative evaluations are present. However, the use of relative performance measures is not as extensive as one would expect from the basic agency theory. One reason could be that executives can protect themselves against systematic risk through private market transactions. A more important reason is that relative evaluations distort economic values and thereby decision making. For instance, an executive who is completely insulated from market risk (i.e. whose compensation depends only the firm's deviation from overall market performance) will care little about factors that affect the market or the industry as a whole. This could obviously lead to

very misguided investment or production decisions.²⁸ Effort-based agency models overlook such implications, because they typically do not include investment or production decisions. This is a variation on the earlier robustness theme and suggests that models with a richer action space for the agent would be desirable to explore.

We have been vague about the nature of the performance measure x , except to say that it should incorporate all informative signals. The most natural measure of performance is profit. However, this variable can be garbled by manipulating accounts. Furthermore, current profit is a poor measure of the manager's true performance, which is equal to the increment in the expected present discounted value of profits (which cannot be measured from accounting data). For instance, investments (in capital or reputation) lower the firm's current profit. This brings us to the standard rationale for giving the manager stock options: the firm's valuation ought to incorporate the present discounted value of the investments that are observable by the market, so that the presence of large stock options in the manager's portfolio aligns his and the owners' preferences.

There are limits to the use of stock options. First, the principal–agent theory emphasizes that incentives conflict with insurance: large stock options conflict with the manager's portfolio diversification. Second, stock options do not necessarily create incentives to make investments, that have benefits that are imperfectly observed by the market. To encourage such investments the manager should be forced to hold stock options after his tenure on the job. But this policy has other problems. It creates a free-rider problem of the type discussed in Subsection 2.1: the manager's return on the stock option depends not only on his performance but also on his successor's performance. Also, the firm and the manager have an incentive to renegotiate when the manager leaves his job so as to let the manager sell his stock options and diversify his portfolio (because his investments are sunk, the stock option imposes *ex post* inefficient risk on the manager). The problem implied by overlapping generations of management and delayed performance measurement are important, but have received little attention in the agency literature.

One extension of agency models that deserves comment is dynamics. Some models seem to suggest that repetition will alleviate moral hazard problems. The idea is that repetition will offer better monitoring capabilities. This notion appears substantiated by results that show that in an infinitely repeated agency model (the agent faces the same technology with independent shocks infinitely often) with no discounting, the agency costs are reduced to zero. The first best can be supported as a self-enforcing equilibrium [Rubinstein (1979) and Radner

²⁸Imagine the decision-making of an oil executive, whose compensation neutralizes the effects of changes in oil price.

(1981)]. Moreover, with little discounting one can come close to first best [Radner (1985)].

The interpretation of these results, as due to better monitoring, may be misleading. Fudenberg et al. (1986) show that in a similar repeated model, the optimal long-term contract coincides with the sequence of optimal short-term contracts, assuming that the agent has free access to the capital market.²⁹ Thus, repetition offers no additional gains. What explains the first-best result is that little discounting will offer the agent a degree of self-insurance such that he will behave essentially as a risk neutral person.

4.2. Labor market discipline

Fama (1980) has suggested that the incentive problems of management that are the focus of agency theory may be greatly exaggerated because dynamic effects are ignored. Fama has in mind the disciplining powers of the managerial labor market. He argues that a manager who misbehaves will show a poor performance and consequently his human capital value will deteriorate. The labor market will settle up *ex post* by paying the manager his perceived marginal product, which will reflect past performance. A concern for reputation alone will take care of any deviant incentives. There is no real need for explicit incentives.

This conclusion is optimistic. Holmstrom (1982b) provides a model explicating Fama's intuition. The essential ingredient in the model is that the manager's productivity is unknown. It pays to work hard, because that influences the market's perception about the unknown productivity. Of course, in equilibrium no one is fooled. Instead, the manager is caught in a rat race where he has to work to prevent an adverse evaluation. However, there is no presumption that returns to reputation coincide with periodic returns from output. A manager may well work excessively in the early periods of developing his reputation and slack off later. If productivity is not fixed forever, but subject to periodic shocks, then a stationary equilibrium will support a level of effort that is positively related to the signal-to-noise ratio (how accurately the market can observe output relative to the variance in productivity) and positively related to the discount factor. With little discounting one comes close to first best.

The model shows that the labor market can induce effort without explicit contracts, but that there is little reason to believe that supply will be optimal. More interestingly, if one introduces other decisions like investment, the manager's choice need not be well guided by reputation concerns. In fact, the presence of a

²⁹See also the closely related work by Malcolmson and Spinnewyn (1988). For repeated principal-agent models without free access to the capital market, see Lambert (1983) and Rogerson (1985).

longer horizon may be the very source of divergent investment preferences even assuming that the manager is naturally industrious. The manager will choose investments that maximize his human capital returns (his reputation) while owners want to maximize the financial value of the firm. These two investment returns can be quite unrelated. Depending on the technology and the uncertainties involved, the manager may choose too much or too little risk. For a particular model specification, Holmstrom and Ricart i Costa (1986) show how contracting can align preferences. The optimal contract in this model is an option on the value of the manager's human capital, which in some cases is well approximated by an option on the value of the firm.

It is quite possible that the real problems with managerial incentives derive from the conflicts that arise due to managerial career concerns rather than effort choice as commonly considered in agency theory. This dimension deserves further investigation.

Wolfson (1985) has done an interesting empirical study of the disciplinary powers of reputation. He investigated the market for limited partnerships in oil-drilling ventures. Because of the tax code, which allows limited partners to deduct initial drilling expenses from their income tax, the contract between the general partner and limited partners is designed so that limited partners bear the main exploratory expenses while the general partner bears the main costs of completing the well. Since both share in the returns if a well is completed, the contract gives the general partner an incentive to complete fewer wells than limited partners desire. However, new ventures come up frequently and new partnerships are formed. One would expect this to have an effect on the general partner's behavior, and it does. Wolfson finds that the general partner completes more wells than myopic behavior would dictate. But the reputation effect is not strong enough to remove all incentive problems; Wolfson finds that share prices of limited partnerships reflect residual incentive problems. This accords broadly with the predictions from the reputation model above. The labor market exerts disciplinary influence, but is not sufficient to alleviate all problems.

4.3. Product market discipline

It is an old theme that the real costs of monopoly may derive more from organizational slack than price distortions [e.g. Leibenstein (1966)]. The easy life of a monopolist may be the greatest benefit of running a monopolistic firm. By implication, then, competition will provide discipline and reduce managerial incentives to slack off.

Jensen and Meckling (1976) take exception with this inefficiency hypothesis. They claim that agency costs are no less in competitive industries than in monopolistic industries. Since the easy life is enjoyed by the manager and his

associates rather than the owners, the owners of monopolistic firms should be as interested in curbing agency costs as the owners of competitive firms.

This reasoning misses one important distinction between competitive and monopolistic industries. In the former there is more information about the circumstances in which the manager operates. In line with the rationale for using relative evaluations, competitive markets provide a richer information base on which to write contracts.

The value of competition is obvious if one imagines explicit incentive schemes in which the manager is compared with other firms in his market. We know that relative evaluations will allow some reduction in the uncontrollable risk that the manager has to bear and this will reduce agency costs.³⁰

It is also easy to see that competition can reduce slack via a concern for reputation. For instance, the model in Holmstrom (1982b) has the feature that a sharper signal about performance will automatically lead to an increased level of effort in equilibrium (since effort responds positively to the signal-to-noise ratio). Observing competitors' performance is one way in which signal strength is increased.

A somewhat subtler channel of incentives is provided by the price mechanism. Suppose costs are uncertain but correlated. Then any rule of price formation will carry information about the other firms' costs and thereby be useful as a signal. An incentive contract that uses price as an index would help in reducing agency costs and possibly slack.

Hart (1983a) has developed a model to study a variation of this argument. In his model there is a continuum of firms. Some are run by managers and therefore subjected to control problems, others are run by the owners themselves. The degree of "competition" is measured by the ratio of entrepreneurial firms to managerial firms. Slack occurs, by assumption, only in managerial firms.

The marginal cost of all firms is the same. Managers are rewarded solely as function of their own profits. Price, which in this case reveals fully the marginal cost, is not a contractual variable by assumption. This hypothesis may be hard to rationalize empirically. One could argue that in some cases the industry is so poorly circumscribed that it is hard to identify which price to look at. Also, prices do reflect other variables like quality. Yet, even weakly correlated price signals should be valuable in contracting. The best argument therefore is a methodological one: one wants to study in isolation the indirect effects of competition through profits. In other words, how effective is the price systems as an implicit incentive scheme?

Hart's model has some special features. Managerial effort is a direct substitute for input costs. More critically, managers only care about reaching a subsistence

³⁰ However, risk reduction does not necessarily lead to less managerial slack in all agency models, though this may be a peculiarity of specific models more than anything else.

level of consumption. Consumption above this level has no value; consumption below it is catastrophic. The implication is that managers, who observe input costs before acting, will always work hard enough to achieve a profit level that will allow them to consume the minimum necessary for subsistence. An increase in productivity translates directly into slack.

With this structure it is intuitive that competition will reduce slack. Competition drives price down. If the manager were to slack the same amount as without competition, he would not be able to reach a sufficiently high profit level to collect the minimum reward that he needs. Hence, he has to work harder.

The complete argument is more complicated, because one has to consider changes in the incentive schemes in response to competition. The particular preference structure that Hart uses plays a critical role here. This was pointed out by Scharfstein (1988a), who considered a more standard preference structure. He found that when managers are more responsive to monetary incentives, Hart's conclusion is precisely reversed: competition increases slack. Apparently, the simple idea that product market competition reduces slack is not as easy to formalize as one might think.

4.4. Capital market discipline

Take-overs are presumed to be the ultimate weapon against managerial misconduct. Take-over threats are often suggested as a rationale for the neoclassical assumption that managers will maximize firm value. A naive argument is that if managers do not maximize value, then somebody can take over the firm, install a new value maximizing management and realize an arbitrage profit. Thus, incumbent management can do nothing less than maximize firm value.

Some authors, like Scherer (1980), have questioned the strength of take-overs as a disciplinary device. He notes that take-overs are quite costly and therefore it seems likely that if there is managerial misconduct it has to be substantial before there is an incentive for somebody to intervene.

There are other problems with the take-over argument. Why can present owners not effect the same change as the raider can? Apparently, the value of a take-over must rest either on private information that a raider holds or on special benefits that the raider, but not the shareholders at large, can capture (or perhaps both). Even so, one must ask: Why would the new management behave any better than the old one? If not, why does the old management have to leave? And if it does not, why would take-over threats change the behavior of present managers?

One of the notable contributions of formal take-over models is the discovery that take-overs by an outsider cannot easily be explained by private information alone. Grossman and Hart (1980) provide the following rationale for why such

tender offers might not succeed. Suppose that the raider knows privately how to improve the performance of the firm. For instance, he may have identified a better management team. If he makes a tender offer that will benefit him if he succeeds, then it must be that his gain comes at the expense of those who tendered. The mere knowledge that a tender offer is valuable to the raider should lead present shareholders to conclude that it is not to their advantage to tender. It is a dominant strategy to hold on to one's shares, assuming that these shares are marginal. (It would be different of course if one held so many shares that tendering them could swing the outcome.)

Another way of expressing the take-over dilemma is in terms of free-riding. Present share holders can free-ride on the raider's efforts to improve the firm. The scenario outlined above is the extreme one in which the raider would have to give away all gains in order to take over. Thus, he has no incentive to take over nor to invest any resources in identifying improvements.

So what explains the occasional success of tender offers? There are several possible changes in the simple story. One is that the raider values the firm differently than the present owners in a subjective sense (ignoring private information). He may desire to run the firm. Or some other firms that he owns could benefit from the take-over. Closely related to this is the possibility that the raider could exploit minority shareholders if he succeeds. In fact, it may be in the interest of shareholders to write a charter that explicitly permits such dilution, because that will make take-over easier, encourage raiders to invest effort into identifying poorly run firms and thereby indirectly provide managers with incentives to act in the interest of its ownership (assuming that a take-over is costly to the manager and hence something to be avoided).

The question of the optimal design of dilution rights is precisely what Grossman and Hart focus on. They prove, in a stylized model, that higher rights to dilute will drive the manager closer to maximizing firm value. Take-over threats will act as an incentive scheme as postulated in less formal accounts. Dilution is not costless for present shareholders, because it will lower their return in case a take-over occurs as well as lower the price at which the raider can successfully bid for the company. The trade-off is between better management and a higher frequency of take-overs versus a lower bid price and less residual income. This determines the optimal dilution rights of the charter. Uncertainty about potential benefits as well as costs for the raider are key factors in the calculation.³¹

³¹For two recent, very interesting entries on the role of the corporate charter in influencing take-overs, see Grossman and Hart (1987) and Harris and Raviv (1987). Both papers try to explain why it might be optimal to have one share/one vote. The main argument is that an equal distribution of voting rights (rather than multiple classes of stock) will place all competitors for corporate control in the same position. An unequal distribution, by contrast, can favor those for whom the private benefits from control are high while the social benefits are not. Thus, one share/one vote may provide for the right transfer of control.

Dilution rights are important in inducing take-overs by forcing the shareholders to tender. Yet they need not cost present shareholders much, if there is more than one bidder. Present takeover regulations (in particular, Williams Amendment from 1968) encourage multiple bids by requiring a minimum number of days before target shares can be purchased and by permitting tendered shares to be withdrawn if a higher bid arrives. The role of multiple bids is discussed in Bradley et al. (1987). They provide empirical evidence showing that take-over bids are frequently front-end loaded and hence coercive, but nevertheless give present shareholders the bulk of the synergistic gains (about 95 percent of the joint increase in firm values). Their study suggests that multiple bids are quite common as well as important for a proper understanding of take-overs.

Dilution is not the only way to provide incentives for a take-over. An alternative explanation for why tender offers can succeed is that the raider holds a substantial share of the firm at the time of the offer [Shleifer and Vishny (1986a)]. In that case he can offer a price which is high enough to compensate present owners for the expected increase in firm value and still be left with a surplus. In other words, the minority shareholders free-ride, but the raider's gains are big enough that free-riding does not hinder a take-over. (Present regulations require that stockholdings (and intent) be disclosed once more than 5 percent of the firm is acquired. However, due to a permitted lag in filing, the initial stake of raiders is on average above 10 percent by the time the tender offer becomes public.) Thus, the prospect of take-overs, hence market monitoring of management, could be substantially improved by the presence of large shareholders. In contrast, one might hypothesize that if management holds a large enough proportion of shares, then take-overs are unlikely to succeed. Of course, management interest in the firm should reduce incentives to slack in accordance with standard moral hazard reasoning [see our discussion of Jensen and Meckling (1976) in Subsection 3.1]. There is a potential trade-off. A small management share will act as a good incentive. A larger share will prevent the market for corporate control from operating effectively. Morck, Shleifer and Vishny (1986) provide preliminary evidence that the best incentives are supported by an intermediate managerial stake (in the range of 5–20 percent).

The models mentioned so far say little about the actual mechanism by which take-overs police management. Scharfstein (1988b) has elaborated considerably on this dimension.³² He analyzes a model in which the manager would want to put less effort into managing if he could. The manager has an opportunity to slack, because he, but not the owners, knows the potential productivity of the firm.

Scharfstein assumes that with an exogenously given probability there is a raider who can observe the change in the technology. If he takes over, he can implement a new and better contract. There is no argument given for why the

³²A related model is in Demski, Sappington and Spiller (1987).

raider might be in the unique position of learning about technology. Ideally, one would like to study the incentives for the raider to invest effort in monitoring the firm.

The problem for present shareholders is to design a optimal contract for the manager, given the knowledge that technology may change and that this may trigger a take-over. Dilution is a parameter in the design. To make matters simple, Scharfstein assumes that dilution is determined by a commitment on behalf of the shareholders to tender their shares if and only if the raider's offer is above a given price. Also to be determined is the severance pay to the manager in the event of a take-over and a loss of the job.

The optimal program is relatively complicated despite the simplifying assumptions. Scharfstein shows, however, that the potential of a raid is helpful in disciplining the manager. Also, in accordance with Grossman and Hart (1980), shareholders will commit to a price that is below the potential value of the firm in order to encourage take-overs. From this he goes on to conclude that efforts to curb a manager's ability to fight take-overs have value, both socially and for incumbent ownership.

Defensive tactics have been a hotly debated legal issue in recent years.³³ Shleifer and Vishny (1986b) argue that it is not always against the interests of present shareholders to prevent defense mechanisms. They focus on greenmail whereby the manager buys out the raider in exchange for a promise not to attempt a take-over for a given period of time. The idea is that excluding a bidder may be a way of inviting even better offers from other bidders later. In their model, by assumption, the manager acts purely in the interest of the shareholders, yet greenmail occurs.

Another rationale for defensive tactics is that the threat of take-overs may induce managers to behave myopically. One could conceive of situations in which take-overs are desirable *ex post*, but undesirable *ex ante*, because of the distortions in decision-making that they induce. Thus, if the threat can be reduced by poison pills and the like, or alternatively, if the likelihood of take-overs can be altered by golden parachutes or other compensation schemes, this may be socially good. Laffont and Tirole (1987b) have studied a model of efficient managerial turnover and optimal defense tactics in case managerial quality is uncertain (management can be either good or bad *ex ante*). The second-best (screening) contract makes transfer of control less likely than under full information and it also provides for time-increasing (incentive) stock plans. Both serve to make the manager value the future more. [Two other papers, investigating managerial myopia and take-overs, are Hermalin (1987) and Stein (1988).]

³³ Defensive tactics include altering the debt-equity ratio [see Harris and Raviv (1985)], invitation of a "White Knight", selling off assets of value to the potential acquirer, acquiring assets that may make the merger illegal on anti-trust grounds as well as litigation of other forms.

The preceding discussion has studied the private incentives to favor or fight take-overs. In conclusion, we note that we have not said much about the social benefits of take-overs. First, the optimal corporate charter from the point of view of initial shareholders may lead to socially too few take-overs (this is akin to monopoly pricing; a similar argument is made in a related context by Aghion and Bolton (1987)). Second, it may be the case that there are too many take-overs, because the latter may simply redistribute rents (away from labor unions, say), as has recently been suggested by Shleifer and Summers (1988).

4.5. Implications for firm behavior

It should be clear by now that many forces operate to discipline management. But what are the implications of these disciplinary measures on firm behavior? In what ways will the neoclassical treatment of the firm be altered?

We discussed in Section 3 one important decision which may be influenced by agency considerations: the choice of financing mode. Here we will focus on investment and production decisions.

Regarding investment choice we note first that the objective function of the firm is typically ill-defined as soon as markets are incomplete, even without separation of ownership and control. Only under exceptional circumstances will shareholders agree on which investment and production decisions the firm should take. Since the question of unanimity has been explored rather exhaustively in the literature by now, with several good summaries available [see, for example, Grossman and Hart (1977)], there is no reason for us to reiterate the findings here.³⁴

Two implications are, however, worth noting. The first is that the question, does the manager maximize profits (or the value of the firm)? is not always a meaningful one, particularly in connection with investment choice. The second implication is that there is little reason to expect that owners will agree on the manager's incentive structure when markets are incomplete. Thus, the abstraction employed throughout, that the owners can be represented by a single principal who designs a scheme for the manager, can well be questioned. Perhaps a better approach, as well as a more realistic one, would be to see the corporate control problem in a political perspective. Grossman and Hart (1977) mention this. They

³⁴We know of little empirical evidence concerning the importance of incomplete markets. Some would argue that with the multitude of instruments presently available, securities markets are effectively complete. Thus, shareholders should not disagree for reasons of market incompleteness about a firm's investment plans. This is not in conflict with the apparent fact that markets for human capital are seriously incomplete. Claims on human capital cannot be sold (slavery is forbidden) and services for human capital face trading impediments for moral hazard and adverse selection reasons as we have discussed.

envision that all corporate decisions are determined by majority rule. More generally we can envision a constituent theory in which managers act much like politicians. Managerial decisions are guided by a concern for constituent support, in particular from owners, but they are not directly controlled by ownership. The voting power of the owners is primarily vested in the right to oust management.

Let us go back to our original question. The implications of incentive problems on investment choice are ambiguous if one looks at general managerial models. This is hardly surprising and partly a modelling problem. For more specific models, sharper predictions can be made. For instance, consider a moral hazard model in which the manager's incentive scheme is linear such as that of Holmstrom and Milgrom (1987). In this model the slope of the manager's incentive scheme is negatively related to the size of risk (the variance of the technology), while the overall agency costs are positively related to risk. Consequently, scale decisions, which increase riskiness in the sense that they make it harder to identify the manager's contribution, entail increased agency costs. Scale will be smaller than in a world with symmetric information.

A more interesting scale effect is present in models where managers can use capital (or any other input, like labor) as a substitute for their own effort and the owner cannot determine whether the manager is asking for capital because prospective returns are high or because the manager intends to slack [see, for example, Hart (1983b) or Holmstrom and Weiss (1985)]. The nature of an incentive compatible scheme in this case is such that the manager is not given as much capital as he would get if the owner could verify the information the manager has. Agency costs manifest themselves in underemployment of capital, because one wants to discourage managers with a high return potential from pretending that it is low.

These findings may have macroeconomic consequences. For instance, one can construct models in which economy-wide resources are underemployed [Grossman, Hart and Maskin (1983)] as well as models in which swings in aggregate economic variables get amplified due to managerial slack [Holmstrom and Weiss (1985)].

Portfolio choice is also influenced by agency considerations. The normative implications of standard portfolio analysis, for instance the capital asset pricing model, are simple. A publicly held firm should only consider contributions to systematic risk in deciding on an optimal portfolio. Idiosyncratic risk should not matter, because investors can diversify in the market. Moreover, all firms should judge projects the same way (assuming of course that the return characteristics are independent of the firm undertaking the investment). In contrast, once agency costs are incorporated, idiosyncratic risk may come to play an important role. The reason is that the manager must bear some idiosyncratic risk, because that is precisely the risk that is informative about the effort he put in. In fact, with relative performance evaluation one may in some cases completely want to

eliminate the systematic risk component from the manager's reward structure.³⁵ This will imply that in making investment decisions, the firm should consider idiosyncratic risk an important factor, which indeed seems to be the case in reality.

Because idiosyncratic risk will have to be borne, diversification by the firm may become desirable. The argument is a bit more delicate than one might think, though, because one has to explain why, instead of diversifying, the firm could not use relative performance evaluation as a substitute. Aron (1988) has studied the problem in more detail.

Quite generally, portfolio choice in a managerial model is influenced by a desire to specialize for technological reasons versus a desire to use a common technology for incentive reasons. As a manifestation of this tension, consider the choice of how much correlation to seek with the market. Agency theory puts a premium on being technologically closer to other firms for the purpose of being able to control the manager better. Thus, firms should bunch together more than they would in a world under symmetric information. Consequently, the social portfolio will be riskier. More interestingly, if the managers feel that they are directly and indirectly compared to the competition, that may lead to bandwagoning effects (the mistakes in loaning extensively to the LDC countries could be one example). Managers, in fear of being too exposed, will choose their activities close to each other. This has not been extensively analyzed, though it is easy to see how it may come about, for instance, in reputational models. We suspect that the racing aspects present in career pursuits may in the end be the ones that have the most profound effects on managerial behavior.

Next, consider implications for output decisions. Will the manager set output so as to maximize firm profits? The answer depends very much on the technology. Suppose managerial effort only affects marginal cost and, somewhat unrealistically, suppose marginal cost can be observed and contracted on. Then it is obvious that the manager will be quite happy to choose the quantity that maximizes profit, *conditional on cost*. In a sense, cost acts as a sufficient statistic and nothing additional is learned from quantity choice. Hence, incentives should not be connected with quantity choice and the profit maximization paradigm remains valid in spite of agency problems. In contrast, of course, if quantity decisions, directly or indirectly, provide additional information for contracting, then distortions will occur in its choice.

Managerial incentives also affect product market competition. Indeed, if the contract between the manager and owner is observed by competitors, the contract should be designed to influence competitor behavior. For instance, in an

³⁵This is true if the technology takes the form $x = a + \theta + \epsilon$, where a is the manager's effort, and the manager cannot privately decide on investments. But recall our earlier observation that if the manager can control investments, then relative evaluations may be less desirable.

oligopoly game, a manager can be given incentives not to lower price or not to increase sales. The agency relationship allows owners to commit to a price in a way that may be infeasible otherwise.³⁶ The effect is that competitors will also raise their price. However, it can be shown [Fershtman et al. (1986) and Katz (1987)] that if the manager's *actions* can be contracted on and managerial contracts are mutually observed, a "Folk Theorem"-like result will obtain: the oligopoly equilibrium is indeterminate. By contrast, if actions cannot be contracted for directly, but rather must be induced via a *performance* plan, agency will matter and need not lead to indeterminacy [Fershtman and Judd (1986)].

The strategic aspects of agency have also been studied when performance-based contracts are impossible, but some contractual choice is still feasible [see Bonanno and Vickers (1986), Mathewson and Winter (1985) and Rey and Stiglitz (1986), all on vertical restraints; see also Brander and Lewis (1986) on debt contracts]. If managerial contracts are not mutually observed, say because they are implicit or entail side-contracting, then agency does not matter if the manager is risk neutral (the owners offer to sell their firm to the manager), but may matter if the manager is risk averse; see Katz (1987). (For instance, risk-sharing provisions may make the manager a tougher bargainer.)

The effect of agency on competition (as well as the feedback of competition on agency contracts) is an interesting topic. Much seems to depend on variations in assumptions such as: Are contracts observed by competitors? Can contracts depend on the agent's action or just his performance? Can contracts be linked to those of competitors? A lot more work remains to be done on the subject.

4.6. The hazards of a black-box view of the firm

In Subsection 4.5 we presented an example of a firm whose manager slacks, and yet the firm is observationally indistinguishable from a profit-maximizing firm. This leads us to consider more generally whether the firm can be represented by a single objective function such as profit maximization, or, as was presumed by the older literature on managerial theories of the firm, a utility function increasing with the firm's profit, size, growth or expenses [see, for instance, Baumol (1959), Penrose (1959) and Williamson (1964)].

Many models have assumed that the firm's managers maximize the firm's size subject to "capital market constraints", which guarantee the owners a minimum profit. Such reduced forms beg the question of why managers care about size. The literature commonly offers psychological reasons. No doubt, managers enjoy the power associated with a large number of subordinates. But it also seems

³⁶The use of agents for purposes of commitment, and hence an improved strategic position, is widespread.

important to investigate whether size concerns could be explained economically. For example, we could identify economic reasons which emanate from the traditional principal–agent model in which the manager has private information about productivity, and in which his effort to obtain a given output target decreases with the productivity parameter and the number of subordinates (a larger number of subordinates may reduce on-the-job pressure, etc.). In such a model, the manager cares about size (the number of his subordinates) *not per se*, but because a large workforce allows him to enjoy an easy life (exert low effort). The size of the firm then exceeds the optimal size (obtained when the owners have perfect information about productivity). Other – more conjectural – economic explanations for the size concern come to mind. (1) The size of one's staff may influence the labor market's perception of one's ability. (2) A large staff may make a manager's function harder to suppress (because relocation or layoff of the staff are more costly). (3) In a framework in which managerial compensation is based on the performance of competitors, it may pay to expand beyond the profit-maximizing point if competitors are on the same product market and are thus hurt by one's expansion. (4) In a dynamic setting, a manager may want the firm to grow to secure promotion opportunities for his subordinates.

These more “primitive” explanations could be read as supporting the general assumption that managers care about size, and interpreted as vindicating the reduced form approach. This is missing the point. Reduced forms are not robust to structural changes. Only a careful consideration of the structural form will indicate whether profit maximization or size maximization are good approximations for positive or normative analysis. It is clear, for instance, that the five “explanations” above of why managers care about size have diverse implications concerning the firm's behavior in the product market. We feel strongly that there is a need to study where black-box representations of the firm's objective yield appropriate approximations of its behavior.

4.7. Concluding remark

Agency considerations affect the behavior of the firm quite generally. As the attentive reader must have noticed, however, there is a significant dilemma in that most of the changes in firm behavior are hard to observe. The modeller sees everything that goes on inside his model, but to an outside observer much of it will go unnoticed. For instance, the fact that the manager chooses more or less risky investments than would be the case without an incentive problem is hardly an observable implication. The same can be said about scale and output decisions. We believe this problem (which is not entirely unique to agency theory) has received way too little attention in the literature to date. One reason is that

agency models are rarely incorporated into an economically richer environment, because such extensions tend to be complicated.

5. Internal hierarchies

The previous section identified the firm with a manager (or a group of perfectly colluding managers), whom shareholders and creditors tried to control through a variety of mechanisms. The rest of the firm was viewed as a black box described by a stochastic technology that transforms the manager's actions and characteristics into an outcome (e.g. profit). We now open this black box and take a look at internal organization. Before addressing the relevant issues, we make two methodological points.

The two-tier capitalist-management model creates the potentially misleading impression that the internal hierarchy (below management) can be summarized in reduced form by an exogenous random production function. To see why this need not be the case, consider a three-tier structure: owner/manager/worker. The manager faces an incentive scheme $s_1(\cdot)$ that is contingent on observable variables and picks some unobservable action a_1 (related to production, supervision, etc.); similarly, the worker faces an incentive scheme $s_2(\cdot)$ and picks an action a_2 . In general, the optimal action $a_i (= 1, 2)$ does not only depend on $s_i(\cdot)$, but also on $a_j (j = 1, 2; j \neq i)$. For instance, the manager and the worker may form a productive team in which the manager picks the technology (a_1), the worker produces using this technology (a_2) and hence output x depends on both sides' actions. Assuming that both sharing rules are tied to output, each party's decision to act depends on what the other party intends to do.³⁷ The outcome will be a solution to a non-cooperative game. Consequently, even if the incentive contracts for the rest of the hierarchy are taken as given, one cannot write a reduced form technology for the firm, $F(x | a_1)$, that maps the manager's effort into a probability distribution over output, and then proceed as in a standard one-agent model (cf. Subsection 4.1). The mistake is in assuming that the manager's action can be described by a maximization over his expected utility given his sharing rule, when in fact it will be determined by the outcome of a non-cooperative game, where the expectation of the worker's action plays a central role. Another way of expressing this is to say that the technology controlled by the manager cannot be defined independently of his incentive scheme as in one-agent models.

³⁷To give another example, suppose that x depends only on a_2 , and that a_1 represents supervisory effort, which provides an estimate of a_2 . Assuming that the sharing rule depends on the supervisory evidence found by the manager (as well as on x , if the latter is verifiable), the hierarchical structure yields a supervision game, in which, again, each party's optimal structure depends on the other party's action.

Our second methodological point concerns the distinction between single comprehensive contracting and multi-lateral contracting. The former assumes that the owners impose a grand contract upon the entire hierarchy, preventing side (or delegated or sequential) contracting between its members. In the context of the previous example, a grand contract, chosen by the owner, would determine $s_2(\cdot)$ and $s_1(\cdot)$ simultaneously, assuming recontracting between the manager and the worker is prohibited. A grand contracting approach is employed in the unified, abstract models of Laffont and Maskin (1982) and Myerson (1984). For instance, Myerson views an organization as a centralized communication system in which at each date a mediator receives information from the various agents and, in turn, tells each agent the minimal amount necessary to guide the agents' actions at that date. Most other work on multi-agent models assumes likewise that the principal can design contracts for all agents without the agents being able to recontract or communicate among themselves.

Clearly, if preventing side-contracts is costless, there is no loss in employing a grand contract design. In fact, one would typically gain strictly by controlling all contract options. Unfortunately, preventing side-contracts is costly or infeasible, which suggests considering the polar case of unlimited side-contracting. Side-contracting will usually add costly constraints to the owners' optimization problem.³⁸ Obviously, both the grand contract design and the unlimited side-contracting model are caricatures, and the reality must lie somewhere in between; we would actually argue that the amount of side-contracting that takes place in organizations is an important aspect of the overall design problem.

Below, we will discuss hierarchies in terms of the services that they provide. Information systems are taken up in Subsection 5.1 and supervision in 5.2. Associated incentive features are discussed in Subsection 5.3. Subsection 5.4 considers the implications of learning on promotion and task assignment as well as on wage structures. Subsection 5.5 looks at how hierarchies limit the costs of side-contracting. We finish with two subsections – one on authority, the other on organizational forms – which make little reference to models for the simple reason that there is almost no formal work on these subjects. Our remarks here are correspondingly more philosophical and intended to bring attention to a big gap in formal theorizing about the firm.

5.1. Hierarchies as information systems

As we have noted before, information is valuable for at least two reasons. It improves decision-making and it permits better control of a subordinate's ac-

³⁸These additional constraints bring technical complications. In a grand design, the optimal contract that implements action a_i by agent i only depends on what the other agents are asked to do, not on what contracts they are on. In contrast, such a partial decomposition is not possible when side-contracting must be considered.

tions. Thus, there is both a decision-making demand and an incentive demand for information, both of which have been well recognized by accountants studying the properties of internal information systems. To isolate the decision-making demand, one can study organizations under the simplifying assumption that all members share the same objective. This is the approach taken in team theory [Marschak and Radner (1972)].

A team theoretic study begins by postulating an underlying organizational decision problem. For instance, the problem could concern how much to produce and distribute to separate markets, each with an uncertain demand. Information about demand is collected by different members. How much information should they collect and how should they communicate with each other? This problem is approached in two stages. In the first stage the value of a given information system is established by solving for the optimal organizational decision rule (which consists of the set of decision rules for its team members) under that particular information structure. One of the central results of team theory is that the optimal decision rule coincides with a person-by-person satisfactory rule under standard concavity assumptions. This means that the overall optimal decision rule is one in which each team members' rule is optimal for that member alone taking the other members' rules as given. (In game theoretic language, person-by-person satisfactory is equivalent to a Bayesian Nash equilibrium, which obviously is a necessary condition for optimality.)

Equipped with a characterization of optimal decision rules, the second stage compares alternative information systems. These correspond to stylized communication structures that might be observed in the real world. The problem is usually too complicated to derive the best information structure given costs of information acquisition and communication. Therefore, most of the theory is focused on discrete comparisons of the benefits, leaving the costs to be evaluated separately. An exception is a recent paper by Geanakopolos and Milgrom (1985) which formalizes information and communication costs in a tractable way and derives closed form solutions for the optimal hierarchical design. The cost of communication is interpreted as stemming from delays in decision-making. One interesting implication is that the optimal hierarchy is finite in size, because the benefits from adding coordinating layers of management eventually go to zero.

One general result on the comparison of information systems emerges from team theory: an information system x is more valuable than information system y if it is more informative in the sense of Blackwell (1953). Loosely speaking, x is more informative than y if it is less garbled, that is the distribution of signal y can be construed as arising from the signal x plus additional noise. This is closely related to the notion that x is a sufficient statistic for y . What this result shows is that Blackwell's analysis of one-person decision problems extends to multi-person settings, assuming that objectives are shared.

In fact, recalling our discussion about information in agency problems in Subsection 4.1, Blackwell's result also extends to situations in which objectives

are not shared. A more informative information system is not only more valuable for decision-making, but also for writing incentive contracts. A quite general treatment of these questions, covering both decision-making and incentive demand for information, is provided in Gjesdal (1982).

These results all have bearing on problems in accounting, particularly managerial accounting. Accountants have taken a keen interest in the information economic literature and developed it for their own needs [see, for instance, Demski and Feltham (1976) and Baiman (1982)]. The informativeness criterion gives us some feel for what type of information is worth recording in an accounting system. Elaborations are provided in several papers. Antle and Demski (1987) view accounting rules as providing a numerical representation of some coarsening of the information and analyze when these rules (in the context of revenue that must be recognized over time) involves a loss of (useful) information. Demski and Sappington (1986), analyze line-item reporting and ask when there are strict gains to auditing one more variable. Holmstrom and Milgrom (1987) rationalize the use of time-aggregated accounts. Caillaud, Guesnerie and Rey (1986) and Melumad and Reichelstein (1986) study gains to communication in an organization in which coordination is not required; communication serves the purpose of selecting a desirable incentive contract from a menu of choices (participatory management). Maskin and Riley (1985) compare the values of alternative information structures, assuming that high measurement costs make them mutually exclusive. For instance, is it better to monitor a worker's output or her input (use of capital, raw materials or possibly labor)? Crampes (1983) offers a similar analysis in a regulatory context.

The central question in accounting is how to aggregate information. Accounting systems aggregate information to a substantial degree. The explanation offered by Blackwell-type results is that nothing is lost from aggregating information into a sufficient statistic. However, it is clear that accounting systems go well beyond such limited aggregation. The obvious explanation is that information is costly to process and communicate. Information of marginal value is not worth including in contracts nor worth communicating further within the organization. Unfortunately, neither team theory nor agency theory have been able to incorporate information costs very effectively. Partly this is due to severe problems in quantifying information. For instance, trying to measure information in terms of "bits" transmitted or processed has met with limited success economically.

It appears that further progress on information costs will require a better understanding of the nature of information and its role in decision-making. In particular, one must come to grips with the difficult concept of bounded rationality. Bounded rationality manifests itself on two levels in organizational decision-making. On an individual level, a decision-maker must (i) isolate the relevant part of the available information and (ii) find the optimal decision. The first step is not a trivial one; we all know that having too much information is as bad as having no information at all if, as is often the case, we do not have the time to

sort out the decision-relevant part. Indeed one of the functions of accounting systems is to aggregate information so that decision-makers can focus on a small number of key variables. The second step is also time-consuming (the typical example is a chess decision, which to be even nearly optimal would require extremely long backward induction computations). In both steps, the decision-maker must trade-off the quality of information and decision-making against the costs (time or other) of improving them. This tradeoff, particularly emphasized by Simon (1957, 1976), is of crucial importance in practice, but little formal progress has been made on examining it or its implications for organizational behavior.

Bounded rationality is important not only on the individual level, but also on the organizational level. An organization cannot afford to remember extensive and detailed information. Instead it attempts to codify information in the form of standardized rules that are meant to help the organization to adapt quickly and relatively efficiently to changes in the environment. Several authors [for instance, Arrow (1974), Nelson and Winter (1982), Kreps (1984), Schein (1984) and Cremer (1986b)] have tried to bring content to the notion of organizational memory. The interesting question of what happens when the environment changes more dramatically, outdated present agendas, operating rules and organizational memory, is largely unexplored, though it must be of considerable importance. This relates closely to the problem of modelling unforeseen contingencies under incomplete contracts, mentioned earlier. Both issues would benefit greatly from a successful formalization of bounded rationality and complexity.

5.2. Supervision in hierarchies

The complexity of the two-tier agency structure may well account for the fact that studies of higher-order hierarchies have been rare. Very special assumptions must be made in order to be able to solve the optimal contract associated with a complex hierarchy. Interesting insights have nevertheless been obtained by Williamson (1967), Calvo and Wellisz (1978, 1979), Rosen (1982) and Keren and Levhari (1983) among others. As an illustration, we will give a simplified exposition of the Calvo–Wellisz model (1978).

Suppose that the firm is organized according to a familiar pyramidal structure. Level 1 forms the productive tier of the firm (workers). Level 2 consists of managers supervising level 1 workers. Because the quality of supervision is a function of the number of workers being supervised, there may be many level 2 supervisors. These in turn need to be monitored by a third level and so on. The top level consists of a single agent (or unit), who is the residual claimant of the firm's profit, net of wage and input payments. For instance, level 3 could be the shareholders (respectively, the executive officers) and level 2 the executive officers

(respectively, the division officers). Note that the numbering of levels is from bottom to top. One of the main questions is how many supervision layers to build above the productive workers on level 1. What determines the size of the firm? What constrains it from growing indefinitely? Potentially, the answer could be a deterioration of supervisory effort as envisioned in Williamson (1967).

Calvo and Wellisz consider the following simple technology. Employees can either work (0) or shirk (1). The monetary disutility of work is g . Supervision involves “checking” a subordinate (employees on the immediately lower level) with some probability. Checking reveals the subordinate’s activity without error.³⁹ If the employee could be punished sufficiently for not working, then the agency problem could be solved costlessly (the threat to punish would act as a sufficient threat at a minimal level of supervision, i.e. probability of checking). To avoid this, one assumes limited liability (or infinite risk aversion below some threshold income if there is any possibility of a monitoring mistake) so that the punishments are restricted. Then the optimal punishment is to bring the employee to her reservation utility (normalized to zero).

If the employee is not checked, or if she works when being checked, she gets a wage w . If p is the probability of being checked, the employee works if and only if $wp \geq g$. Thus, the “efficiency wage” equals g/p . Note that it decreases with the probability of monitoring. Also, the employee earns rents because of limited punishments; the rent is $w - g = g(1 - p)/p$. The supervision technology is described as follows: the probability of being checked is a decreasing function of the total number s of employees supervised by her supervisor; for instance, $p = 1/s$. The employee at the top of the pyramid is exogenously supposed to choose $e = 1$ (presumably because of the labor or capital market incentives discussed in Section 4). Note that this is an Alchian–Demsetz-type model in which monetary incentives based on individual performance are infeasible and control must rest on supervision of input supply.

In the optimal contract, all employees work (shirking at any level implies shirking at all lower levels of the hierarchy). Suppose that a level- k employee together with the employees under her, brings profit Π_k (where $\Pi_1 = x$ is the output per worker). Π_k is defined gross of the wage required to induce the employee to work. A level- $k + 1$ employee should supervise n_k level- k employees, where n_k maximizes $n_k(\Pi_k - gn_k)$; so $n_k = \Pi_k/2g$. And $\Pi_{k+1} = \Pi_k^2/4g$. Given this, the top manager is willing to add a $(k + 1)$ th layer of employees (pushing herself to the $(k + 2)$ th layer) only if $\Pi_{k+2} > \Pi_{k+1}$ or $\Pi_{k+1}^2/4g > \Pi_k^2/4g$, that is, $\Pi_{k+1} > \Pi_k$.

In this model, we thus obtain an optimal firm size equal to either one (self-employment) or infinity. This is not very satisfactory, but the conclusion is very sensitive to the supervision technology specified; see Calvo and Wellisz

³⁹For an interesting analysis of imperfect monitoring, see Baiman and Demski (1980).

(1978). More interesting is the observation that the span of control increases with the rank in the hierarchy; because $\Pi_{k+1} > \Pi_k$, $n_{k+1} > n_k$. This implies that the wage also increases with the rank in the hierarchy, even though all employees are identical, and all jobs equally hard to perform.

An important question for owners of a private firm or supervisors of public enterprises is how incentives extend down the hierarchy. Top managers form only a small part of the organization; indeed, much of the productive work is done by layers that have limited financial incentives and whose rewards are not determined directly by the owners (engineers, marketing staff, product analysts, and especially production workers). Top managers are crucial because a failure in their supervisory, coordination and arbitration functions has severe consequences for how the rest of the organization behaves. This is well illustrated in the Calvo–Wellisz model, where lack of supervision by one manager implies shirking by all employees below her. The way in which lower units, and therefore the performance of the firm, respond to changes in the upper units' incentive schemes is an important question that hierarchical models of the kind just described could shed useful light on.

Let us turn next to some other questions concerning supervision. As mentioned in Section 2, technological non-separabilities and the concomitant problem of identifying individual performance, create a problem of moral hazard in teams.⁴⁰ Monetary incentives based on joint performance, which involve a source that breaks the budget balancing constraint, may work in some circumstances [Holmstrom (1982a)]. This solution may be limited by coalition formation and risk aversion.⁴¹

An alternative is to obtain further measures of individual performance by establishing the input supply (effort) of each agent. Supervision serves that role. We would expect supervision to be more prevalent in parts of the firm where individual contributions to output would otherwise be hard to measure.

The type of evidence that the supervisor collects is of central importance. One must distinguish between hard evidence, which is data that can be verified in case of a dispute and soft data, which cannot be verified. Examples of hard data include accounting information and information about the number of units that an employee has produced. Hard data can be used in explicit incentive schemes such as piece rates.

More often the supervisor can only obtain soft data by judging the employee's performance by direct observation. The supervisor must then be trusted to report findings in an honest fashion. This can pose special problems. Honest reporting

⁴⁰"Team" here has a different meaning from the one given in Subsection 5.1. Parties in a team do have conflicts of interest in this subsection.

⁴¹The agents may collude against the source by coordinating to produce beyond the non-cooperative level. See Eswaran and Kotwal (1984).

may not be in the supervisor's interest for several reasons. First, to induce her to exert supervisory effort, she may be paid according to the number of mistakes or failures she records; she may thus have an incentive to overstate the frequency of shirking. Second, a supervisor is often a member of the team herself through her non-supervisory activities (coordination, management, communication). Hence, she may be tempted to assert facts that reflect poorly on the other members in order to emphasize her own contribution to the team's performance.

There are mechanisms that can "harden" soft information (make it more reliable). Suppose the supervisor monitors many agents (or a single agent over time). A "quota" system entitles the supervisor to distribute a given number (or maximum number) of sanctions or rewards among the agents. She still has the authority to announce which agents shirked and which did well, but now she cannot influence sanctions or rewards as freely. Examples of quota systems include a coach who picks the players for a game, or a school teacher who decides who should enter the next grade. The point of a quota is that it circumscribes the supervisor's ability as well as desire to misrepresent facts.

This important observation originated in the tournament literature [Bhattacharya (1983), Carmichael (1983) and Malcolmson (1984)]. Tournaments, in which a set of prizes are distributed to team members based on rank-order performance, can be viewed as a variation on the quota system. The essential characteristic is that the sum of prizes is constant, which has desirable incentive properties for the principal. She cannot escape payment by distorting observations. Furthermore, when there is a large number of agents in the team, optimal tournaments may approximate closely optimal general incentive schemes [Green and Stokey (1983)]. (Note, however, that the large numbers case has the drawback of yielding a large span of control and therefore a poor quality of supervision.) We will return to tournaments in the next subsection.

The use of quotas has potential drawbacks. It may have perverse effects on the supervisor's incentive to exert supervisory effort. Why should she care about whether Mr. A did better than Mr. B? This problem is partly curbed by the supervisor's reputation. To take an analogy, consider the case of a policeman handing out tickets for speeding. The policeman's word is trusted by authorities (police department or courts) over the driver's. Presumably, this is only because the policeman has a more frequent relationship with the authorities than the driver, and therefore is more able to develop trust with those authorities. And, indeed, if too many drivers complained of unfair ticketing by the same policeman, the authorities would become suspicious and would launch an inquiry. Similar considerations may be important in firms.⁴²

⁴²Note that this mechanism is similar to allowing a maximum number of complaints over some length of time.

5.3. Hierarchies as incentive structures

Hierarchies can act as incentive structures by inducing competition among agents. We will discuss two channels through which members of a hierarchy may be led to compete with each other. First, an agent's performance may be usefully compared to the other agent's performance for monitoring purposes when agents face correlated shocks (this will be referred to as yardstick competition). Second, the agents may be induced to compete in the same market.

The tournament literature has discussed an interesting incidence of yardstick competition [Lazear and Rosen (1981), Green and Stokey (1983), Nalebuff and Stiglitz (1983), Mookerjee (1984) and Shleifer (1985)]. Tournaments compare agents by rank, which provides both insurance and flexibility. Agents will have to carry less risk, because their performance rank is insensitive to common uncertainties (cf. our earlier discussion of relative performance evaluation; Subsection 4.1). They will also be induced to adapt more efficiently to common changes in the environment. In special cases, tournaments duplicate optimal insurance as well as work effort despite non-variable changes in circumstances. The fact that rewards are paid based on ordinal rather than cardinal measures can be a further advantage when measurement costs are high or when measures are hard to quantify (for instance, because they are based on supervisory judgement; see the discussion above). Indeed, tournaments are commonplace in firms. A prize for "the most valuable employee of the month" is a quite explicit example. More importantly, promotions induce a tournament or a sequence of tournaments.

The optimal design of prizes as well as the composition of agents in a tournament have been analyzed in this literature. One question of interest is the following: Can the strongly skewed distribution of earnings across ranks commonly observed in firms [Lydall (1968)] be explained as an optimal tournament design? Lazear and Rosen (1981) find that single tournaments do not yield sufficient skewness. Subsequently, Rosen (1986) has reconsidered the question in a model of elimination tournaments. He notes that most top managers come through the ranks and that the process may be similar to sports contests in which players are eliminated in each round. In sports (e.g. tennis) the distribution of prize money is very skewed as well. Rosen finds that sequential tournaments can explain skewness better. The intuition is that managers who come close to the top of the hierarchy see their advancement opportunities shrink (assuming that the size of the hierarchy is fixed). To preserve their work incentives higher rewards must be provided.

The tournament literature provides other interesting insights into the design of hierarchies. We think one should be cautious, however, in interpreting a hierarchy as designed uniquely for the purpose of providing agents with an incentive structure. Why should job assignment be part of the reward system? Why not let the supervisor distribute prespecified monetary rewards instead? It seems more

likely that promotions primarily serve the purpose of moving people to tasks where their comparative advantage is highest and that incentive properties are derivative. On the margin promotion rules could be influenced by incentive considerations, but incentives could hardly be the driving force. [In fact, promotions with associated large wage increases could have rather detrimental effects on the continued incentives for losers; see Dye (1984).]⁴³ In the next subsection we will discuss learning models of job assignment and provide independent reasons why wages might be attached to jobs. It would be desirable to mix the adverse selection/job assignment literature with the moral hazard/tournament literature to obtain a more consistent theory of hierarchical mobility as an incentive device. [For a start, see MacLeod and Malcolmson (1985).]

We turn next to product market competition. When an agent's incentives cannot be based on a reliable measure of performance, product market competition can supplement imperfect compensation schemes and act as an implicit incentive device. We describe two examples, in which the principal may want to induce product market competition between the agents (this contrasts with the literature reviewed in Subsection 4.3, in which the market structure is exogenous).

Rey and Tirole (1986) offer a model of retailers serving a given geographical area or more generally a market (within a firm, one might think of competing salespersons, divisions, or marketing teams). The retailers (agents) sell the goods produced by a monopolist supplier (principal). They may either compete throughout a geographical area or market, or alternatively be allocated a territory or market segment over which they have a local selling monopoly (exclusive territories). Their performance is not directly observable. Competition on the product market (in the price or services) has the advantage of partly insuring the agents. A shock on demand or retail cost is likely also to affect one's competitors and therefore gets partly absorbed through the competitive mechanism (cf. our earlier discussion in Subsection 4.3). By contrast, under exclusive territories, no such compensatory mechanism exists, and the agents are therefore more exposed to demand and cost fluctuations.

Competition has desirable insurance properties similar to general relative performance schemes. And, analogously, in the course of providing insurance it acts as an incentive device. There is, however, a flip-side to competition. It constrains the way agents can exploit monopoly power in the product market. As is usual in industrial organization, strategic behavior to appropriate monopoly rents destroys some in the process. By contrast, an agent who is granted a monopoly through exclusive territories, is free of strategic constraints and can

⁴³Another problem with tournaments is that they are detrimental to cooperation [Lazkar (1986)]. This relates closely to the problem with relative performance evaluation raised in Section 4; managers may be led to make wrong production and investment decisions if relative values are distorted.

exploit his monopoly power fully. Hence, product market competition may or may not be optimal.⁴⁴

Farrell and Gallini (1988) and Shepard (1987) introduce another reason why product market competition may be desirable. In their models, competition acts as a commitment to supply non-contractible quality. Recall the buyer/seller paradigm of Subsection 2.1, in which the buyer must make some specific investment before the seller delivers. Suppose that the value of this investment depends on the ex post quality chosen by the seller, and that this quality is observable but not verifiable. With only one seller (source), he has an incentive to choose ex post the minimum quality he can get away with legally. This quality is in general much too low, and alternative incentives must be provided to yield an efficient level of quality. The mechanism envisioned by Farrell and Gallini, and Shepard, is licensing by the upstream firm to create a competitor. Upstream competition for the downstream market yields an ex post incentive to supply acceptable quality. This explanation seems to fit some licensing practices (like Intel in the semi-conductor industry).⁴⁵

Tournaments and product market competition are just two ways in which incentives are structured so that information about relative performance gets exploited. The same principles can be applied to explore job structures more generally. One implication is that it may be desirable to design jobs so that they overlap even though such overlap is technologically wasteful. Consultants on organizations have occasionally emphasized the use of job duplication without seeing the economic merits in it [Peters and Waterman (1982)]. A theoretical analysis of a similar phenomenon has been provided in the context of second sourcing (use of two suppliers) by Demski et al. (1987).

Closely related to job duplication is the question of job rotation. Having the same employee perform one task for a long time may be technologically desirable, but the longer he does the job, the harder it may become to know whether he is performing to the potential of that job. Also, setting standards for such an employee may become problematic because of the well-known ratchet effect. In fear of raised standards in the future, the employee may underperform deliberately today [see, for instance, Weitzman (1980), Freixas et al. (1985) and Laffont and Tirole (1988)]. Job rotation provides some relief. The knowledge that the job is temporary induces the employee to perform harder – the cost of higher

⁴⁴Similar effects arise for other competition-reducing restraints, for instance resale price maintenance, when these are feasible. Caillaud (1986) formalizes the effect of unregulated product market competition on the control of a regulated firm.

⁴⁵Unlike the previous model, competition does not have any direct costs in the Farrell–Gallini–Shepard theory. In order to have a single upstream production unit, Shepard introduces increasing returns to scale. One might also be able to construct models in which quality competition would have some drawbacks; for instance, if quality is measured by several attributes, competition may well yield a mix of attributes that is not optimal from the point of view of exploiting monopoly power.

standards will not be borne by him. Also, job rotation offers an alternative source of information about potential. Against these benefits one has to weigh, of course, the costs of training and learning about the task, as well as the intertemporal free-rider problems that may emerge [Fudenberg et al. (1986)].

Finally, we want to mention that incentive concerns influence organizational design in other ways as well. Task assignment can change the opportunity cost of agents. For instance, doing a job at home can be more costly incentive-wise than doing it on the job, because the temptation to slack is greater. The use of time-cards, which permit flexible working hours, which are more sensitive to opportunity costs are another example. The agency literature has paid scant attention to these issues.

5.4. Hierarchies as internal labor markets

Hierarchies are composed of a variety of jobs. An important task of labor management is to assign the right employees to the right jobs. In this subsection we will briefly review the literature on internal labor markets that deals with job assignment and its implications for the wage structure. For the most part, jobs are taken as given here. A related and important question is the design of an efficient job structure, which has received little attention in this literature.

The simplest case of job assignment is one in which job and worker characteristics are known and the environment is static (one period). A basic question is the following: Is it optimal to assign the most able employees to the top of the hierarchy? This need not always be the case. Counter-examples are provided in the communication models studied by Geanakopolos and Milgrom (1985). There is reason to believe, however, that talent is commonly valued more highly at the top, because of the pyramidal structure of the hierarchy. Paraphrasing Rosen (1982), if a soldier makes a mistake he may die, if a colonel makes a mistake his division may be captured, but if the general makes a mistake the whole war can be lost. In other words, the value of correct decision-making multiplies as one goes up the hierarchy and with it the marginal product of ability. This logic carries through in an extension of the supervision model of Subsection 5.2 that incorporates different abilities [see Calvo and Wellisz (1979)].

An interesting implication of matching higher level jobs with higher ability is that it skews the earnings profile. Suppose output in job level i is $x_i = \rho a_i + b_i$, where ρ is a measure of worker ability and a_i is a measure of the importance of the job level. In each job level, more able workers earn more because they are more productive – the relationship is linear in our example. But if it is also the case that higher ability workers are assigned to higher level jobs, the difference in productivity is magnified. The overall relationship between wage and ability

becomes convex. This is emphasized by Rosen (1982) and is also a feature of the Calvo–Wellisz model.

Static models overlook important questions of job mobility. Mobility is of interest only if worker or job characteristics are imperfectly known. This brings us to learning models in which the hierarchy acts as an information acquisition filter. The problem is intricate, because the question is not just to match workers with jobs in a myopic fashion based on currently available information, but also to consider the implications of current assignments on what might be learned for the benefit of future assignments. Obviously, organizations are well aware of this dynamic dimension. Careers are partly designed with learning about ability in mind. A lot of experimentation goes on, particularly with young workers. Correspondingly, older workers may never be given a chance to prove themselves; they may get stuck in jobs that are below their true potential.⁴⁶

Learning models in which workers are merely passive participants – “pawns” moved around by the company – have been studied by Prescott and Visscher (1980), MacDonald (1982) and Waldman (1984) among others. As an illustration, let us give some details of the Waldman model, because it brings out some interesting strategic aspects of job assignment.

Waldman introduces the idea that what is observable to outside firms is not the same as what the present employer learns. Specifically, in his model the worker’s performance (output) can be observed internally, while the outside market only can observe the worker’s job assignment and wage. He sets up a two-period model in which a worker performs a routine job in the first period, which reveals ability. The firm, based on this information, decides whether to promote the worker to a job in which output grows with ability.

Ignoring the market, the optimal promotion policy would be to assign a worker whose ability is above a cut-off level ρ^* to the ability-contingent job and leave him in the routine one otherwise. But one has to consider the fact that the market can bid away the worker contingent on the promotion, which reveals partial information about ability. Assuming that the worker can quit without penalty (because involuntary servitude is prohibited, including the posting of bonds), a promotion implies a wage increase as well to meet the outside bid. Consequently, a worker whose ability turns out to be just above the cut-off level ρ^* is not worth promoting, because of the implied (discrete) wage increase.

⁴⁶The so-called two-armed bandit models [Rothschild (1974)] can explain why some workers may not reach their optimal level. In these models one stops experimenting before the true value is learned with certainty. It is commonly claimed that the reverse phenomenon – known as “Peter’s Principle” – is true. This empirical principle states that everybody eventually rises to a level at which s/he is incompetent. Learning models provide a natural vehicle for studying Peter’s Principle, but we are unaware of any work on this. One of the biggest problems such a model would have to address is why workers, who are found incompetent in their present task, are not demoted. Are the reasons sociological or can an economic rationale be found?

The equilibrium in this model will therefore exhibit fewer promotions than would be optimal if ability information were symmetric.⁴⁷ Actual wage patterns depend on contractual options. Waldman considers two cases: one-period contracts (spot wage) and long-term contracts. The most interesting feature, present in both cases, is that wages are attached to jobs, not to ability. The literature on internal labor markets has made frequent note of this important empirical regularity [Doeringer and Piore (1971)]. The rationale here is that discrete (and sometimes substantial) wage increases accompany promotions, because promotions involve considerable changes in market beliefs. This works in tandem with Rosen's idea, described above, that task assignment magnifies productivity differences due to ability.

In the models discussed so far, the employees do not act strategically. Ricart i Costa (1986) has analyzed a variation of Waldman's model in which the employee also learns his ability, but can use this information to solicit better job offers. The market offers a menu of output-contingent wage contracts in the second period, such that the employee's choice reveals his true ability. The present employer will foresee this and offer a matching wage. The upshot is that wages will be somewhat sensitive to ability in addition to jobs.⁴⁸

Another strategic aspect of importance is that employees, whose promotions will depend on inferences about ability obtained from performance, may change their behavior to influence perceptions. That was already mentioned in Subsections 2.2 and 4.2. Career concerns can be beneficial in inducing effort to excel [Holmstrom (1982b)] as well as detrimental, because they may lead to undesirable influence activities [Milgrom (1988)] or to undesirable investment, production and other decisions [Holmstrom and Ricart i Costa (1986)]. Promotion policies can have a profound impact on employee behavior in this regard. Policies which place relatively larger weight on seniority (the alleged practice in Japan), remove a built-in pressure to compete and may be desirable, because they reduce unwanted influence activity.

From the notion that promotion policies have significant effects on the process of learning as well as employee behavior, it is a short step to realize that job design should be guided by these considerations. Even when employees act non-strategically, job structure matters for learning. Sociologists [e.g. Jacobs (1981)] have argued that the depth of the job hierarchy may reflect the need to become informed about the true characteristics of workers. For instance, in

⁴⁷In a similar spirit, Milgrom and Oster (1984) argue that employers may bias promotion policies in favor of "visible" employees, whose characteristics are better known to the market for some other reason (so that the act of promoting the employee has a lower information content). They suggest that women earn less than men, because they are less visible publicly.

⁴⁸Strategic use of information by the employee has also been investigated in the large labor literature on screening, though most of it makes no reference to hierarchies. [See, for instance, Spence (1973), Guasch and Weiss (1980).] A recent interesting paper on screening is Hermalin (1986).

activities where errors are rare, but disastrous when they occur, a long career path is implied (e.g. airline pilots). These questions could well be addressed more formally with the learning apparatus we already have in place.

A related design question is the degree to which tasks are performed by groups. Sharing of praise or blame changes individual preferences in career models, possibly in an advantageous way. Obviously, there are many other reasons why team work is efficient. But the point is that individual performance measurement, which is prescribed by all effort-based agency models, need not be necessary, nor even desirable when career concerns are considered. This has been overlooked in the past and deserves more attention.

5.5. The hierarchy as a nexus of contracts

The grand mechanism design envisioned in the introduction to this subsection, as well as in Subsections 5.2 through 5.4, is appealing from the point of view of tractability. If it were feasible and costless to design a single contract for the whole organization, it would also be optimal. However, the single-contract paradigm is obviously a fiction. Organization theorists [see, for example, Cyert and March (1963) and Nelson and Winter (1982)] have emphasized the multi-lateral nature of contracting in real world hierarchies suggesting that the firm is a *nexus of contracts* [Jensen and Meckling (1976)].

There are many reasons why contracting is necessarily multi-lateral. For instance, all relevant parties cannot meet at the same time. This is clearly the case when contracting takes place across generations. Future workers cannot be part of a labor contract until they enter the firm (certainly not before they are born). More generally, it is hard to envision who the future partners will be even if they are acting agents in the economy at present. It is also true that, by choice, parties may decide to use short-term contracts if informational asymmetries are present [Hermalin (1986)]. Incomplete contracting, which will require subsequent updating and renegotiation, does not fit the single-contract paradigm either if new parties will enter later.

A major reason for multi-lateral contracting is that agents can enter into side-contracts with each other. On an informal basis this is commonplace in all organizations. Personal relationships and the like fall in this category.⁴⁹ More generally, reciprocation in the conduct of tasks represents side-contracting that cannot fully be controlled by a comprehensive contract. The most explicit form of side-transfers are bribes. They may be paid as monetary compensation for services or they may take more subtle forms – a promotion in exchange for another favor, for instance. It has been alleged that auditing firms occasionally

⁴⁹Such indirect transfers have been emphasized by the Human Relations School; see Etzioni (1964).

obtain favorable contracts from their clients in exchange for good audits. Civil servants are known to have received lucrative job offers after they have quit their government jobs. The list could be extended. The point is that side-contracting in the form of bribes, personal relationships and promises of reciprocation are prevalent. How does this affect the design of incentives and tasks in a hierarchy?

This question, which is truly a major one, has hardly been studied at all, partly because of the analytical complexity. The grand contract design leads to a tractable optimization program, with a manageable set of constraints. By contrast, side-contracting will involve subdesigns by the agents, which generally complicate the analysis considerably. Some headway has been made recently, however, in certain simple agency settings. The approach that is taken is to view the side-contracts as incorporated into the grand design. The principal designs the contract outright so that it leaves no opportunity for the agents to engage in further side-transfers. This approach is not meant to be descriptive of the real situation. Typically, the principal will not be able to control information flows to the extent required for exhausting side-contracting opportunities. However, it is a useful technical device and provides an initial evaluation of the costs of side-contracting. The very fact that there are additional restrictions on the design makes it clear that side-contracting is, in general, costly.

Tirole (1986b) considers side-contracting between an owner, a supervisor and a worker.⁵⁰ The structure is the following. The worker observes the productivity of the technology (which can take only two values) after the contract is drawn. Depending on a random event, the supervisor may or may not observe the productivity. Thus, there are four information states. The owner designs a contract for both the worker and the supervisor, but cannot prevent his two employees from colluding via a side-contact (the owner could also collude with an employee, but this is proved to be worthless in the optimal contract). The owner's contract specifies that the worker and the supervisor report the productivity and as a function of the reports, payments are made and production is ordered.

⁵⁰ Cremer and Riordan (1986) is another paper using a similar approach. They consider a special case in which side-contracting can be made innocuous by a judicious organizational design. In their first model (their analysis holds for more complex hierarchical models), a group of downstream firms contract with an upstream supplier for the procurement of some input. Over the course of their relation the supplier becomes privately informed about its production cost while each customer gets private information about its value for the input. All parties are risk neutral. Cremer and Riordan solve for the optimal grand design, and show that by using expected externality payments, the optimal contract is immune to side-contracting. The intuition is that such payments force each party to internalize the externality imposed by its decisions on the other parties. By adding payments, a group internalizes the externality imposed by its decisions on the rest of the organization.

Demski and Sappington (1984) consider a model in which agents can collude about which equilibrium to play. No side-contracts are involved, because both prefer a different equilibrium than the principal desires. This triggers a design change, which is costly for the principal. However, see Ma and Moore (1985) and Turnbull (1985) on mechanisms that can avoid the problem costlessly.

It is shown that the optimal contract indeed looks different than if the supervisor and agent could be prevented from side-contracting. The contract provides for efficient production in all states, except the one in which the worker alone is aware that productivity is low. Thus, the supervisor is useful. To prevent side-contracting, information rents have to be shared between the worker and the supervisor. Most interestingly, the solution can be interpreted as one in which the worker and the supervisor collude so that the latter acts as an advocate for the former towards the owner. This is a phenomenon that is not surprising to observers of firms. It also points to the general idea that collusion occurs at the nexus of informed parties: shared secrets act as a catalyst for collusion.

The importance of collusion and side-contracting is heavily documented in the sociology literature [e.g. Crozier (1963) and Dalton (1959)]. Collusion is partly issue-dependent and is argued to be conditioned by the structure of information in the way indicated by the supervision model in Tirole (1986b).

To alleviate problems with side-contracting, the organization can try to curb transfers in various ways. This is routinely done for monetary transfers by direct prohibition. Limiting personal relationship (through isolation) is sometimes used as well, but it has obvious drawbacks. Functional transfers are often restricted; the threat of collusion may provide an explanation for limited use of supervisory reports, or for the widespread use of rough and inflexible bureaucratic rules and referral to a superior authority who resolves conflicts due to unforeseen contingencies (bureaucracies are organizations mainly run by rules). This points to some costs of using a grand contract that is coalition-proof, that is, which eliminates the incentives to collude. Supervision and flexibility may be lost in the process.

Finally, the organization can try to restrict reciprocity by promoting short-run relationships between its members through mobility. For instance, consider the extensive use of consulting firms, independent boards, anonymous refereeing or frequent permutations in the civil service and diplomatic corps. Of course, promoting such short-run relationships has drawbacks. They may prevent specific investments in work relationships or the development of trust that is so crucial for cooperation [Tirole (1986b)]. More work needs to be done to formalize how these internal reciprocity games interfere with efficient organizational behavior and how they influence the organizational design.

Side-contracting is a special case of multi-lateral contracting. An interesting multi-lateral contracting problem occurs when one agent works for many principals. This case has been studied by Bernheim and Whinston (1985, 1986). They assume that each principal contracts independently with the common agent; contracts between principals are excluded. The main issue is efficiency of the agent's action, as a result of the efforts of the principals to influence his choice. They show that if the agent is risk neutral, the efficient action is selected. Intuitively, the agent can be made the residual claimant for each of the prin-

cipals' interests. When the agent is risk averse, however, bilateral contracting generally leads to an inefficient action. (The aggregate incentive scheme is efficient conditional on the choice of action.) For this result to obtain it is critical that principals know enough to forecast each others' incentive schemes (in a Nash equilibrium). It would be interesting to consider the case in which other parties' incentives are not known.

Bernheim and Whinston have opened a useful alley of research. Common agency is an important phenomenon, which can be found in wholesaling, government, and central service functions of firms, to name but a few examples.⁵¹ It is also of interest in view of the recent organizational trend towards matrix management in firms. In matrix management subordinates are responsible to several superiors simultaneously. It remains to be seen how well such organizations can cope with the emerging problems of common agency.

5.6. The hierarchy as an authority structure

As in the two-tier case, a major obstacle in designing contracts for a complex organization is the impossibility or the high cost of specifying all the relevant future contingencies. Contracts will necessarily be incomplete and as new contingencies arise, gaps in the contract must be filled through bargaining. One role of authority within the organization is to constrain the bargaining process by designating a decision-maker in case of disagreement. Authority – its scope and entitlements – is a rather elusive concept. For instance, consider scope. Because contingencies are not precisely specified in an initial contract, neither is the exact set of decisions from which the party with the authority can choose. An engineer may have the authority to introduce a new technology for workers, but at the same time he may be prevented from choosing exhausting or potentially dangerous technologies.

The rights of authority at the firm level are defined by ownership of assets, tangible (machines or money) or intangible (goodwill or reputation). The distribution of authority rights comes from the delegation of the owner's authority to lower level functions (managers, foremen, etc.), usually in a nested fashion (one manager can delegate forward within his set of rights). A production manager or foreman is free, within limits, to reorganize his shop to adjust to new circumstances. A production worker can decide on how he carries out a task. The

⁵¹Baron (1985) analyzes a common agency problem arising from the regulation of a public utility by the Environmental Protection Agency (EPA) as well as a public utility commission. He shows that in a non-cooperative equilibrium, the EPA chooses more stringent abatement standards and maximum allowable emission fees than those which the two regulators would choose in a cooperative equilibrium. The PUC, which must provide the firm with a fair return, chooses higher prices than in the cooperative equilibrium.

allocation of decision rights within the firm is obviously a central issue and one could envision an approach to hierarchies based on the analysis of incomplete contracts. At this point in time such a theory is still to emerge. Hence, we will restrict ourselves to comments on some features of authority that can be expected to play an important role in any analysis.

The notion of authority through asset ownership is more distinct than delegated authority. The former has a fairly clear-cut legal meaning and is conferred by written document. By contrast, delegated authority is in most cases conferred orally and is revocable by simple declaration. The legal implications of delegated authority have been discussed by some legal scholars [Conrad et al. (1982)]. The very purpose of delegated authority may well be to avoid constant recourse by third parties to the principal. This implies that third parties must be able to transact with the agent with a minimum of inquiry as to her authority. The delegation is thus based on a common understanding of how the organization works, which must be shared by the principal, the agent and the third party. Because incomplete contracts are the basis for authority, we must look for a rule that gives legitimacy to non-contractible actions taken by the agent on behalf of the principal. Legal systems generally define authority by usage: either the authority is implied by the position or it is circumstantial; in both cases, authority is thought to be legitimate if it corresponds to good practice or prevailing customs.⁵² Of course, this common usage definition of authority is still ambiguous, as witnessed by the courts' very diverse interpretations of the powers of a CEO.

The struggle for a clear legal definition of delegated authority has its counterpart in corporate organizations, where internal arbitration may replace the judicial system. Making all members understand who is deciding what in yet unforeseen circumstances is a perilous, but important, exercise in organization behavior, and its outcome can be seen as a part of corporate culture or organizational capital. If a common understanding fails, disagreements, conflicts of authority, and noncommittal delays and use of upper-management time result. An important aspect of authority within an organization is its vertical structure. Most conflicts between divisions or employees are solved by higher authorities, for instance chief executives, rather than by courts. Williamson (1975), in particular, has emphasized the superiority of internal organization in dispute settling matters.

This leads us to enquire about the requisite qualities of an arbitrator (be she a court or a superior). First the arbitrator must have a good knowledge of the situation to try to duplicate the outcome of the missing optimal comprehensive contract. Second, she must be independent. With respect to the first quality,

⁵²The German procura system goes much further and allows the agent to bind the principal in all transactions but transfers of real estate.

external arbitrators, like courts, are likely to incur a cost of becoming informed. This cost also exists for superiors in an organization; in particular, in large firms the chief executives may be overloaded with decisions to arbitrate between their subordinates and have little *a priori* knowledge of each case; but because of everyday interaction, as well as a past familiarity with various jobs within the firm, internal arbitrators may incur a lower information cost. The second quality, independence, requires that the arbitrator not be judge and party, so as to value aggregate efficiency beyond the interest of any party. Side-contracting with the arbitrator must be prevented. Independence may fail, for instance, when the arbitrator has kept close ties with one of the involved divisions. More generally, arbitrators must have a reputation for settling disputes "fairly" (understand: "efficiently").

5.7. *Organizational forms*

As mentioned in the Introduction, organizational forms are related to several factors. At a given point of time, a firm's organization is meant to promote communication and incentives. The capital structure, the outside visibility of managers, the internal job market, the auditing and supervisory designs, the structure of competition between agents are all geared to this purpose. The organizational model is also conditioned by the current knowledge of how various types of organization work. The cost of experimenting (associated with both the possibility of mistake and the cost of training employees to learn the new rules of the game) explains both the predominant role of history and the existence of fads in organizational innovations. Last, the organizational model depends on the economic environment, including factor prices, and on the growth of the firm.

Examples of organizational innovations are the apparition of the U- and M-forms documented by Chandler (1966) and Williamson (1975). A reading of these innovations in the light of our survey might go as follows.

The U-form (unitary form) gathers activities according to their function within the firm: for example, auditing, marketing, finance, materials procurement, production. The Viner-like rationale for this gathering is to avoid a duplication of costs associated with each function.

The drawback of the U-form is, of course, the team problems (*à la* Alchian and Demsetz) that it may create. The performance of a product in a market depends on its design (R&D department), the quality of manufacturing (production), post-sale services (maintenance department), marketing efforts (marketing department), and so on. The difficulty of measuring individual performance is that it requires careful supervision and a good understanding of each functional division by the CEO and top managers. The latter become easily overloaded as

the firm grows. And, indeed, the U-form collapsed with the horizontal expansion of firms [Chandler (1966)]. It was replaced by the M-form (multidivisional form), which resembles a collection of scaled-down U-form structures. In the M-form, divisions are organized so that their performance can be reliably measured. Distinction by product categories is most likely to achieve this goal. The role of the top management is then reduced to advising, auditing and allocating resources between the competing divisions. Within a division, by contrast, the supervisory mode is more prevalent and allows some assessment of the relative contributions of functional subdivisions.

As can be seen, the switch from the U-form to the M-form was partly triggered by a changing environment. A more recent example of this phenomenon is the matrix organization. Among other things, matrix organizations try to promote horizontal communication and decision-making (for instance, between marketing, R&D and production managers). The need for joint decision-making was made more acute by the gradual shortening of the life-cycle of products. In an industry where products become obsolete within a year, firms must be particularly quick at finding the right market niches. One may predict that future organizational innovations will follow. Already some discontent has been recorded concerning, for instance, the high number of authority conflicts between the joint decision-makers, which creates a high demand for time-consuming arbitration by top executives. [For an account of recent organizational developments, see Piore (1986) and Piore and Sabel (1985).] We are unaware of formal, agency-related work modelling these organizational forms.

6. Conclusion

This chapter has been deliberately issue-oriented. Rather than recapitulating existing techniques and results, we outlined the main lines of research and unveiled many open questions. Despite the tremendous progress made by organization theory over the last fifteen years, we still have a weak understanding of many important facets of organizational behavior.

There are at least three outstanding problems that need attention. A first (theoretical) step is to develop and apply techniques that deal with non-standard problems such as incomplete contracts, bounded rationality and multi-lateral contracting. The second step ought to integrate observations from neighboring fields such as sociology and psychology – in a consistent (not ad hoc) way into the theoretical apparatus. The third step will be to increase the evidence/theory ratio, which is currently very low in this field. While informational asymmetries, contractual incompleteness or imperfect communication will typically be hard to measure, empirical research such as Joskow's (1985) (applying Williamson's ex-post bilateral monopoly problem to contracts between coal mines and electric

utilities) or Wolfson's (1985) (applying incentive theory to oil drilling) raise hopes that the economic approach to organizations will be more carefully tested in the near future.

References

- Admati, A. and Perry, M. (1987) 'Strategic delay in bargaining', *Review of Economic Studies*, 54:345–364.
- Aghion, P. and Bolton, P. (1986) 'An 'incomplete contracts' approach to bankruptcy and the optimal financial structure of the firm', University of California, Berkely, mimeo.
- Aghion, P. and Bolton, P. (1987) 'Contracts as a barrier to entry', *American Economic Review*, 77:388–401.
- Akerlof, G. (1970) 'The market for 'lemons': Quality and the market mechanism', *Quarterly Journal of Economics*, 84:488–500.
- Alchian, A. (1968) 'Corporate management property rights', in: *Economic policy and the regulation of securities*. Washington, D.C.: American Enterprise Institute.
- Alchian, A. and Demsetz, H. (1972) 'Production, information costs, and economic organization', *American Economic Review*, 62:777–795.
- Antle, R. and Demski, J. (1987) 'Revenue recognition', Yale University, mimeo.
- Antle, R. and Smith, A. (1986) 'An empirical investigation of the relative performance evaluation of corporate executives', *Journal of Accounting Research*, 24:1–39.
- Aron, D. (1988) 'Ability, moral hazard, firm size and diversification', *Rand Journal of Economics*, 19:72–87.
- Arrow, K. (1974) *The limits of organization*. New York: Norton.
- Arrow, K. (1985) 'The economics of agency', in: J. Pratt and R. Zeckhauser, eds., *Principals and agents: The structure of business*. Boston: Harvard Business School Press, 37–51.
- Baiman, S. (1982) 'Agency research in managerial accounting: A survey', *Journal of Accounting Literature*, 1:154–213.
- Baiman, S. and Demski, J. (1980) 'Economically optimal performance evaluation and control systems', *Journal of Accounting Research*, 18(Supplement):184–220.
- Baron, D. (1985) 'Noncooperative regulation of a nonlocalized externality', *Rand Journal of Economics*, 16:553–568.
- Baumol, W. (1959) *Business behavior, value and growth*. New York: Macmillan.
- Baumol, W., Panzer, J. and Willig, R. (1982) *Contestable markets and the theory of industry structure*. New York: Harcourt Brace Jovanovich.
- Bernheim, D. and Whinston, M. (1985) 'Common marketing agency as a device for facilitating collusion', *Rand Journal of Economics*, 16:269–281.
- Bernheim, D. and Whinston, M. (1986) 'Common agency', *Econometrica*, 54:932–942.
- Bhattacharya, S. (1983) 'Tournaments and incentives: Heterogeneity and essentiality', research paper no. 695, Graduate School of Business, Stanford University.
- Blackwell, D. (1953) 'Equivalent comparisons of experiments'. *Annals of Mathematical Statistics*, 24:265–272.
- Bonanno, G. and Vickers, J. (1986) 'Vertical separation', Nuffield College, Oxford, mimeo.
- Bradley, M., Desai, A. and Kim, E. (1987), 'Synergistic gains from corporate acquisitions and the division between the stockholders of target and acquiring firms,' *Journal of Financial Economics*, forthcoming.
- Brander, J. and Lewis, T. (1986) 'Oligopoly and financial structure: The limited liability effect', *American Economic Review*, 76:956–970.
- Caillaud, B. (1986) 'Regulation, competition and asymmetric information', *Journal of Economic Theory*, forthcoming.
- Caillaud, B., Guesnerie, R. and Rey, P. (1986) 'Noisy observation in adverse selection models', INSEE, mimeo.
- Caillaud, B., Guesnerie, R., Rey, P. and Tirole, J. (1988) 'Government intervention in production and incentives theory: A review of recent contributions', *Rand Journal of Economics*, 19:1–26.

- Calvo, G. and Wellisz, S. (1978) 'Supervision, loss of control, and the optimal size of the firm', *Journal of Political Economy*, 86:943–952.
- Calvo, G. and Wellisz, S. (1979) 'Hierarchy, ability, and income distribution', *Journal of Political Economy*, 87:991–1010.
- Carmichael, L. (1983) 'Firm-specific human capital and promotion ladders', *Bell Journal of Economics*, 14:251–258.
- Chandler, A. (1966) *Strategy and structure*. New York: Doubleday and Company (Anchor Book Edition).
- Clark, R. (1985) 'Agency costs versus fiduciary duties', in: J. Pratt and R. Zeckhauser, eds., *Principals and agents: The structure of business*. Cambridge, Mass.: Harvard Business School Press.
- Conrad, A., Knauss, R. and Spiegel, S. (1982) *Enterprise organization*, 3rd ed. Foundation Press.
- Corey, R. (1978) 'Should companies centralize procurement?' *Harvard Business Review*, 56:102–110.
- Crampes, C. (1983) 'Subventions et régulations d'une entreprise privée', *Annales de l'INSEE*, 51:47–63.
- Cramton, P. (1984) 'Bargaining with incomplete information: An infinite horizon model with two-sided uncertainty', *Review of Economic Studies*, 51:529–593.
- Crawford, V. (1986) 'Long term relationships governed by short term contracts', working paper no. 205, Industrial Relations Section, Princeton University.
- Cremer, J. (1986a) 'Cooperation in ongoing organizations', *Quarterly Journal of Economics*, 101:33–50.
- Cremer, J. (1986b) 'Corporate culture: Cognitive aspects', Virginia Polytechnic Institute, mimeo.
- Cremer, J. and Riordan, M. (1986) 'On governing multilateral transactions with bilateral contracts', discussion paper no. 134, Studies in Industry Economics, Stanford University.
- Crozier, M. (1963) *Le phénomène bureaucratique*. Paris: Éditions du Seuil; (translated by Crozier as *The bureaucratic phenomenon*, University of Chicago Press, 1967).
- Cyert, R. and March, J. (1963) *A behavioral theory of the firm*. Englewood Cliffs, N.J.: Prentice-Hall.
- Dalton, M. (1959) *Men who manage*, New York: Wiley.
- De Alessi, L. (1983) 'Property rights, transaction costs, and *X*-efficiency', *American Economic Review*, 73:64–81.
- Demski, J. and Feltham, G. (1976) *Cost determination: A conceptual approach*. Ames, Iowa: The Iowa University Press.
- Demski, J. and Sappington, D. (1984) 'Optimal incentive contracts with multiple agents', *Journal of Economic Theory*, 17:152–171.
- Demski, J. and Sappington, D. (1986) 'Line item reporting, factor acquisition and subcontracting', *Journal of Accounting Research*, 24:250–269.
- Demski, J., Sappington, D. and Spiller, P. (1987) 'Managing supplier switching', *Rand Journal of Economics*, 18:77–97.
- Diamond, D. (1984) 'Financial intermediation and delegated monitoring', *Review of Economic Studies*, 51:393–444.
- Doeringer, P. and Piore, M. (1971) *Internal labor markets and manpower analysis*. Lexington, Mass.: Heath.
- Dye, R. (1984) 'The trouble with tournaments', *Economic Inquiry*, 22:147–149.
- Eswaran, M. and Kotwal, A. (1984) 'The moral hazard of budget-breaking', *Rand Journal of Economics*, 15:578–581.
- Etzioni, A. (1964) *Modern organizations*. Englewood Cliffs, N.J.: Prentice-Hall.
- Evans, D. (1986) 'Tests of alternative theories of firm growth', research report no. 86-36, C.V. Starr Center for Applied Economics, New York University.
- Fama, E. (1980) 'Agency problems and the theory of the firm', *Journal of Political Economy*, 88:288–307.
- Farrell, J. and Gallini, N. (1986) 'Second-sourcing as a commitment: Monopoly incentives to attract competition', *Quarterly Journal of Economics*, forthcoming.
- Fershtman, C. and Judd, K. (1986) 'Strategic incentive manipulation in rivalrous agency', IMSSS TR 496, Stanford University.
- Fershtman, C., Judd, K. and Kalai, E. (1986) 'Cooperation and the strategic aspect of delegation', Northwestern University, mimeo.
- Freixas, X., Guesnerie, R. and Tirole, J. (1985) 'Planning under incomplete information and the ratchet effect', *Review of Economic Studies*, 52:173–191.

- Friedman, B. (1985) *Corporate capital structures in the United States*. Chicago: University of Chicago Press.
- Fudenberg, D. and Tirole, J. (1983) 'Sequential bargaining with incomplete information', *Review of Economic Studies*, 50:221–247.
- Fudenberg, D., Holmstrom, B. and Milgrom, P. (1986) 'Short term contracts and long term agency relationships', School of Organization and Management, Yale University, draft.
- Galbraith, K. (1967) *The new industrial state*. Boston: Houghton Mifflin.
- Gale, D., and Hellwig, M. (1985) 'Incentive compatible debt contracts: The one period problem', *Review of Economic Studies*, 52:647–664.
- Geanakopoulos, J. and Milgrom, P. (1985) 'A theory of hierarchies based on limited managerial attention', Cowles Foundation paper no. 775, Yale University.
- Gertner, R., Gibbons, R. and Scharfstein, D. (1987) 'Simultaneous signalling to the capital and product markets', Massachusetts Institute of Technology, mimeo.
- Gibbons, R. (1985) 'Optimal incentive schemes in the presence of career concerns', Massachusetts Institute of Technology, mimeo.
- Gjesdal, F. (1982) 'Information and incentives: The agency information problem', *Review of Economic Studies*, 49:373–390.
- Green, J. and Stokey, N. (1983) 'A comparison of tournaments and contests', *Journal of Political Economy*, 91:349–364.
- Grossman, S. and Hart, O. (1977) 'On value maximization and alternative objectives of the firm', *Journal of Finance*, 32:389–440m.i.t.2.
- Grossman, S. and Hart, O. (1980) 'Takeover bids, the free rider problem, and the theory of the corporation', *Bell Journal of Economics*, 11:42–64.
- Grossman, S. and Hart, O. (1982) 'Corporated financial structure and managerial incentives', in: J. McCall, ed., *The economics of information and uncertainty*. Chicago: University of Chicago Press, 107–137.
- Grossman, S. and Hart, O. (1983) 'An analysis of the principal–agent problem', *Econometrica*, 51:7–45.
- Grossman, S. and Hart, O. (1986) 'The costs and benefits of ownership: A theory of vertical and lateral integration', *Journal of Political Economy*, 94:691–719.
- Grossman, S. and Hart, O. (1987) 'One share/one vote and the market for corporate control', working paper, Department of Economics, Massachusetts Institute of Technology.
- Grossman, S., Hart, O. and Maskin, E. (1983) 'Unemployment with observable aggregate shocks', *Journal of Political Economy*, 91:907–928.
- Guasch, L. and Weiss, A. (1980) 'Wages as sorting mechanisms in competitive markets with asymmetric information: A theory of testing', *Review of Economic Studies*, 47:653–664.
- Harris, M. and Raviv, A. (1979) 'Optimal incentive contracts with imperfect information', *Journal of Economic Theory*, 20:231–259.
- Harris, M. and Raviv, A. (1985) 'Corporate control contests and capital structure', working paper, Northwestern University.
- Harris, M. and Raviv, A. (1987) 'Corporate governance: Voting rights and majority rules', working paper, Department of Finance, Northwestern University.
- Hart, O. (1983a) 'The market mechanism as an incentive scheme', *Bell Journal of Economics*, 74:366–382.
- Hart, O. (1983b) 'Optimal labour contracts under asymmetric information: An introduction', *Review of Economic Studies*, 50:3–35.
- Hart, O. and Holmstrom, B. (1987) 'The theory of contracts', in: Trueman Bewley, ed., *Advances in economic theory, Fifth World Congress*. Cambridge: Cambridge University Press.
- Hart, O. and Moore, J. (1988) 'Incomplete contracts and renegotiation', *Econometrics*, 56:755–786.
- Hermalin, B. (1986) 'Adverse selection and contract length', Massachusetts Institute of Technology, mimeo.
- Hermalin, B. (1987) 'Adverse effects of the threat of takeovers', Massachusetts Institute of Technology, mimeo.
- Hermalin, B. and Weisbach, M. (1987) 'The determinants of board composition', Massachusetts Institute of Technology, mimeo.
- Holmstrom, B. (1979) 'Moral hazard and observability', *Bell Journal of Economics*, 10:74–91.

- Holmstrom, B. (1982a) 'Moral hazard in teams', *Bell Journal of Economics*, 13:324–340.
- Holmstrom, B. (1982b) 'Managerial incentive problems – A dynamic perspective', in: *Essays in economics and management in honor of Lars Wahlbeck*. Helsinki: Swedish School of Economics.
- Holmstrom, B., and Milgrom, P. (1987) 'Aggregation and linearity in the provision of intertemporal incentives', *Econometrica*, 55:303–328.
- Holmstrom, B. and Ricart i Costa, J. (1986) 'Managerial incentives and capital management', *Quarterly Journal of Economics*, 101:835–860.
- Holmstrom, B. and Weiss, L. (1985) 'Managerial incentives, investment and aggregate implications: Scale effects', *Review of Economic Studies*, 52:403–426.
- Hopenhayn, H. (1986) 'A competitive stochastic model of entry and exit to an industry', University of Minnesota, mimeo.
- Jacobs, D. (1981) 'Towards a theory of mobility and behavior in organizations: An inquiry into the consequences of some relationships between individual performance and organizational success', *American Journal of Sociology*, 87:684–707.
- Jensen, M. (1986) 'Agency costs of free cash flow, corporate finance and takeovers', *American Economic Review*, 76:323–329.
- Jensen, M. and Meckling, W. (1976) 'Theory of the firm: Managerial behavior, agency costs, and capital structure', *Journal of Financial Economics*, 3:305–360.
- Joskow, P. (1985) 'Vertical integration and long term contracts: The case of coal-burning electric generating plants', *Journal of Law, Economics and Organization*, 1:33–80.
- Jovanovic, B. (1982) 'Selection and the evolution of industry', *Econometrica*, 50:649–670.
- Katz, M. (1987) 'Game-playing agents: Contracts as precommitments', Princeton University, mimeo.
- Keren, M. and Levhari, D. (1983) 'The internal organization of the firm and the shape of average costs', *Bell Journal of Economics*, 14:474–486.
- Kihlstrom, R. and Laffont, J.-J. (1979) 'A general equilibrium entrepreneurial theory of the firm based on risk aversion', *Journal of Political Economy*, 87:719–748.
- Klein, B., Crawford, R. and Alchian, A. (1978) 'Vertical integration, appropriable rents and the competitive contracting process', *Journal of Law and Economics*, 21:297–326.
- Kreps, D. (1984) 'Corporate culture and economic theory', Graduate School of Business, Stanford University, mimeo.
- Kreps, D. and Wilson, R. (1982) 'Reputation and imperfect information', *Journal of Economic Theory*, 27:253–279.
- Laffont, J.-J. and Maskin, E. (1982) 'The theory of incentives: An overview', in: W. Hildenbrand, ed., *Advances in economic theory*. New York: Cambridge University Press.
- Laffont, J.-J. and Tirole, J. (1986) 'Using cost observation to regulate firms', *Journal of Political Economy*, 94:614–641.
- Laffont, J.-J. and Tirole, J. (1987a) 'Auctioning incentive contracts', *Journal of Political Economy*, 95:921–937.
- Laffont, J.-J. and Tirole, J. (1987b) 'Repeated auctions of incentive contracts, investment and bidding parity, with an application to takeovers', discussion paper no. 463, Massachusetts Institute of Technology.
- Laffont, J.-J. and Tirole, J. (1988) 'The dynamics of incentive contracts', *Econometrica*, September.
- Lambert, R. (1983) 'Long-term contracting and moral hazard', *Bell Journal of Economics*, 14:441–452.
- Lazear, E. (1986) 'Pay equality and industrial politics', Hoover institution, Stanford University, mimeo.
- Lazear, E. and Rosen, S. (1981) 'Rank-order tournaments as optimum labor contracts', *Journal of Political Economy*, 89:841–864.
- Leibenstein, H. (1966) 'Allocative efficiency as 'X-efficiency'', *American Economic Review*, 56:392–415.
- Leland, H. and Pyle, D. (1977) 'Information asymmetries, financial structure, and financial inatermediaries', *Journal of Finance*, 32:371–387.
- Lippman, S. and Rumelt, R. (1982) 'Uncertain imitability: An analysis of interfirm differences in efficiency under competition', *Bell Journal of Economics*, 13:418–438.
- Lucas, R. (1967) 'Adjustment costs and the theory of supply', *Journal of Political Economy*, 75:321–339.
- Lucas, R. (1978) 'On the size distribution of business firms', *Bell Journal of Economics*, 9:508–523.

- Lydall, H. (1968) *The structure of earnings*. New York: Oxford University Press.
- Ma, C. and Moore, J. (1985) 'Stopping agents cheat', working paper, London School of Economics.
- MacCaulay, S. (1963) 'Nob-contractual relations in business', *American Sociological Review*, 28:55–70.
- MacDonald, G. (1982) 'Information in production', *Econometrica*, 50:1143–1162.
- MacDonald, G. (1984) 'New directions in the economic theory of agency', *Canadian Journal of Economics*, 17:415–440.
- Mace, M. (1971) 'Directors: Myth and reality'. Graduate School of Business Administration, Harvard University.
- MacLeod, B. and Malcomson, J. (1985) 'Reputation and hierarchy in dynamic models of employment', Queen's University, mimeo.
- Malcomson, J. (1984) 'Work incentives, hierarchy, and internal labor markets', *Journal of Political Economy*, 92:486–507.
- Malcomson, J. and Spinnewyn, F. (1988) 'The multiperiod principal agent problem', *Review of Economic Studies*, 55:391–408.
- Mansfield, E. (1962) 'Entry, Gibrat's law, innovation and the growth of firms', *American Economic Review*, 52:1023–1051.
- Marschak, J. and Radner, R. (1972) *Economic theory of teams*. New Haven: Yale University Press.
- Maskin, E. and Riley, J. (1985) 'Input vs. output incentive schemes,' *Journal of Public Economics*, 28:1–23.
- Masten, S. (1986a) 'Institutional choice and the organization of production: The make-or-buy decision', *Journal of Institutional and Theoretical Economics*, 142:493–509.
- Masten, S. (1986b) 'The institutional basis for the firm', working paper no. 473, Graduate School of Business Administration, The University of Michigan.
- Mathewson, F. and Winter, R. (1985) 'Is exclusive dealing anticompetitive?', working paper no. 8517, University of Toronto.
- McAfee, P. and McMillan, J. (1987) 'Competition for agency contracts', *Rand Journal of Economics*, 18:296–307.
- Melumad, N. and Reichelstein, S. (1986) 'Value of communication in agencies', Graduate School of Business, Stanford University, mimeo.
- Milgrom, P. (1988) 'Employment contracts, influence activities and efficient organization design', *Journal of Political Economy*, 96:42–60.
- Milgrom, P. and Oster, S. (1987) 'Job discrimination, market forces and the invisibility hypothesis', *Quarterly Journal of Economics*, 102:453–476.
- Milgrom, P. and Roberts, J. (1982) 'Predation, reputation and entry deterrence', *Journal of Economic Theory*, 27:280–312.
- Milgrom, P. and Roberts, J. (1987) 'Bargaining and influence costs and the organization of economic activity', research paper no. 934, Graduate School of Business, Stanford University.
- Mirrlees, J. (1974) 'Notes on welfare economics, information and uncertainty', in: M. Balch, D. McFadden and S. Wu, eds., *Essays in economic behavior under uncertainty*. Amsterdam: North-Holland, 243–258.
- Mirrlees, J. (1976) 'The optimal structure of authority and incentives within an organization', *Bell Journal of Economics*, 7:105–131.
- Modigliani, F. (1963) 'Corporate income taxes and the cost of capital: A correction', *American Economic Review*, 53:433–443.
- Modigliani, F. and Miller, M. (1958) 'The costs of capital, corporation finance, and the theory of investment', *American Economic Review*, 48:261–297.
- Mookerjee, D. (1984) 'Optimal incentive schemes with many agents', *Review of Economics Studies*, 51:433–446.
- Morck, R., Shleifer, A. and Vishny, R. (1986) 'Ownership structure and corporate performance: An empirical analysis', NBER: Cambridge, Mass., mimeo.
- Murphy, K. (1984) 'Incentives, learning and compensation: A theoretical and empirical investigation of managerial labor contracts', *Rand Journal of Economics*, 17:59–76.
- Myers, S. and Majluf, N. (1984) 'Corporate financing and investment decisions when firms have information that investors do not have', *Journal of Financial Economics*, 13:187–221.
- Myerson, R. (1984) 'Multistage games with communication', discussion paper no. 590, CMSEMS, Northwestern University.

- Nalebuff, B. and Stiglitz, J. (1983) 'Prizes and incentives: Towards a general theory of compensation and competition', *Bell Journal of Economics*, 14:21–43.
- Narayanan, M.P. (1985) 'Managerial incentives for short-term results', *Journal of Finance*, 40(5):1469–1484.
- Nelson, R. and Winter, S. (1982) *An evolutionary theory of economic change*. Cambridge, Mass.: Harvard University Press.
- Penrose, E. (1959) *The theory of the growth of the firm*. New York: Wiley.
- Peters, T. and Waterman, R. (1982) *In search of excellence*. New York: Warner Books.
- Picard, P. (1987) 'On the design of incentive schemes under moral hazard and adverse selection', *Journal of Public Economics*, 33:305–332.
- Piore, M. (1986) 'Corporate reform in American manufacturing and the challenge to economic theory', Massachusetts Institute of Technology, mimeo.
- Piore, M. and Sabel, C. (1985) *The second industrial divide*. New York: Basic Books.
- Prescott, E. and Visscher, M. (1980) 'Organization capital', *Journal of Political Economy*, 88:446–461.
- Radner, R. (1981) 'Monitoring cooperative agreements in a repeated principal–agent relationship', *Econometrica*, 49:1127–1148.
- Radner, R. (1985) 'Repeated principal–agent games with discounting', *Econometrica*, 53:1173–1198.
- Rey, P. and Salanié, B. (1986) 'Long term, short term, and renegotiation', INSEE, mimeo.
- Rey, P. and Stiglitz, J. (1986) 'The role of exclusive territories in producers' competition', Princeton University, mimeo.
- Rey, P. and Tirole, J. (1986) 'The logic of vertical restraints', *American Economic Review*, 76:921–939.
- Ricart i Costa, J. (1986) 'Managerial task assignment and promotions', Universitat Autònoma de Barcelona, mimeo.
- Rogerson, W. (1985) 'Repeated moral hazard', *Econometrica*, 53:69–76.
- Rosen, S. (1982) 'Authority, control, and the distribution of earnings', *Bell Journal of Economics*, 13:311–323.
- Rosen, S. (1986) 'Prizes and incentives in elimination tournaments', *American Economic Review*, 76:701–715.
- Ross, S. (1973) 'The economic theory of agency: The principal's problem', *American Economic Review*, 63:134–139.
- Ross, S. (1977) 'The determination of financial structure: The incentive signalling approach', *Bell Journal of Economics*, 8:23–40.
- Ross, S. (1978) 'Some notes on financial incentive signalling models, activity choice and risk preferences', *Journal of Finance*, 33:777–792.
- Rothschild, M. (1974) 'A two-armed bandit theory of market pricing', *Journal of Economic Theory*, 9:185–202.
- Rubinstein, A. (1979) 'Offenses that may have been committed by accident – An optimal policy of retribution', in: S. Brahm, A. Shotter and G. Schrodauer, eds., *Applied game theory*. Wurtzburg: Physica-Verlag, 406–413.
- Sappington, D. (1984) 'Incentive contracting with asymmetric and imperfect precontractual knowledge', *Journal of Economic Theory*, 34:52–70.
- Scharfstein, D. (1988a) 'Product market competition and managerial slack', *Rand Journal of Economics*, 19:147–155.
- Scharfstein, D. (1988b) 'The disciplinary role of takeovers', *Review of Economic Studies*, 55:185–200.
- Schein, E. (1984) 'Coming to a new awareness of organizational culture', *Sloan Management Review*, 25:3–16.
- Scherer, F. (1980) *Industrial market structure and economic performance*, 2nd edn. Chicago: Rand McNally College Publishing Company.
- Shavell, S. (1979) 'Risk sharing and incentives in the principal and agent relationship', *Bell Journal of Economics*, 10:55–73.
- Shepard, A. (1987) 'Licensing to enhance demand for new technologies', *Rand Journal of Economics*, 18:360–368.
- Shleifer, A. (1985) 'A theory of yardstick competition', *Rand Journal of Economics*, 16:319–327.
- Shleifer, A. and Summers, L. (1988) 'Hostile takeovers and breaches of trust', in: A. Auernbach, ed., *Corporate takeovers: Causes and consequences*. Chicago: The University of Chicago Press, 33–56.

- Shleifer, A. and Vishny, R. (1986a) 'Large shareholders and corporate control', *Journal of Political Economy*, 94:461–488.
- Shleifer, A. and Vishny, R. (1986b) 'Greenmail white knights and shareholders' interest', *Rand Journal of Economics*, 17:293–309.
- Simon, H. (1951) 'A formal theory of the employment relationship', *Econometrica*, 19:293–305.
- Simon, H. (1957) *Models of man*. New York: Wiley.
- Simon, H. (1976) *Administrative behavior*, 3rd edn. London: Macmillan.
- Sobel, J. and Takahashi, I. (1983) 'A multistage model of bargaining', *Review of Economic Studies*, 50:411–426.
- Spence, M. (1973) 'Job market signalling', *Quarterly Journal of Economics*, 87:355–374.
- Spence, M. and Zeckhauser, R. (1971) 'Insurance, information and individual action', *American Economic Review Papers and Proceedings*, 61:380–387.
- Stein, J. (1987) 'Efficient stock prices, inefficient firms: A signal-jamming model of myopic corporate behavior', Harvard Business School, mimeo.
- Stein, J. (1988) 'Takeover threats and managerial myopia', *Journal of Political Economy*, 96:61–80.
- Stiglitz, J. (1975) 'Incentives, risk and information: Notes towards a theory of hierarchy', *Bell Journal of Economics*, 6:552–579.
- Tirole, J. (1986a) 'Procurement and renegotiation', *Journal of Political Economy*, 94:235–259.
- Tirole, J. (1986b) 'Hierarchies and bureaucracies', *Journal of Law, Economics and Organization*, 2:181–214.
- Turnbull, S. (1985) 'The revelation principle and the 'principal pet' in multiple agency problems', working paper, the Ohio State University.
- Vassilakis, S. (1985) 'On the division of labor', Johns Hopkins University, mimeo.
- Viner, J. (1932) 'Cost curves and supply curves', *Zeitschrift für National-ökonomie*, 3:23–46.
- Waldman, M. (1984) 'Job assignments, signalling, and efficiency', *Rand Journal of Economics*, 15:255–270.
- Weitzman, M. (1974) 'Prices versus quantities', *Review of Economic Studies*, 41:477–491.
- Weitzman, M. (1980) 'The 'ratchet principle' and performance incentives', *Bell Journal of Economics*, 11:302–308.
- Williamson, O. (1964) *The economics of discretionary behavior: Managerial objectives in a theory of the firm*. Englewood Cliffs, N.J.: Prentice-Hall.
- Williamson, O. (1967) 'Hierarchical control and the optimal firm size', *Journal of Political Economy*, 75:123–138.
- Williamson, O. (1975) *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.
- Williamson, O. (1985) *The economic institutions of capitalism*. New York: Free Press.
- Wilson, R. (1969) 'The structure of incentives of decentralization under uncertainty', in: M. GUILBAUD, ed., *La décision*. Paris: CNRS.
- Wilson, R. (1986) 'The theory of syndicates', *Econometrica*, 36:119–132.
- Wolfson, M. (1985) 'Empirical evidence of incentive problems and their mitigation in oil and tax shelter programs', in: J. Pratt and R. Zeckhauser, eds., *Principals and agents: The structure of business*. Boston: Harvard Business School Press, 101–125.