The Internal Organization of Firms

Margaret A. Meyer

Nuffield College

Oxford OX1 1NF, England

August, 1992

Acknowledgment: This paper was first presented as the *Review of Economic Studies* Lecture at the Royal Economic Society Conference, London, March 1992. I am grateful to the audience there and to one at Oxford for helpful comments.

Keywords: internal organization, job design, career design, delegation, integration, divestiture

1. Introduction

This paper discusses a selection of recent developments in the economic analysis of the internal organization of firms.

Analyses of principal-agent relationships have been extremely influential in demonstrating the importance of illuminating the "black box" that the internal organization of firms used to be. Principal-agent analyses demonstrated that differences in objectives and in information could together result in deviations from allocatively efficient outcomes, and hence from profit-maximizing behavior. They thus convincingly showed that compensation of employees and managers (and, by implication, industrial relations and corporate governance) could significantly affect firm performance.

Nevertheless, early principal-agent analyses had relatively few implications for organizational structure. At a micro level, the division of labor between principal and agent, or among agents, was taken as exogenously determined. Therefore, the implications of informational considerations for the design of jobs or the extent of delegation of responsibility were left largely unexplored. At a more macro level, the organizational context of the contracts studied was in fact immaterial. Contracts were assumed to be negotiated once and for all, and never to need revision or augmentation. It was therefore irrelevant whether the principal and the agent were members of the same or different firms, that is, it was irrelevant whether the transaction studied was conducted through the market or within an organizational hierarchy. While this irrelevance allowed the key insights from principal-agent theory to be applied to a wide range of interactions, including taxation and regulation, it did limit the theory's ability to explain the determinants of the boundaries of firms.

This paper is organized around four headings: job design, career design, internal structure, and the boundaries of firms. For each heading, I have selected one particular issue to discuss. These issues have been explored in recent theoretical work, by myself and others. Each one is also of current interest to managers and management educators, and current practice in these areas has undergone changes in the last decade. While my discussion of recent trends concerns private-sector organizations, most of the issues, and the formal analysis of them, are relevant to public-sector organizations as well.

2. Microstructure: Job Design and Career Design

2a. Job Design: Flexible vs. Rigid Jobs

Within the area of job design, I will analyze the question: What determines how flexibly or rigidly jobs are defined?

If firms could perfectly monitor how their employees spent their time and knew their employees' abilities, then the design of jobs would be a straightforward matter: firms would calculate the profit-maximizing allocation of time across activities for each employee and ensure that it was implemented. If, however, an employee's allocation of time cannot be monitored, then new issues arise. To induce employees to allocate their time appropriately across different activities, how much use should the firm make of performance-related pay and how much should it rely on rigid restrictions on conduct?

Holmstrom and Milgrom (1991) have analyzed this question in a model in which a risk-averse employee privately chooses how much attention to devote to a company task and how much to personal business activities. Returns from time spent on the company task accrue to the firm, and performance on the company task can be measured, though with error. By contrast, for personal business activities, although they may be socially valuable, the returns accrue to the employee. Furthermore, the firm has no way of monitoring the outcome of personal business activities. It is assumed, though, that the firm can prohibit the employee from spending time on designated personal business activities, by imposing rigid rules and procedures. For example, one type of personal business activity is networking within the firm. The firm could prevent the employee from networking during working hours by restricting his access to other departments. If the firm could fully monitor the employee's allocation of time, it might well be desirable to allow him some time to network, since the value of this time to the employee could exceed the value to the firm of having him spend it on the company task. What happens if networking, and other such activities, cannot be monitored?

Holmstrom and Milgrom show:

Result 1a: *As measuring performance on the company task becomes easier, pay should become more sensitive to performance, and more personal business activities should be permitted.*

Result 1b: *If measuring performance is sufficiently difficult, some personal business activities which are socially valuable should be prohibited.*

These results can be explained and interpreted as follows. There are two ways for the firm to limit the extent to which the employee diverts attention from the company task to personal business activities: prohibit the latter through rigid rules, and tie his pay closely to performance on the company task. Since the employee is risk averse, the cost of the second method is larger the greater the error in measuring performance, and therefore larger measurement error tilts the firm's preference towards rigid rules and away from pay-for-performance. Sufficiently large measurement error can make the risk-bearing cost of using performance pay so great that it is better to control the employee by prohibiting some personal business activities which are socially valuable. Thus bureaucratic rigidity may be an optimal strategy for job design when measuring performance is very difficult. Improved measurement of performance should lead simultaneously to an increased emphasis on results and a reduced emphasis on procedures. In other words, employees' financial responsibility and their authority over own their own activities should move together.

This analysis may help explain the recent trends (reported by Piore (1989), Kanter (1989), and others) towards more flexible job definitions for middle managers and towards more performance-related pay. These changes have occurred along with two significant improvements in performance measurement: First, the conceptual development of effective measures of non-financial aspects of performance such as quality, customer satisfaction, and market share, and second, the enhanced ability to produce and use these, as well as financial measures, through improved information technology (Eccles (1991)).

2b. Career Design: Fast Tracks vs. Uniform Treatment

Another issue in the design of an organization's microstructure is the design of

careers—selecting the procedures that determine how employees move from job to job within the organization and when they leave.

I will focus on one specific aspect of career design: the choice between what can be termed "fast track policies" and "uniform early treatment". By fast track policies, I mean policies which designate individuals as "stars" at an early stage and provide extra opportunities (e.g. training or job assignments) that enhance the likelihood that they will advance faster than their peers. There is both descriptive and statistical evidence for the use of fast track policies in American firms. For example, in a detailed study of a large corporate hierarchy, Rosenbaum (1984) found that even when controls for ability were included, later promotions and earnings were correlated with early promotions and earnings. Uniform early treatment describes a form of career development much more common in Japan: in large Japanese firms, workers are typically not differentiated from their cohort until 12-15 years after they join—during this period they are promoted solely according to seniority (the "escalator system"); after this point, however, there is considerable differentiation according to ability (Aoki, 1988; Prendergast, 1991). Recently, though, Japanese practices have begun to change. Especially in the staffing of new ventures, established companies have been giving able young employees opportunties to move into management positions much faster than in the past.

What determines how early in employees' careers the firm takes actions that have long-lasting effects on their earnings and promotion opportunities? I will discuss both the advantages and the disadvantages of fast track policies and attempt to assess in which organizational environments they can be most useful.

I have developed a model (Meyer (1991a)) which shows that fast tracks, implemented through a policy of biasing promotion contests in favor of those who have performed well early on, can increase the value of the information firms acquire about their employees' abilities.

Suppose the firm hires two employees, i and j, who are ex ante identical but whose exact abilities are initially unknown to everyone. The firm has two periods in which to observe the employees' performance, after which it tries to promote the more able of the two. In each of the two periods, the difference in the employees'

outputs, Δx^t, equals the difference in their abilities, $\Delta\eta$, plus an exogenous shock, ϵ^t. Suppose that not only are output differences a noisy signal of the ability difference, but also the firm has access only to rank-order information on outputs. However, the firm can costlessly choose, in each period, the criterion according to which the outputs are ranked, that is, it can choose the extent to which each rank-order contest is biased. Formally, it chooses a number c^t in each period, and it learns whether or not j's output exceeds i's by at least c^t, i.e. whether

$$\Delta x^t < -c^t \qquad \text{or} \qquad \Delta x^t \geq -c^t$$

In practice, contests can be biased by treating employees asymmetrically, for example, giving one of them more training or equipment, or more access to senior personnel. Implementing bias in this way (by differentiating the inputs into employees' production) is feasible even if no individual in the firm receives information on outputs that is finer than rank-order information.

To complete the formal assumptions, assume that the noise terms ϵ^t are distributed independently and symmetrically about 0—in the absence of explicit bias, the employees' jobs are identical. Also, assume that a larger value of Δx^t signals a larger value of $\Delta\eta$, so is better news about i's relative ability—this is a monotone likelihood ratio condition.

We are now in a position to see why the firm gains by biasing the later contest according to the outcome of the early one. For simplicity, suppose the two contests are intrinsically equally informative—ϵ^1 and ϵ^2 are identicallly distributed– and suppose the first contest is unbiased. *If no bias is used in the second contest, the second contest is of no value to the firm in its promotion decision.* For if i wins both contests, the firm will promote i, and if i wins the first and loses the second, the firm is also willing to promote i. Thus, promoting the winner of the first contest, completely ignoring the outcome of the second contest, is an optimal strategy. A policy of biasing the second contest in favor of the first-period loser has the same problem: it is still optimal to ignore the second contest. For if i wins the unbiased first contest and j wins the second with bias in his favor, i's overall record is stronger, because his victory was achieved without the help of bias. Only if the second contest is biased in favor of the first-period winner can the second contest have positive value for the firm: in this

example, if i wins the first contest and is given any positive bias in the second, then a win by j in the second is a strong enough signal to make the firm want to promote j. Therefore, the optimal use of bias in the second contest is to favor the first-period winner.

This result is more general than this simple example: it holds whether or not the contests are intrinsically equally informative and whatever the bias in the first period:

Result 2a: *The quality of the promotion decision that follows the second contest is maximized by setting the bias in the second contest to favor the first-period winner.*

The rule for the optimal level of second-period bias has a simple form: The bias should be set so the firm's promotion decision, as a function of the difference in second-period outputs, is exactly the same as if the firm could observe this difference precisely.

The implication of Result 2a is that the optimal use of bias reinforces the likely edge in ability of the employee who wins the first contest and so enhances the correlation between success at the first stage and success at the second.

The model also yields:

Result 2b: *The quality of the promotion decision that follows the second contest may be maximized by biasing the first contest, even though the employees are ex ante identical.*

Thus, it *may* be optimal to arbitrarily single out one of the employees for special treatment—even before he has done anything to deserve it. The reason why first-period bias may be beneficial is that makes the informativeness of the first contest variable—if the favored employee wins, the firm learns little, but if the disadvantaged employee wins, this victory is a strong signal about his ability. Since the firm can adjust the second-period bias according to how much it has learnt from the first contest, the increase in the variability of first-period informativeness is, per se, valuable. Whether a non-zero level of bias in the first contest is actually optimal depends on whether the benefit from the greater variability of first-period informativeness outweighs the cost from the lower average level of first-period informativeness.

A second explanation for fast tracks, which reinforces that just given, has been provided by Bernhardt (1991). The argument builds on the premise that current employers learn more about their employees' abilities than do potential employers. Hence promoting an employee provides a signal to the labor market that he is able, so it leads the market to raise the wage it is willing to offer him. Promotion policies may thus be influenced by the desire to minimize wage raises as well as to maximize output. It follows that in choosing between two employees for a second promotion, it *may* be cheaper to award it to the one who received his first promotion earlier, and whose current wage is therefore higher, even if this individual is now known by the firm to be slightly less able. Hence the firm's desire to exploit its private information about employees' abilities can lead it to give unduly heavy weight to early performance in determining career profiles.

A third explanation for fast tracks, suggested by Prendergast (1991), is based on the idea that ability, training, and job level are complementary inputs—the marginal return to an increase in any one of these dimensions is higher, the higher the levels of the others. Prendergast supposes that firms initially acquire better information than workers about their abilities and that firms must induce workers to invest in acquiring firm-specific skills, by convincing them that their promotion prospects will be improved if they do train. Given the complementarity assumed, giving a worker an early promotion is a credible signal to him that he has high ability and therefore that the likelihood of further advancement if he trains is high. As a result, those promoted early *will* train and will consequently move ahead faster than their peers. Prendergast shows that the signaling benefits of early promotion in inducing training may be large enough to lead the firm to promote able workers even before they are qualified. That is, the firm may make distinctions in its treatment of workers even before such distinctions are productively efficient.

A potentially important disadvantage of fast tracks is that they demoralize slow

starters. This effect is suppressed in the models of Meyer and Bernhardt, in which employees make no decisions about effort or training. The demoralization is captured in Prendergast's analysis, since those who do not receive early promotions realize that it is not worth their while investing in training. A complete analysis of the costs and benefits of fast tracks for incentives must, however, consider not only the incentives of employees *after* selections for the fast track have been made, but also their incentives before this point. If selections for the fast track are based on early performance, then the prospect of joining a fast track enhances early incentives for everyone. If the efforts so induced benefit the firm, then this effect of fast tracks opposes the later demoralization of slow starters.

I have examined the tradeoff between these early and later incentive effects in a second model of the design of a sequence of (possibly) biased contests (Meyer, 1992). I have shown that if risk-averse employees are known to be identical, so the firm's only concern is to provide incentives at the lowest possible cost, then it is again optimal for the firm to use a fast track, that is, to commit to biasing the later contest in favor of the early winner. A small second-period bias in favor of the first-period winner produces only a second-order reduction in second-period incentives, while it generates a first-order increase in first-period incentives. Thus, in the absence of inherent differences in ability among employees, the demoralization that follows the selection for the fast track is outweighed by the enhanced motivation to qualify for the fast track.

The analysis in Meyer (1992) suggests that the overall incentive effects of fast tracks are most likely to be negative when early performance is very informative about ability. Yet Meyer (1991a) shows that it is in such environments that fast tracks are valuable in improving the quality of job assignments. Hence in these environments, the choice of whether to use fast tracks entails trading off concerns for incentives against concerns for efficient job assignments.

When such a tradeoff arises, the choice between fast tracks and uniform early treatment may depend on the extent of delegation within the firm. The more delegation of decision-making authority to lower levels, the more important it is to maintain the motivation of employees at these levels, and therefore the less attractive are fast tracks. On the other hand, the more centralized is authority, the faster the returns to ability and training rise with job level, and hence the greater the importance of identifying the most able, ensuring that they train, and placing them in the most responsible jobs. In highly centralized organizations, then, fast tracks should be relatively more attractive.

3. Macrostructure: Internal Structure and Boundaries of Firms

This brings us naturally to the question of what determines whether activities within the firm are coordinated in a centralized or a decentralized fashion. In the final part of my talk, I want to consider this question, and a closely related one: What determines whether the assets used to perform a given set of activities are part of one firm or another or constitute an independent enterprise?

I will argue that under specified, idealized conditions on contracting, the answers to both questions are, at a formal level, trivial. Under these conditions, centralization is at least as good as any arrangement involving delegation And having all activities performed within an integrated firm is at least as good as any division of the activities among separate firms. However, I will show that attention to limitations on contracting allows the development of meaningful answers to both questions, answers that can help explain observed variations in internal structure and in the boundaries of firms.

3a. Internal Structure: Centralization vs. Delegation

Piore (1989), Kanter (1989), and the MIT Commission on Industrial Productivity (1989) report several recent trends in firms' internal structure, each of which

represents a move towards greater use of delegation.

First, there has been an expansion in the number of independent profit centers within firms. When a unit is designated as a profit center, it is given a profit target and its managers' compensation is affected by how far actual profits deviate from the target; the managers are given flexibility to decide what inputs to procure and how to organize production. Firms are designating not just operating units as profit centers; they are also organizing staff support services, such as training and business research departments, as profit centers.

Second, there has been an increased use of cross-functional or cross-departmental teams. For example, it is increasingly common for new products to be developed by teams of engineers, including product, process, and industrial engineers, all working simultaneously. These cross-boundary teams tend to be coordinated laterally, by the team members themselves, rather than by members' superiors.

Third, Piore reports a new, more decentralized strategy for managing innovation. Instead of acquiring and fully integrating an outside company producing a new product, firms are moving toward "arm's length financial participation which establishes a collaborative relationship without abrogating the separate identity and personality of the outside organization" (Piore (1989), p. 8).

These trends towards delegation do, however, represent something of a puzzle for economic theorists. Suppose that employees have private information about the tasks they are to perform, and that top management can solicit reports from employees about their private information. If we assume that 1) there are no limitations on communication, and 2) top management can commit in advance as to how it will respond to every conceivable set of reports by employees, then we have the following implication, known as the Revelation Principle and due in this form to Myerson (1982):

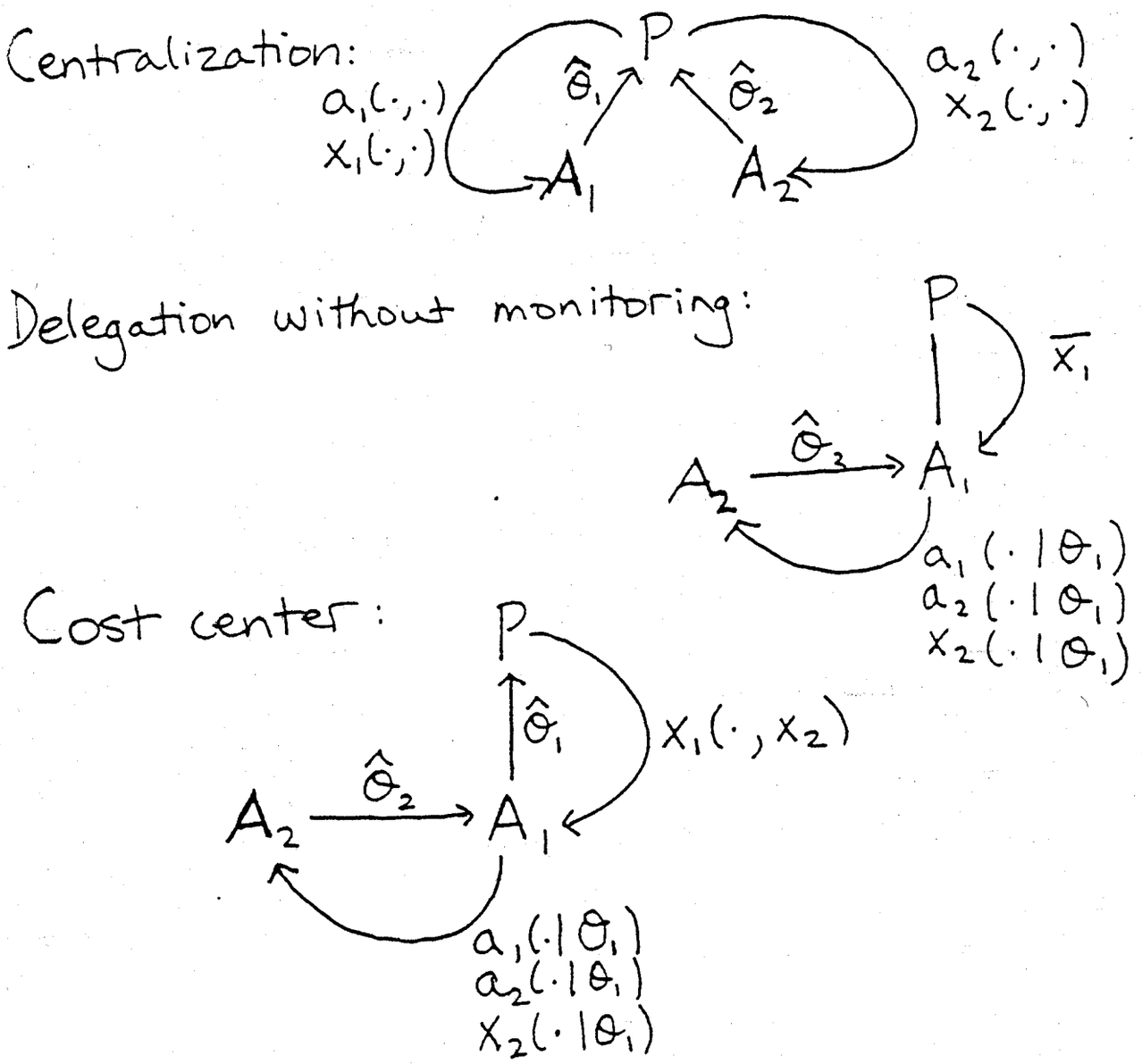
Revelation Principle: Any non-cooperative equilibrium outcome of an arbitrary organizational structure can be replicated by a centralized two-tier structure, in which each employee is given incentives to communicate truthfully all of his private information directly to top management.

For delegation of decision-making to be valuable, then, it must be that either communication is costly or that management's ability to commit itself is in some way limited. I will pursue the implications of incomplete commitment power in discussing the boundaries of firms. Here, I will discuss how Melumad, Mookherjee, and Reichelstein (1990, 1991) have modeled limits on communication and the implications of their analysis for when delegation can be useful.

Melumad, Mookherjee, and Reichelstein (1990, 1991) analyze how top management (referred to as the principal, P) will choose to coordinate the production activities of two employees (referred to as agents 1 and 2, A_1 and A_2). All individuals are risk neutral. Agent i produces output a_i at cost $a_i\theta_i$, and the payment he receives is denoted x_i. The marginal cost parameter θ_i is privately observed by agent i before contracting; the total cost incurred, $a_i\theta_i$, is also private information to agent i. The random variables θ_1 and θ_2 are independently and continuously distributed. Finally, the principal has a benefit function $B(a_1, a_2)$, and he chooses a decision-making structure to minimize the total cost of producing a benefit level $\bar{B}$.

Three organizational structures are compared (see Figure 1). The first involves *centralization* of decision-making. The principal commits in advance to rules determining how much each agent will be asked to produce, and what payment he will receive, as a function of the reports that the agents simultaneously make to the principal about their costs. The second structure is *delegation without monitoring*. The principal makes a fixed payment to agent 1 and delegates to him both the allocation of production between himself and agent 2 and the determination of agent 2's pay-

Figure 1 : Three organizational structures



ment. Agent 1 makes these decisions on the basis of a report from agent 2. In the third structure, agent 1 is designated as a "cost center". He is given a cost target and the flexibility to decide how to allocate production to try to meet the target; his compensation depends on how far the actual payments he makes deviate from the target. (In this simple model, the only payment agent 1 makes is to agent 2, but other expenditures on inputs could easily be incorporated.) Formally, this third structure is *delegation with monitoring* of agent 1's costs; now the payment from the principal to agent 1 depends on the amount agent 1 pays to agent 2.

In the absence of limitations on communication, the Revelation Principle states that centralization must be at least as good as either form of delegation. Melumad et al. show:

Result 3a: *With unlimited communication, delegation to a cost center can replicate the outcome under centralization, but delegation without monitoring does strictly worse.*

This result is easily explained through an example. Suppose that the agents' outputs are perfect substitutes in the principal's benefit function, so

$$B(a_1, a_2) = a_1 + a_2,$$

and that their marginal cost parameters are independently and identically distributed. Under centralization, the principal orders the agent with the lower marginal cost to produce all of the required output. Under delegation without monitoring, agent 1 receives his fixed payment and chooses how to allocate production between himself and agent 2. Because agent 1 does not know agent 2's marginal cost, the monetary cost to agent 1 of procuring output from agent 2 includes not just the production cost but also the "informational rent" (or bribe) to agent 2, which is necessary to induce agent 2 to report his cost truthfully. The cost to agent 1 of producing the output

himself, however, is just the production cost. Consequently, agent 1 will bias the allocation of production towards himself, and this is inefficient from the principal's point of view. (For (θ_1, θ_2) in the shaded region in Figure 2, agent 1 produces all of the required output under delegation without monitoring, whereas under centralization, agent 2 would be the only one to produce.) The principal thus suffers a "control loss" from delegation without monitoring and incurs higher costs as a result. Under a cost center, the principal partially reimburses agent 1 on the margin for purchases from agent 2. This subsidy to agent 1 eliminates the bias in his decisions and results in the same production allocation, and therefore the same costs for the principal, as under centralization.

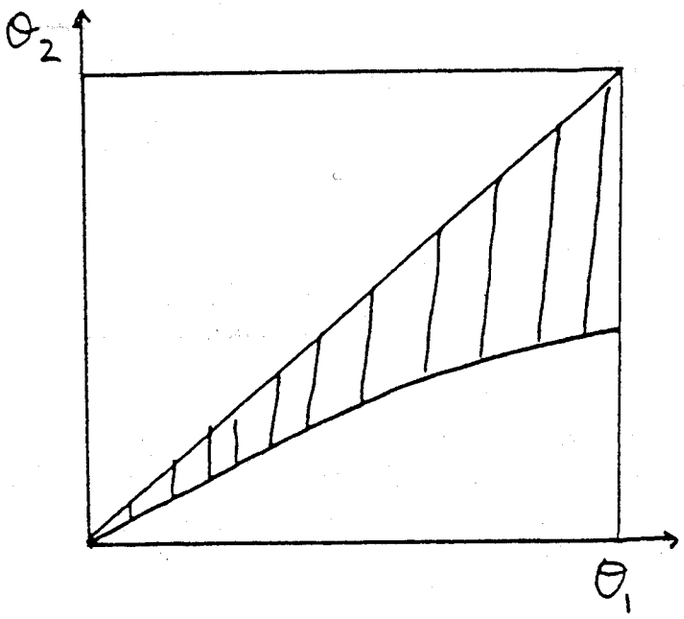
Thus a cost center can do as well as centralization when communication is unrestricted. If there are limits on communication, can it do better?

Melumad et al. model limitations on communication by assuming that agents can transmit only a coarse summary of their private information. Formally, it is assumed that agents can make only a finite number of possible reports, even though their marginal cost parameters θ_i are continuous random variables. In this setting, delegation brings a "flexibility gain", since agent 1 can base the allocation of production between himself and agent 2 on the *exact* value of θ_1, whereas the principal, under centralization, could use only a coarse summary of θ_1.

Result 3b: *With limited communication, delegation to a cost center performs strictly better than centralization, which in turn performs strictly better than delegation without monitoring.*

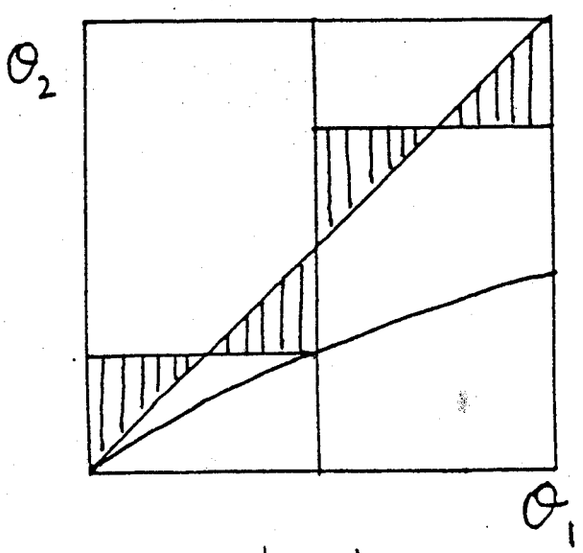
Returning to the earlier example, suppose each agent can make only two possible reports. Under centralization, the principal will prefer to have the agents report sequentially rather than simultaneously. Even so, because of the limitations on communication, the criterion for allocating production will differ from the 45 de-

Figure 2



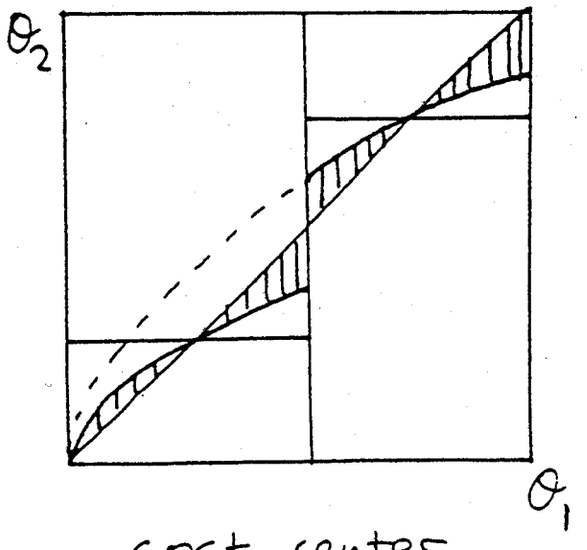
delegation without monitoring, under unlimited communication

Figure 3



centralization under limited communication

Figure 4



cost center under limited communication

gree line—there will be efficiency losses relative to the centralized mechanism with unlimited communication (see Figure 3). Under delegation without monitoring, the criterion for allocating production can depend precisely on θ_1, but agent 1, and not the principal, will capture this flexibility gain. Under delegation to a cost center, a subsidy to agent 1 can no longer completely eliminate the control loss (and hence the inefficiency in production), because the subsidy can depend only on a coarse summary of θ_1. However, for each range of θ_1, the principal can, through delegation to a cost center, induce agent 1 to use a criterion for allocating production that is closer to the 45 degree line, and hence more efficient, than the criterion that results under centralization (see Figure 4). Therefore a cost center is strictly preferable.

I have focused on delegation to a cost center for simplicity; precisely analogous results hold for profit centers. In this setting, a profit center involves delegating to agent 1 the further choice of the benefit level $\bar{B}$ itself and basing his compensation on the difference between $\bar{B}$ and the monetary costs he incurs.

The implication of this analysis is that delegation to a responsibility center can be more efficient than a centralized method of decision-making, when the responsibility center's financial performance can be monitored *and* when the relevant private information is difficult to fully transmit to top management.

This result may help to rationalize not just the general expansion in the use of profit centers but also the two specific instances of the increased use of delegation discussed earlier. As the activities to be coordinated come to involve employees from different deptartments or functions, the relevant private information may well become harder for an individual senior manager to fully understand. Consequently, coordination is more likely to be achieved through delegation, and horizontal channels of communication will partially replace vertical ones. Secondly, since the private information relevant to the production of new products is likely to be particularly complex, decentralization, provided it can be accompanied by close financial moni-

toring, is likely to be preferable for managing innovations.

3b. Boundaries of the Firm: Integration vs. Divestiture

I turn now to my final topic, the determinants of the boundaries of the firm.

The fundamental theoretical puzzle has been expressed powerfully by Williamson (1985), who asked: If organizing transactions within a firm allows parties to escape the transactions costs of using markets (as Coase argued), why isn't it always optimal to merge two independent firms and run the new entity using a policy of "selective intervention"—intervening when and only when intervention is beneficial? An even more basic argument is that if all transactions were governed by contracts that were complete, in the sense of never needing revision or augmentation, then it would not matter whether the parties to the contract were members of the same or different firms. Consequently, the boundaries of firms (both vertical and horizontal) would be irrelevant.

The most compelling resolution to these puzzles starts from the recognition that contracts are necessarily incomplete. This idea is the basis of several recent theories of the determinants of the boundaries of firms.

The first approach identifies ownership with possession of residual rights of control over physical assets—residual rights are those rights not explicitly included in contracts (Williamson (1979, 1985), Klein, Crawford, and Alchian (1978), Grossman and Hart (1986), Hart and Moore (1990)). The allocation of ownership affects the ex post bargaining positions of parties to contracts and hence the ex post distribution of surplus from the transactions. Through this route, ownership affects the parties' incentives to make investments ex ante in relationship-specific assets. This theory's prediction that high specificity of physical assets favors vertical integration has successfully rationalized electric utilities' ownership of coal mines and publishers' ownership of printing presses.

A second approach identifies ownership with the rights to return streams from assets (Holmstrom and Milgrom, 1991). Changes in the value of an asset used by an agent in a transaction, especially an intangible asset, may be very difficult to measure and hence difficult to contract upon. The only way to provide the agent with incentives to maintain the asset may be to transfer ownership of it to him, that is, to make him an independent contractor rather than an employee. This change, however, will divert the agent's efforts away from other activities, such as producing output, unless he is simultaneously provided with output incentives. This theory can thus explain the empirical finding that the easier it is to measure the sales performance of individual salespeople, the more likely they are to be independent agents rather than employees (Anderson and Schmittlein, 1984).

The explanatory successes of these two theories concern cross-sectional variation in asset ownership. Neither theory can easily explain time-series variation in the ownership of business units—for a start, the very large volume of transfers of control. It seems implausible that patterns of asset specificity or costs of measurement should shift so much as to generate the frequent divestitures we observe.

I will now sketch a third theory, based on consideration of "influence costs" (Milgrom and Roberts (1988, 1990)), which, I will argue, can explain several facts about divestitures.

This theory identifies as the distinguishing features of a firm the *authority* and *autonomy* of its top management—that is, their broad rights to intervene in a discretionary fashion in lower-level decisions and their relative immunity to intervention from outside. Top managers are vested with discretionary authority because it would be impossible for them to write contracts with all subunits that specified what actions would be taken in all conceivable situations. That is, contractual implementation of the optimal policy of selective intervention would simply not be feasible. Expansion of the activities carried out within a particular firm (as, for example, when a formerly

separate organization is merged into it) thus increases the range over which discretionary authority may be exercised. Increases in the range of discretionary authority have both benefits and costs. The benefits stem from the enlargement in the set of possible interventions—for example, personnel transfers and capital reallocations. The costs arise because such interventions can have redistributive consequences, transferring rents and quasi-rents among employees. Those employees most affected by particular decisions are often those who possess the most relevant information. Consequently, employees will have incentives to engage in "influence activities"—attempts to distort the information they provide in order to affect the distributive results of management's discretionary decisions. The resulting "influence costs" borne by the firm include the opportunity costs of the resources expended on influence, the efficiency costs of distortions in decisions, and the degradation in performance when organizational policies or structure are altered to limit opportunities for influence.

This is the broad outline of influence cost theory. What observations about divestitures call for explanation?

The first significant feature of divestitures is their volume: In recent years, 40-45 precent of merger and acquisitions transactions in the U.S. have involved divestitures—the sale of individual product lines, divisions, or subsidiaries to new owners (as opposed to the sale of entire firms). Divestitures include sales to other firms, management buyouts, sales to the public through stock offerings, and spinoffs. The value of divestitures reached over 70 billion dollars in the US in 1989.

The second observation concerns the nature of the divested units: There are two issues here—the growth prospects of the units that are divested and the relationship of these units to the other activities of the firms that sell them. Both issues are addressed by the quotation, which seems to summarize conventional wisdom: "Companies have been divesting slow growing, cash-draining, and other non-core businesses" (*Mergerstat Review 1990*). Though I am not aware of direct, systematic evidence on the

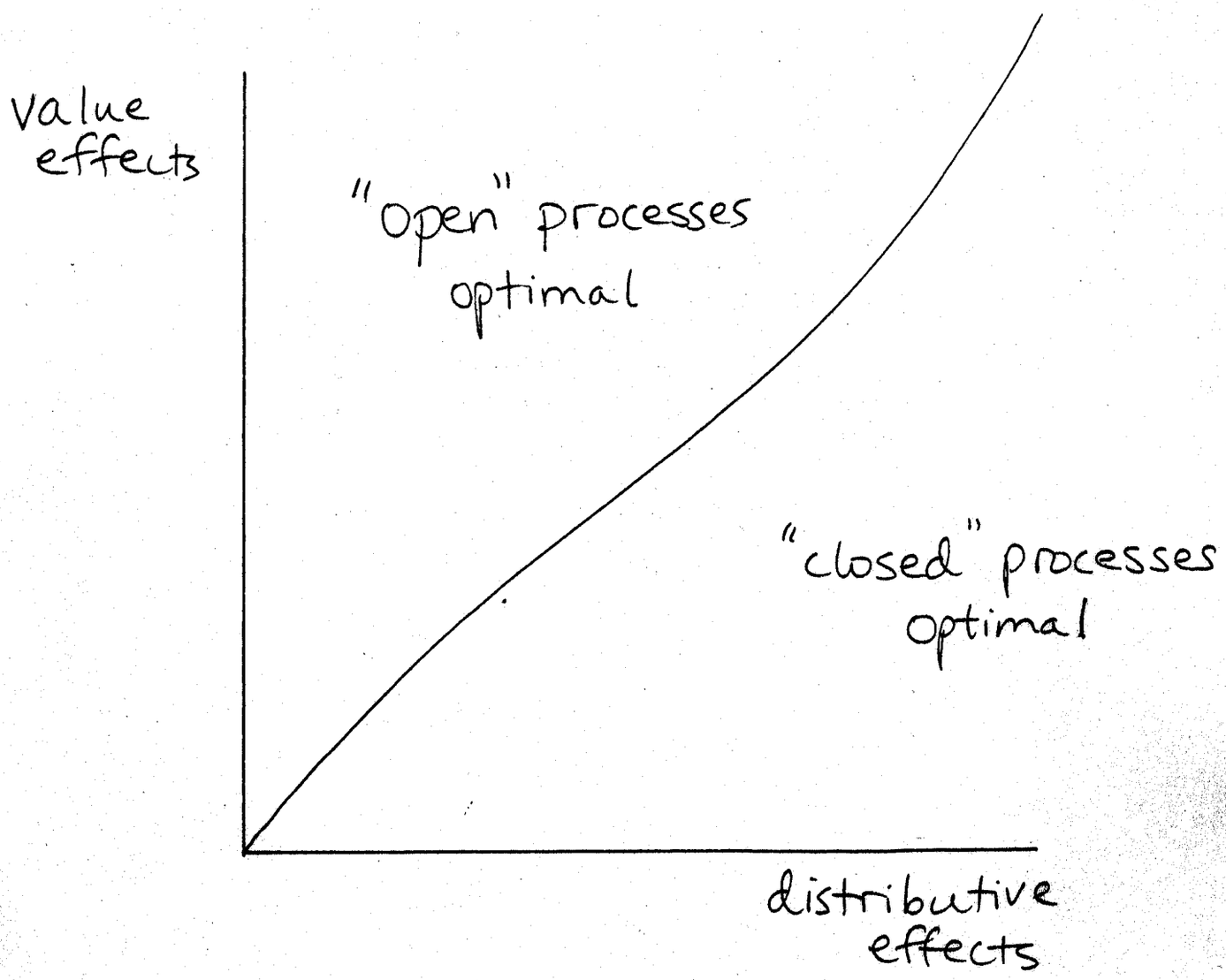
growth prospects of divested units, there is indirect evidence that weak units are over-represented in divestitures from Kaplan and Weisbach's (1990) study of divestitures following acquisitions. Furthermore, examples of decline triggering divestiture are easy to find. There is, however, something of a puzzle in the idea that units with poor prospects are especially likely to be divested. Why should it be easier in the case of weak units to find buyers willing to pay more for them than they are worth to the current owners? A satisfactory theory must answer this question. Finally, with respect to the relationship between the divested units and their parent firms, work by Markides (1990) and Lichtenberg (1990) has documented the trend in the 1980's toward "corporate refocusing".

The third aspect of divestitures is the buyers of the divested units: The evidence for the last decade suggests that when divested units are acquired by other firms, the purchasers tend to be already in the divested unit's line(s) of business (Bhagat, Shleifer, and Vishny, 1990; Kaplan and Weisbach, 1990). (This pattern is consistent with the trend toward refocusing.)

The theory that Milgrom, Roberts, and I (Meyer, Milgrom, and Roberts (1992)) have developed to explain these features of divestitures can be outlined using Figure 5. The horizontal coordinate measures the distributive effects on employees of decisions by top management and the vertical coordinate the effects of top management's decisions on the value of the firm. Under "open" decision processes, management solicits information from employees and retains discretion as to how to use what it learns. Under "closed" processes, no information is solicited, and discretion is replaced by rigid rules. Open processes incorporate information better, but they are more subject to organizational politics. Closed policies will be preferable when the value of information in decision-making is low and the distributive potential of the decisions, and hence the incentives for influence activities, are large.

The key aspect of our theory is that there is a basic asymmetry between the effects

Figure 5


value effects
"open" processes optimal
"closed" processes optimal
distributive effects

of growth and decline on the distributive consequences of decisions about employment. In a unit where prospects are relatively bad and there is a positive probability of layoffs, any job-related quasi-rents received by the employees are at risk. They will have incentives to manipulate their reports to try to save their jobs, by trying to convince top management to transfer extra resources to their unit from elsewhere in the firm. In contrast, in a unit which faces no threat of layoffs, the job-protection motive for influence activities would be absent.

One way to avoid the influence costs in a declining unit is to isolate the declining unit so that it can no longer attempt to claim corporate resources. But accomplishing isolation while maintaining the unit within the parent firm may be very difficult—it will be difficult for top management to prevent the unit's managers from using existing communication channels to convey information about the unit's prospects and difficult for top management to commit not to use any information that does get through to them. In effect, then, isolation necessitates separation of the unit from the parent firm, for example, through a spinoff, a sale to the public, or a management buyout.

But what can separation achieve? We argue that separation can effect the transition from open to closed decision processes, and hence reduce the incentives for influence activities, because the capital market is likely to be significantly less responsive to influence attempts at the margin than is corporate management. There are fewer opportunities to influence the capital market than to influence corporate management. Furthermore, as analyses of financial contracting under asymmetric information have shown, an independent firm's ability to tap outside capital markets may be very limited.

An alternative strategy for achieving isolation of a declining unit is to sell the unit to a leveraged-buyout (LBO) firm. As Jensen (1989) has emphasized, LBO associations often have contractual mechanisms in place that deny the LBO firm

itself (the top management) the right to transfer resources among the LBO'd units.

The conclusion of this line of argument is that separation of a declining unit can be an optimal response by the parent firm, because the gains from reducing influence costs can exceed the losses from eliminating decision options. Separation of a growing unit, on the other hand, would not bring comparable gains.

Influence costs in the declining unit could also be limited by reducing the danger of layoffs or the degree of competition for resources among units of the firm. These reductions could be accomplished by selling the unit to a firm which, as the unit shrinks, could absorb the affected employees into its own operations or to a firm whose units were more focused around a common business and relied more on one another for services.

Taken together, then, these arguments offer an explanation for why declining units and units in non-core businesses should be over-represented in divestitures and why, when the units are acquired by other firms, the buyers should tend to be already operating in the business of the divested unit.

To test these arguments in a more formal way, we have developed a model of the performance of a unit when integrated within a larger firm and managed according to an "open" decision process. We analyze a game between the top management of the parent firm (referred to as "the firm") and the management team of the unit (referred to as "the managers"). (In order to highlight the consequences of the conflict of interest between the top management and the division's managers, we suppress any conflict of interest among the division's managers themselves, assuming that they act as a team.) In each period, the firm adjusts employment in the unit, on the basis of an informative report by the managers about the shock to the unit's profitability. In each period, the currently employed managers manipulate their report, at a cost to themselves and to the firm, to protect the quasi-rents earned in their jobs. Managers choose their level of influence before learning the realization of the profitability shock.

Their quasi-rents stem from search or relocation costs which the managers would incur if they were laid off. Importantly, reductions in employment expose the managers to the risk of losing their quasi-rents, whereas increases in employment leave their quasi-rents unchanged. The analysis focuses on how the equilibrium level of influence costs varies with the unit's prospects, which are parameterized by the distribution of the profitability shocks.

In this model, since the only possible benefit of influence activities for the managers is to reduce the risk of layoff, it is immediate that if, in equilibrium, there are never any layoffs, then the level of influence is zero. But even when such an equilibrium exists, there may be others as well. The firm may expect high levels of influence, leading it to reduce its estimate of the value of maintaining the unit's size, thereby encouraging the very influence activities that it fears. Thus a favorable business environment is not by itself a guarantee of low influence costs.

To perform comparative statics, we focus on the Pareto-dominant equilibrium. We investigate whether, as the unit's prospects worsen, the equilibrium level of influence increases monotonically. The asymmetric effects of growth and decline on managers' quasi-rents create a force towards monotonicity: the higher the probability of employment contracting, the higher the probability that the marginal return to influence is positive rather than zero. We find, though, that there are two possible sources of non-monotonicity. One has already been suppressed by constructing the model so that a manager's quasi-rents, contingent on remaining employed, are independent of the rate of growth or decline of the unit. If, instead, more rapid decline reduced the value of remaining employed, this effect per se would reduce the incentives for influence. A second potential source of non-monotonicity is technological: with more rapid decline, the firm's employment decision could become less sensitive to the variable about which the managers provide information, thereby reducing the marginal returns to managers from influence activities.

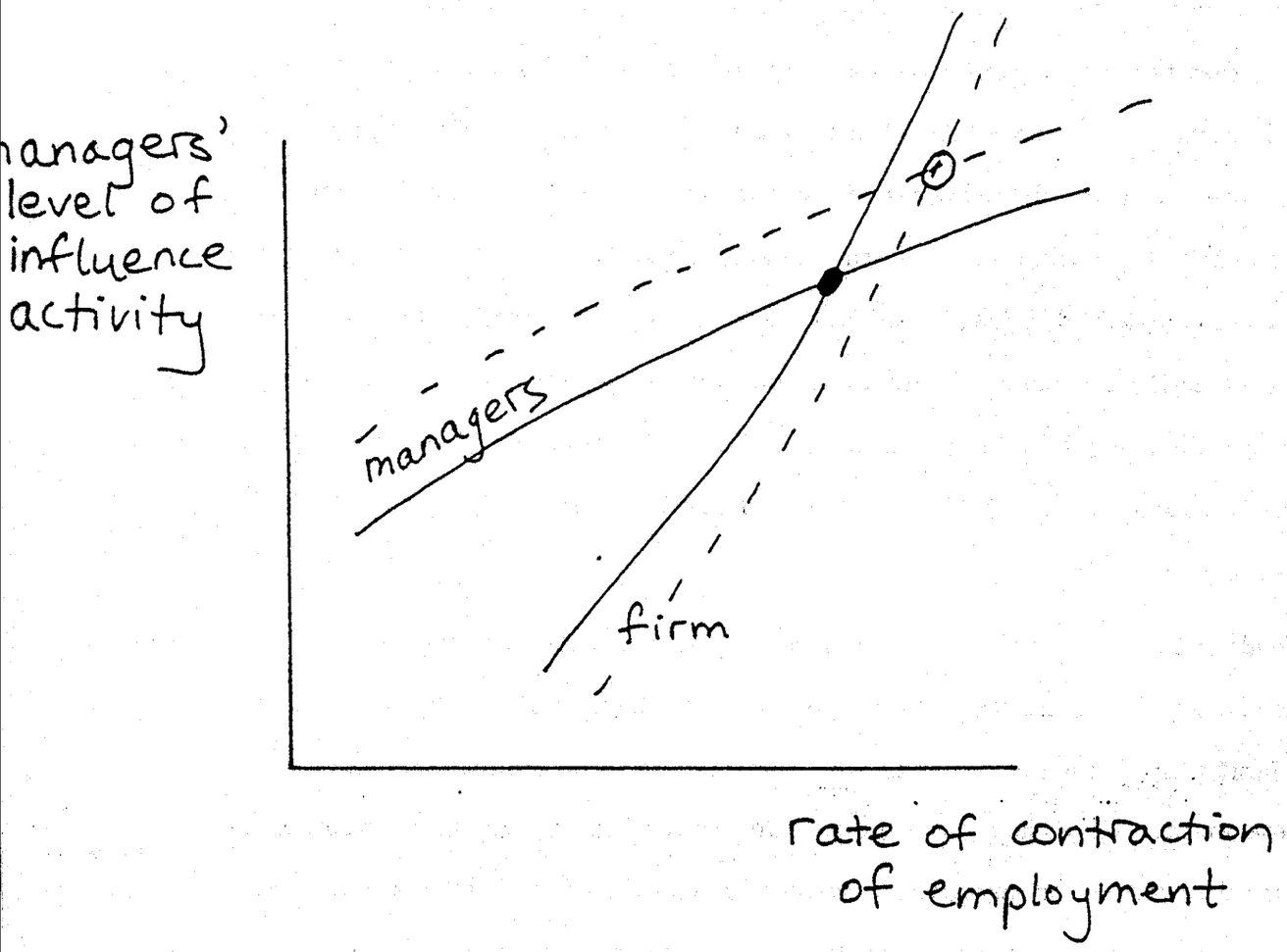
We show that these are, in fact, the only two effects which can interfere with the monotonicity of influence costs as the unit's prospects worsen. Specifically, if these two effects are neutral, then the level of influence and the expected rate of contraction of employment are strategic complements, and the Pareto-dominant equilibrium can be represented as the intersection of the upward-sloping "reaction functions" in Figure 6. The higher the rate of contraction anticipated by the managers, the higher the level of influence they choose. And the higher the level of influence anticipated by the firm, the faster the firm chooses to contract employment. As the distribution of the profitability shock worsens, both reaction functions shift outwards—for any given rate of contraction, influence increases, and for any given level of influence, the rate of contraction increases. Therefore as the unit's prospects worsen, it contracts faster and influence costs rise in equilibrium. The negative effects of decline on the unit's performance are thus exacerbated by the firm's inability to prevent influence activities.

We also analyze how the equilibrium level of influence costs varies with the degree of focus of the parent firm. We parameterize focus by the probability that a manager who is laid off from the unit has skills which enable him to be productively employed elsewhere in the parent firm. An increase in focus shifts both reaction functions in Figure 6 inwards—for any given rate of contraction, influence is reduced, since managers' probability of remaining employed is less sensitive to the fate of their unit, and for any given level of influence, the rate of contraction falls, because the wage that managers demand is lower. In consequence, in the Pareto-dominant equilibrium, greater focus implies a lower level of influence costs. Thus, any direct, productive benefits of focus are supplemented by the indirect benefit of reduced influence activity.

Finally, we model the performance of the unit following separation from the parent firm. To capture the idea that separation closes off decision options, we assume that employment is no longer adjusted according to the managers' report. Hence,

Figure 6

"Reaction functions" in game between firm and managers of unit



——— before
- - - - after
} the distribution of the profitability shock worsens

employment is less sensitive to the profitability shock, but the managers' incentives for influence are eliminated. We show that separation is never optimal in the absence of a risk of layoffs but may become optimal as prospects worsen: the value lost from closing off the option of fine-tuning employment can be outweighed by the gain from reduced influence costs.

The arguments and the model I've sketched can thus help to account for the features of divestitures I listed earlier. The influence cost approach also has a number of additional implications for the boundaries of firms. To mention just one, firms do sometimes divest especially fast-growing, profitable units. Such units generate lots of rents, for example in the form of superior promotion opportunities. These rents would be the object of influence activities arising in other units of the firm, whose employees would like to have these opportunities for themselves. The parent firm in such a situation may thus prefer to curtail these influence costs by spinning off the fast-growing unit. Turning this argument around, it may help account for the frequent failure of acquisitions of small, innovative, fast-growing firms by larger, more traditional ones.

4. Conclusion

I have considered briefly four aspects of organizational design, sketching what light recent theoretical developments can shed on some trends and debates in current management practice. Within each of the four categories I identified, there are of course other questions of interest. Under the heading of job design, there is the classic question of the division of labor among employees at the same hierarchical level, an issue which, until recently, Japanese and American firms approached very differently. Other issues in career design include the extent to which management positions should be filled from within the firm and whether or not to adopt a policy of lifetime employment. A recent trend in the internal structure of American firms,

which has accompanied the move toward greater delegation, is the reduction in the number of levels of middle management. Finally, a further challenge to theories of the determinants of firm boundaries is to understand the causes and consequences of recent moves by American firms to 1) include key suppliers on product development teams; 2) create competition between internal units and outside vendors; and 3) develop closer and longer-term relationships with customers.

My discussion has shown that these various aspects of organizational design cannot be fully understood in isolation from one another. For example, whether or not fast tracks are desirable depends on the nature of jobs at different levels and the skills they require, which in turn depends on the degree of delegation of decision-making. And, especially when influence costs are potentially large, the firm's internal structure—the nature of the interactions among its units—will affect the benefits and costs of keeping a given unit within the firm's boundaries.

Furthermore, choices about organizational structure are closely linked to other choices firms face—such as choices about technology, corporate governance, and strategy. For example, the recent shift in technology away from classic mass production towards a capacity to switch rapidly among many different product designs, and the accompanying reduction of in-process inventories, have numerous implications for the flexibility of jobs and the extent of delegation.

A promising approach to understanding these interlinkages is to identify sets of activities or practices which are complementary, in the sense that a high level on one dimension should be accompanied by high levels on the others. As Milgrom and Roberts (1990b) have observed, when there are complementarities and nonconvexities, there can be multiple patterns which are potential optima, and yet no variation in only a few dimensions at a time would be beneficial. This observation can help explain the insistence by many commentators (for example, the MIT Commission (1989)) that improvements in management systems cannot be accomplished through

piecemeal changes but instead require an integrated strategy. In turn, the importance of coordinating changes in numerous dimensions highlights a role for the central management of a firm which goes beyond the role of the principal in resolving each of the individual issues I have discussed.

References

Anderson, E. and Schmittlein, D., "Integration of the Sales Force: An Empirical Examination", *Rand Journal of Economics*, 1984.

Aoki, M. *Information, Incentives, and Bargaining in the Japanese Economy.* Cambridge: Cambridge University Press, 1988.

Bernhardt, D., "Strategic Promotion and Compensation", Working Paper, Queen's University, April, 1991. Published in *Review of Economic Studies*, April 1995.

Bhagat, S., Shleifer, A., and Vishny, R., "Hostile Takeovers in the 1980's: The Return to Corporate Specialization", *Brookings Papers on Economic Activity: Microeconomics*, 1990.

Dertouzos, M., Lester, R., Solow, R., and the MIT Commission on Industrial Productivity. *Made in America: Regaining the Productive Edge.* Cambridge: MIT Press, 1989.

Eccles, R. "The Performance Measurement Manifesto", *Harvard Business Review*, Jan.-Feb., 1991.

Grossman, S. and Hart, O., "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", *Journal of Political Economy*, 1986.

Hart, O. and Moore, J., "Property Rights and the Nature of the Firm", *Journal of Political Economy*, 1990.

Holmstrom, B. and Milgrom, P. "Multi-Task Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization,* Special Issue, 1991.

Itoh, H., "Incentives to Help in Multi-Agent Situations", *Econometrica*, May, 1991.

Itoh, H. "Cooperation in Hierarchical Organizations: An Incentive Perspective", Working Paper, Faculty of Economics, Kyoto University, March, 1991. Published in *Journal of Law, Economics and Organization*, April 1992.

Jensen, M., "The Eclipse of the Public Corporation", *Harvard Business Review*, 1989.

Kanter, R. M. *Men and Women of the Corporation.* New York: Basic Books, 1977.

Kanter, R. M., "The New Managerial Work", *Harvard Business Review*, Nov.-Dec., 1989.

Kaplan, S. and Weisbach, M., "The Success of Acquisitions: Evidence from Divestitures", NBER Working Paper 3484, 1990. Published in *Journal of Finance*, March 1992.

Klein, B., Crawford, R., and Alchian, A., "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", *Journal of Law and Economics*, 1978.

Lichtenberg, F., "Industrial De-Diversification and Its Consequences for Productivity", NBER Working Paper 3231, 1990. Published in *Journal of Economic Behavior and Organization*, August 1992.

McAfee, P. and McMillan, J., "Organizational Diseconomies of Scale", Working Paper, U.C. San Diego, Feb., 1992. Published in *Journal of Economics and Management Strategy*, Fall 1995.

Markides, C., "Corporate Refocusing and Economic Performance, 1981-87", unpublished doctoral dissertation, Havard Business School 1990. Published as *Diversification, Refocusing, and Economic Performance.* Cambridge and London: MIT Press, 1995.

Melumad, N., Mookherhee, D., and Reichlestein, S., "Hierarchical Decentralization of Incentive Contracts", Working Paper, Stanford Graduate School of Business, Nov., 1990. Published in *Rand Journal of Economics*, Winter 1995.

Melumad, N., Mookherjee, D., and Reichlestein, S., "A Theory of Responsibility Centers", Working Paper, Stanford Graduate School of Business, July, 1991. Published in *Journal of Accounting and Economics*, Dec. 1992.

Mergerstat Review 1990, Schaumberg, IL: Merrill Lynch Business Brokerage and Valuation, 1991.

Meyer, M., "Learning from Coarse Information: Biased Contests and Career Profiles", *Review of Economic Studies*, Jan., 1991a.

Meyer, M., "The Dynamics of Learning with Team Production: Implications for Task Assignment", Technical Report No. 30, Stanford Institute for Theoretical Economics, Oct., 1991b. Published in *Quarterly Journal of Economics*, Nov. 1994.

Meyer, M., "Biased Contests and Moral Hazard: Implications for Career Profiles", *Annales d'Economie et de Statistique*, Special Issue on "Organizations and Games", March, 1992.

Meyer, M., Milgrom, P., and Roberts, J., "Organizational Prospects, Influence Costs, and Ownership Changes", *Journal of Economics and Management Strategy*, Spring, 1992.

Milgrom, P., "Employment Contracts, Influence Activities, and Efficient Organization Design", *Journal of Political Economy*, 1988.

Milgrom, P. and Roberts, J., "An Economic Approach to Influence Activities and Organizational Responses", *American Journal of Sociology*, Supplement, 1988.

Milgrom, P. and Roberts, J., "Bargaining Costs, Influence Costs, and the Organization of Economic Activity", in Alt, J. and Shepsle, K., eds., *Perspectives on Positive Political Economy.* Cambridge: Cambridge University Press, 1990a.

Milgrom, P. and Roberts, J. "The Economics of Modern Manufacturing; Technology, Strategy, and Organization", *American Economic Review*, 1990b.

Milgrom, P. and Roberts, J. *Economics, Organization and Management.* Englewood Cliffs, NJ: Prentice Hall, 1992.

Mirrlees, J. "The Optimal Structure of Incentives and Authority within an Organization" *Bell Journal of Economics*, 1976.

Myerson, R., "Optimal Coordination Mechanisms in Generalized Principal-Agent Relationships", *Journal of Mathematical Economics*, 1982.

Piore, M., "Corporate Reform in American Manufacturing and the Challenge to Economic Theory", Working Paper No. 533, MIT, 1989. Published in Allen, T. and Scott-Morton, M., eds., *Information Technology and the Corporation of the 1990s.* Oxford; New York; Toronto and Melbourne: Oxford University Press, 1994.

Prendergast, C., "Career Development and Specific Human Capital Collection", Working Paper, Graduate School of Business, Unviersity of Chicago, 1991. Published in *Journal of the Japanese and International Economics*, Sept. 1992.

Rosenbaum, J. *Career Mobility in a Corporate Hierarchy.* Orlando: Academic Press, 1984.

Williamson, O., "Transaction Cost Economics: The Governance of Contractual Relations", *Journal of Law and Economics*, 1979.

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