Journal of Economic Literature
Vol. XXX (September 1992), pp. 1382–1415

Hierarchy: The Economics of Managing

By Roy Radner
AT&T Bell Laboratories and New York University

This paper was prepared for the Marshall Lectures, Cambridge University, October 25–26, 1989. I am grateful to C. V. Kuh for helpful discussions and suggestions during the preparation of this paper, and to U. Pagano and two referees for comments on a previous draft. The views expressed here are those of the author, and not necessarily those of AT&T Bell Laboratories.

1. *Introduction*

One hundred years ago, at the time of the publication of Alfred Marshall's *Principles of Economics*, the typical British or U.S. firm was a small enterprise, managed by the owner, and perhaps a few assistants. There were, of course, larger enterprises, such as railroads, mines, and shipyards, but even in those firms relatively few persons specialized in the activity of managing. In the United States, according to Thomas K. McCraw (1984, p. 64),

> Although the profound economic movement that has become known as the "rise of big business" began with the railroads in the 1850s, it continued to move forward, in vastly expanded form, causing revolutions in manufacturing and distribution. These changes occurred between about 1880 and 1920. Prior to this period, no single manufacturing enterprise, indeed no entire manufacturing industry, had attained sufficient size to affect masses of people. Before the 1880s, even major factories customarily employed no more than a few hundred workers. . . . Yet, within a single generation after 1880, all this changed. By 1890, each of several large railroads employed more than 100,000 workers . . . and in 1901, the creation of the United States Steel Corporation climaxed a $1.4 billion transaction. This sum, which far exceeded the imaginations of most contemporary citizens, became a symbol of the new giantism in the American economy.

According to Joan Robinson, "Marshall had a picture, based on observation, of the family business in British manufacturing industry. . . . He observed that in many cases the fortunes of a business are bound up in the life of a family. An individual sets it going and it prospers, but by the third generation its vigor is lost" (Robinson 1977, p. 1324). On the other hand, Marshall was not unaware of "big business."

> And as with the growth of trees, so it was with the growth of businesses as a general rule before the development of vast joint stock-companies, which often stagnate, but do not readily die. . . . Nature still presses on the private business by limiting the length of life of its original founders, and by limiting even more narrowly that part of their lives in which their faculties retain full vigor. And so, after a while, the guidance of the business falls into the hands of people with less energy and less creative genius, if not with less active interest in its prosperity. If it is turned into a joint stock-company, it may retain the advantages of division of labor, of

specialized skill and machinery: it may even increase them by a further increase in its capital; and under favorable conditions it may secure a permanent and prominent place in the work of production. But it is likely to have lost so much of its elasticity and progressive force, that the advantages are no longer exclusively on its side in its competition with younger and smaller rivals. (Marshall 1920, p. 316)

The economies of today's industrialized nations are dominated by giant firms, each with thousands or even hundreds of thousands of employees. For example, the largest private firm in the world, General Motors, has more than 700,000 employees. In such firms, more than a third of the employees may be working full time in activities that are part of—or support—the management process. Thus, quantitatively as well as functionally, "managing" has become a significant activity in our economy.

This phenomenon has not escaped the attention of our colleagues in schools of business and management, where many courses are devoted to the subject. The pure science of economics, however, has been slower to focus on this phenomenon, and pure theory even slower. The picture of the firm in most economic textbooks, and in much current economic research, is still that of a unitary "entrepreneur," bent on maximizing profits. But there is also a growing body of research that views the modern firm as an *organization* of economic agents. There are two aspects of this relatively "new look" that I shall discuss here: (1) the sense in which *the large business enterprise is a small economy* (and sometimes not so small), and (2) the *central role of managing* in that economy. More particularly, I have set myself three tasks:

1. to convince the reader that managing is an activity worthy of economic analysis;
2. to review some insights that economic theory has provided into the economics of managing, and thereby also to persuade you that this activity is amenable to economic analysis; and
3. to sketch a number of theoretical problems waiting to be solved.

There is also a subtheme in my paper, namely, the significance of hierarchy. Large firms are widely perceived to be organized hierarchically, whatever that means precisely. Indeed, a well-known book opposes "markets" and "hierarchies" as the predominant forms of modern economic organization, the latter, of course, referring to the organization of firms (Oliver Williamson 1975). Students of management are well aware that many interactions in a typical firm are not organized hierarchically, even if the formal organization chart looks that way. Nevertheless, it is an important principle of organization, both in its prevalence and in the prevalence of attempts to circumvent it. So in addition to the three tasks I have just described, I shall focus on the question: *What is the economic significance of hierarchy in the organization and management of large firms?*

It will be seen, then, that my topic falls under the heading, the theory of the firm, but falls far short of encompassing the whole of that subject. In particular, I shall have nothing to say about why firms are owned by stockholders rather than workers or customers or the state, or how stockholders' voting rights should be exercised, or when one should expect vertical integration to take place. I shall be pleased if what I have to say has some interesting implications for these issues, but I leave it to other occasions, or even other investigators, to make the inferences. I suspect that, when it comes to the economics of managing large firms, capitalist, socialist, cooperative, and labor-managed firms have much in common, and it is that common element I hope to stress here.

Here, then, is a brief outline of my paper. To begin, I shall show you a few statistics to try to persuade you that managing has become a significant economic activity. Given the size of modern firms, and given the bounds on individuals' capabilities for information processing and decision making—"bounded rationality," if you like—it is obvious that the labor of managing must be divided among many persons in the firm. Although not inevitable, it is not surprising that this division of labor is accompanied by specialization, as with other kinds of labor. I shall use the general term *decentralization* to describe the division of labor of managing among several persons in a firm. As we shall see, this decentralization takes many specific forms.

First, *the processing of information must be decentralized,* i.e., divided among many persons. I shall use some ideas from computer science to argue that many information-processing activities in a firm can be accomplished efficiently by a hierarchical structure. Of course, in order to define "efficiency" I shall have to be precise about what is being economized.

Second, the "management sector" of the firm makes vast numbers of decisions, and bases these decisions on vast numbers of observations, or information variables. It is not economical for all decisions to be based on all the information available to the firm; that is to say, different decisions will typically be based on different sets of information variables. Put another way, different decision makers in the firm will typically have different information. I shall call this the *decentralization of information.* In principle, one could have decentralized information processing without the decentralization of information, but in large firms we see both. The efficient decentralization of information is the subject of the *theory of teams.*

Third, all the members of a firm will not have exactly the same goal. Indeed, even if no one were greedy or lazy, members would still be likely to disagree about what is best for the firm. With the decentralization of information (and of information processing), and a divergence of interests among the members of the firm—i.e., with the *decentralization of incentives*—goes a loss of control. Even if they wanted to, members of the firm could not credibly bind themselves to reveal information honestly or to follow prescribed decision rules. We do not need the theory of games to predict that misrepresentation and moral hazard are likely to lead to inefficiencies, but recent research on principal-agent and partnership models yields some theoretical understanding of the extent to which clever mechanism design and the exploitation of long-term relationships can remedy these inefficiencies. On the other hand, it is just in the context of models of sequential decision making under incomplete or imperfect information—not to mention bounded rationality—that the inadequacies of the present state of noncooperative game theory are most mercilessly exposed. Furthermore, I know of no theoretical research to date that compares the relative efficiency of hierarchical and nonhierarchical organizations within a common model.

Finally, I shall have to admit that research to date has not provided an adequate explanation on economic grounds alone of the conditions under which one expects to see a hierarchical organization of business firms. In fact, the explanation of hierarchy may in many cases be more sociological and psychological than purely "economic" in the mainstream sense. Furthermore, one sees in the current management literature articles that call for a less hierarchical organization of business, whatever that might mean. I suspect that, for economists to contrib-

TABLE 1
1988 "Fortune 500" U.S. Industrial Corporations with More than 100,000 Employees

	Empl. in 1000s	Sales Rank
1. General Motors	766	1
2. IBM	387	4
3. Ford	359	2
4. General Electric	298	5
5. Pepsico	235	26
6. United Technol	187	16
7. Philip Morris	155	10
9. Boeing	147	19
9. Chrysler	146	7
10. Eastman Kodak	145	18
11. DuPont	141	9
12. Digital Equipment	122	30
13. McDonnell Douglas	121	25
14. Westinghouse	120	27
15. RJR Nabisco	117	20
16. Goodyear	114	32
17. Xerox	113	22
18. Rockwell International	112	28
19. Allied Signal	110	29
21. General Dynamics	103	41
21. Motorola	102	52
22. Pillsbury	102	70
23. Exxon	101	3

Source: Fortune, Apr. 24, 1989.

TABLE 2
Twelve Largest Employers among the 1988 "Fortune 500" Non-U.S. Industrial Corporations

	Country	Empl. in 1000s	Sales Rank
1. Coal India	India	670	335
2. IRI	Italy	418	4
3. Siemens	W. Germany	353	7
4. Daimler-Benz	W. Germany	339	5
5. Philips	Netherlands	310	13
6. Unilever	U.K./Netherlands	291	11
7. Fiat	Italy	277	8
8. Hitachi	Japan	264	6
9. Volkswagen	W. Germany	252	10
10. CGE	France	204	24
11. Steel Auth. India	India	201	216
12. Nestlé	Switzerland	198	14

Source: Fortune, Apr. 24, 1989.

ute to that discussion in a scientific way, we shall have to modify our model of economic behavior, not merely—as is beginning to happen on a small scale—to take account of bounded rationality, but also to enrich our model of human motivation.

Let me interject here a note on the style of this paper. I shall not burden you with any formal mathematics, except for some tables and graphs. But I want to emphasize that I shall be talking primarily about the contributions of *formal theorists*, not of historians, astute observers, or even informal theorists. Hence you must be prepared to deal with a number of abstract ideas, even if they are deceptively clothed in the English language.

2. *Managing in the Economy*

Here are some statistics about the largest firms in the world, and about the fraction of the labor force devoted to the activity of managing. Although these few statistics hardly constitute a systematic study, I hope they will persuade the reader that a significant fraction of the industrialized labor force is employed in rather large firms, and that a significant fraction—perhaps more than 40 percent—is devoted to the activity of managing. In fact, the latter fraction has grown steadily during the past century.

Table 1 shows the largest 23 employers among the 1988 "Fortune 500" United States industrial corporations, the ones with at least 100,000 employees. The largest of these is General Motors, with 766,000 employees, which makes it the largest private industrial corporation in the world. Incidentally, the last time I checked, General Motors had about as many employees as there are persons employed in manufacturing in the Netherlands! Table 2 shows the largest 12 em-

TABLE 3
1988 "FORTUNE 500" U.S. INDUSTRIAL CORPORATIONS

Employees	Non-U.S.	U.S.	Combined
≥ 200K	11	5	16
≥ 100K	44	23	67
Median	19.2K	—	—
Mean	37.8K	25.4K	—
Total	18.9M	12.7M	—

Source: Fortune, Apr. 24, 1989.

TABLE 4
U.S. ESTABLISHMENTS AND EMPLOYEES BY EMPLOYEE SIZE CLASS, 1986

Establishments with number of employees at least	Percentage of establishments	Percentage of employees
1000	0.1	12.9
500	0.2	19.8
100	2.0	44.1
20	12.4	73.3

Source: U.S. Dept. of Commerce 1989, p. 512.

ployers among the 1988 "Fortune 500" industrial corporations outside the United States; the smallest of these, Nestlé, has almost 200,000 employees.

Table 3 provides some summary statistics about both U.S. and non-U.S. "Fortune 500" industrial corporations. We see that altogether there are 16 firms with more than 200,000 employees, and 67 firms with more than 100,000 employees. In all the "Fortune 500" non-U.S. industrial corporations, the median employment is 19,200, and the mean is 37,800. In the corresponding 500 U.S. firms, the mean is 25,400.

We are clearly dealing with some very large firms here, but how representative are they? I cannot answer this directly but U.S. government statistics on *establishments* (rather than firms) give an indirect answer. Table 4 shows that establishments with at least 1000 employees account for almost 13 percent of the employees but only 0.1 percent of the establishments. Similarly, 0.2 percent of establishments have at least 500 employees, and account for almost 20 percent of all employees. Of course, the typical large firm will have more than one establishment under its management.

We should not be surprised if the management of such large enterprises re-

TABLE 5
U.S. FULL-TIME WAGE AND SALARY WORKERS, 1987
(MILLIONS)

	Male	Female	All
Managerial and professional	11.6	9.3	20.9
Technical and related support	1.5	1.3	2.8
Admin. support, incl. clerical	3.1	11.1	14.2
Total managing	16.2	21.7	37.9
Total employees	47.2	33.7	80.8
Managing	34.4%	64.4%	46.9%

Source: U.S. Dept. of Commerce 1989, p. 406.
Note: Not all sums and percentages appear consistent, because of rounding.

quires a lot of effort and resources. For example, AT&T currently has about 300,000 employees; of these, about 125,000 or 42 percent, are classified as "exempt from the provisions of the Fair Labor Standards Act" which is roughly equivalent to having a managerial rank. Now some of these "exempt" employees are salespersons, attorneys, and scientists; on the other hand, many "nonexempt" personnel are doing secretarial and clerical work, as well as maintenance of the buildings that house administrators and corporate staffs. A more global picture is reflected in the next table. In 1987 there were approximately 81 million full-time wage and salary workers in the U.S.; of these about 47 percent were engaged in occupations that probably formed part of the activity of managing, either as managers or in support of the management effort. Incidentally, Table 5 also reveals that women make up more than half of the workforce devoted to managing, but also suggests that more than half of managers in the narrow sense are men (the latter statement is confirmed by other sources).

The next table shows how the fraction of the labor force devoted to managing has increased since the beginning of the century, from about 12 percent in 1900 to more than 43 percent in 1980. (Recall that the corresponding figure for 1987 was 47 percent, although the data are not exactly comparable. In constructing Table 6, I was able to make use of relatively detailed occupational classifications.)

I have been unable to extend the last table to years before 1900, but it is possible to trace the growth of the *clerical* workforce from 1870 to 1970, which is shown in Figure 1. This figure shows the percentage that "clerical and kindred workers" made up in the total labor force, starting from less than one percent in 1870, and rising to about 18 percent in 1970. It is also interesting that, although the percentage of clerical workers who were women started out very small, it had passed the 50 percent mark in 1940, and was almost 74 percent in 1970.

TABLE 6
MANAGING IN THE U.S. LABOR FORCE, 1900–1980

	Managing (millions)	Total Experienced Labor Force	Percent Managing
1900	3.4	29.0	11.6
1910	5.6	37.3	15.1
1920	7.8	42.2	18.5
1930	10.2	48.7	20.9
1940	11.5	51.7	22.2
1950	16.3	59.0	27.6
1960	20.9	68.0	30.7
1970	30.7	79.8	38.5
1980	45.6	104.1	43.8

Sources: U.S. Bureau of the Census 1975; 1984, V.1, Ch. D, Part 1, Sec. A, Tables 253–310.

I shall become more precise about what I mean by managing in the next section.

3. *Managing, Decentralization, and Hierarchy*

What is managing? In a nutshell, we might say that it is "figuring out what to do," in contrast to "doing it." This was expressed more eloquently in 1921 by Frank Knight, who stressed the importance of uncertainty and the role of the entrepreneur as a specialist in decision making:

> When uncertainty is present, and the task of deciding what to do and how to do it takes the ascendancy over that of execution, the internal organization of the productive groups is no longer a matter of indifference or a mechanical detail. (Knight 1921, p. 268)

By an extension of this point of view, the managing activity of a firm might be visualized as contained in a large "black

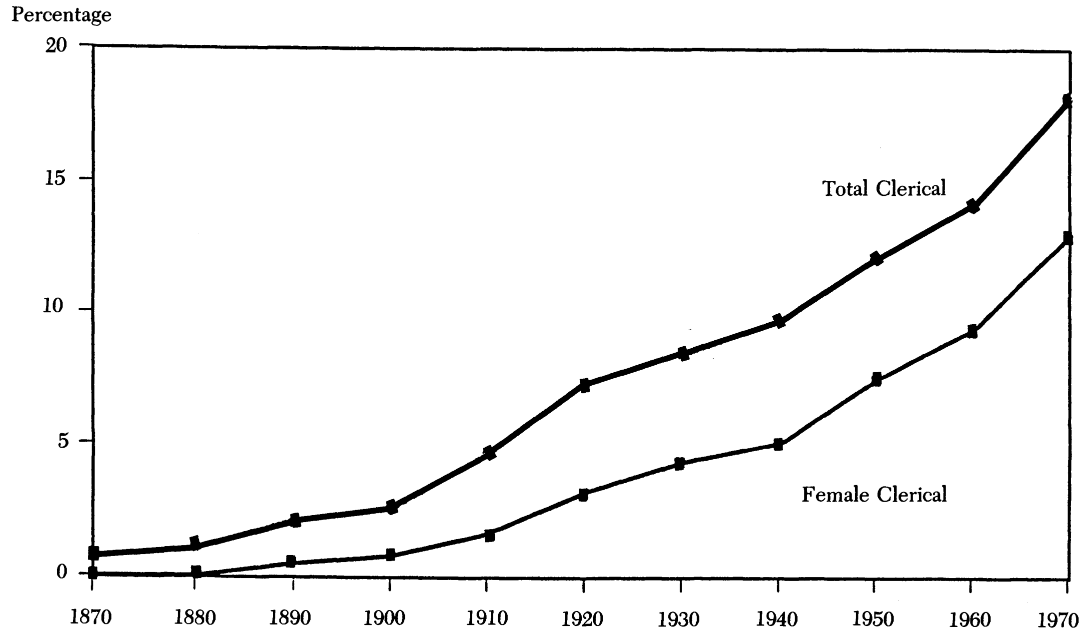

Figure 1. Clerical and Kindred Workers in the U.S. Labor Force, 1870–1970
Sources: U.S. Bureau of the Census 1943, Table XXI, p. 100.

box," into which are periodically fed observations about the firm's environment and the results of its past actions, and out of which come decisions for actions the firm is to take: inputs, outputs, choice of technique, etc. In the modern firm this black box is no longer a single entrepreneur, but is a complex organization including many specialists of different kinds.

At this point, I should say that I am aware that no job in a firm, no matter how routine, is completely devoid of decision-making activity. The blue collar worker on the most routinized assembly line must repeatedly make decisions about how to handle nonstandard situations, and in particular when to call one to the attention of the supervisor. On the other hand, sales managers, in addition to managing their salespersons, often spend considerable amounts of time with clients, engaged in selling, and thus in "doing." Nevertheless, in a large firm it is a useful approximation to divide activities between managing and doing.

Notice that I have used the word "managing" rather than the word "management." In ordinary parlance, the latter word has two meanings: (1) the act of managing, and (2) the collection of persons in the firm called "managers." The first meaning can be "what managers do," or it can have a more general connotation. What do managers do? Here is a partial list:

1. observe the environment and results of past actions
2. process and communicate information
3. make decisions
4. monitor the actions of other firm members
5. hire and fire
6. train and teach

7. plan
8. solve problems
9. exhort, persuade, set goals and values

These are things that would naturally fall under the rubric of managing. In addition, many managers engage in other activities that look more like "doing;" for example, they try to persuade the financial community that the company stock is a good investment, they negotiate loans, they interact with customers and regulators, etc.

In carrying out their (management) activities, managers usually receive help, from staff, secretaries, clerks, equipment, buildings, people who operate and maintain the equipment and buildings, etc. All of these activities and resources I include under the heading of *managing*. In all but the smallest firms, the activities of managing are carried out by more than one person, a phenomenon I have called the *decentralization of managing*.

Activities 1–4 in the above list are usually part of our economic models of statistical decision making and game theory, although we do not usually explicitly consider the resources devoted to them in our "production" or "payoff" functions. Activity 5 (hiring and firing) may also be included in a model of explicit or implicit labor contracts. Training and teaching is sometimes renamed the "production of human capital."

The seventh and eighth activities, planning and problem solving, begin not to fit so comfortably into our standard model. Planning and problem solving might be identified with the choice of strategy, but these activities are not usually modeled as requiring the expenditure of economic resources. (In this connection, where do the activities of research and development fit in?) Finally, the last item makes us downright uncomfortable; in the standard model, peoples' preferences and beliefs (or in the jargon of decision theory, their utility functions and prior probabilities) are exogenously given.

By the *economics of managing* I shall mean the consideration of the resources that go into the activity of managing, and the ways in which different organizations of managing do a better or worse job of economizing those resources and producing good results. To study seriously the economics of managing, one must face squarely the boundedness of rationality of economic decision makers. This phenomenon has long been recognized by theorists, if rarely acted upon. One need not go to critics of modern decision theory, such as Herbert Simon, to find an acute awareness of the boundedness of rationality. In his book, *The Foundations of Statistics* (1954), Leonard J. Savage provided economic theory with the most complete and coherent model of rational economic behavior under uncertainty (in single-person decision problems), but in the same book he emphasized the limitations of the theory as a realistic basis for the description—or even *pre*scription—of rational behavior. In his chapter on "Preliminary Considerations on Decision in the Face of Uncertainty" he comments on the dilemma of decision theory that is reflected in the two proverbs, "Look before you leap," and "You can cross that bridge when you come to it:"

> Carried to its logical extreme, the "Look before you leap" principle demands that one envisage every conceivable policy for the government of his whole life (at least from now on) in its most minute details, in the light of the vast number of unknown states of the world, and decide here and now on one policy. This is utterly ridiculous, not—as some might think—because there might later be cause for regret if things did not turn out as had been anticipated, but because the task implied in making such a decision is not even remotely resembled

> by human possibility. It is even utterly beyond our power to plan a picnic or play a game of chess in accordance with the principle, even when the world of states and the set of available acts to be envisaged are artificially reduced to the narrowest reasonable limits. (p. 16)

(Later in the book, Savage tried to explore some conditions under which out of the grand decision problem of life one could, with approximate accuracy, isolate smaller and more manageable decision problems, but—as I think he recognized—this effort was not entirely successful.)

From a philosophical point of view, it might be argued that an analysis of the economics of managing that is based on the hypothesis of bounded rationality is doomed to failure as a bootstrap operation, or an infinite regress. After all, one way to translate "the economics of managing" is "the management of managing!" This point of view is taken by Pavel Pelikan (1989), who argues that the scarcity of what he calls "economic competence" makes it impossible to determine whether firms are organized to use economic competence efficiently.

Although I might agree with this point of view in principle, in the practice of theorizing I believe that some progress can be made with a more modest and pragmatic approach. Thus, in Section 4, I shall take the point of view that a management organization does implement *some* kind of decision function—or behavior—and that it is interesting to inquire as to what organization structures do this efficiently, i.e., economize on the resources needed to implement the given class of decision functions, given a particular model of the boundedness of rationality. In Section 5, I push this a little farther: given that information processing is costly, what strategies for economizing on information are effective in reducing the amount of information actually used without unduly reducing its effectiveness (value)? In particular, *it is interesting to inquire whether—or under what conditions—there are increasing returns to scale in information-processing and in the use of information.*

Now that I have given some idea of what I mean by the economics of managing, I should explain the remaining word in my title: "hierarchy." Although this is a word that can be given a unique abstract—or mathematical—meaning, in the context of this paper it will take on several different concrete meanings. At the risk of losing part of my audience, I shall start out with the abstract meaning; in fact I shall start by defining a more general concept, that of a *tree*. (I fear that the mathematical name may be misleading, since it corresponds more or less to the botanical object with the same name, but upside-down!) Formally, a *tree* is a collection of objects, together with a relation among them, to be called here "superior to." This relation has the following properties:

1. Transitivity—if A is superior to B, and B is superior to C, then A is superior to C.
2. Antisymmetry—if A is superior to B, then B is not superior to A; in this case I shall say that B is *subordinate* to A.
3. There is exactly one object, called the *root*, that is superior to all the other objects.

I shall say that A is the *immediate superior* of B if there is no object that is "between" A and B in the relation. A fourth property required of a tree is:

4. Except for the root, every object has exactly one immediate superior.

Figure 2 shows a tree with 5 objects. The root is at the top! Notice that not all the objects in the tree need be

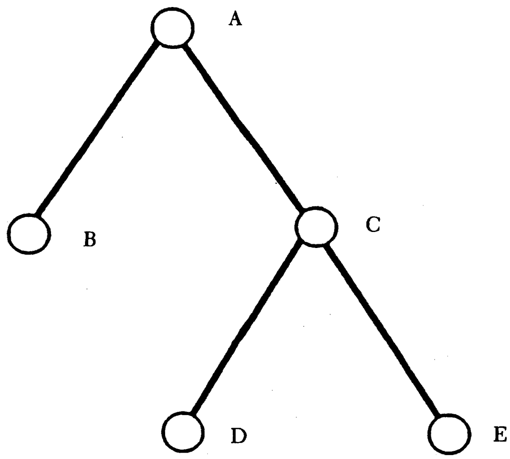

Figure 2. A tree

"comparable;"[1] for example, B is not superior to D, nor is D superior to B.

In everyday language, the word hierarchy not only connotes an upside-down-tree-like structure, but also an assignment of rank or level. By a *ranking* of tree I shall mean an assignment of a number (rank) to each object such that:

1. if A is superior to B, then it has a higher rank (larger number);
2. if A and B have the same rank, then they are not comparable, i.e., A is not superior to B, nor is B superior to A.

I shall adopt the convention that the lowest rank is 1.

I can now define a *hierarchy;* it is a ranked tree. I note that there may be more than one way to rank a tree (in a way that satisfies properties 1 and 2 above); Figure 3 illustrates this. The hierarchies in Figure 3 look like organization charts (for a small organization!). With this interpretation, the relation "superior to" is that of formal authority. In this paper, I shall be discussing other kinds of hierarchies, as well. In fact, the first kind of hierarchy I shall consider is one that represents an organization of tasks or work.

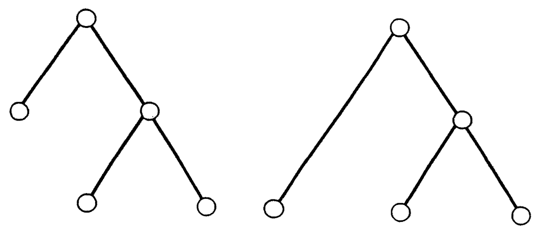

Figure 3. The same tree organized in levels in 2 different ways

In his article, "The Architecture of Complexity" (1962), Herbert Simon gives a wide-ranging discussion of the significance of hierarchy in the structure of complex systems. I shall have to be satisfied here to quote from his parable of the two watchmakers (called Tempus and Hora), which he uses to introduce his discussion of the evolution of complex systems:

> The watches the men made consisted of about 1,000 parts each. Tempus had so constructed his that if he had one partly assembled and had to put it down—to answer the phone, say—it immediately fell to pieces and had to be reassembled from the elements. The better the customers liked his watches, the more they phoned him and the more difficult it became for him to find enough uninterrupted time to finish a watch.
>
> The watches that Hora made were no less complex than those of Tempus. But he had designed them so that he could put together subassemblies of about ten elements each. Ten of these subassemblies, again, could be put together into a larger subassembly; and a system of ten of the latter subassemblies constituted the whole watch. Hence, when Hora had to out down a partly assembled watch to answer the phone, he lost only a small part of his work, and he assembled his watches in only a fraction of the man-hours it took Tempus (Simon 1981, ch. 7, p. 200)

The theme of this parable is repeated in "The Science of Design":

> To design . . . a complex structure, one powerful technique is to discover viable ways of

[1] In mathematical jargon, a tree is a partially ordered set, but need not be completely ordered.

> decomposing it into semi-independent components corresponding to its many functional parts. The design of each component can then be carried out with some degree of independence of the design of others, since each will affect the others largely through its function and independently of the details of the mechanisms that accomplish the function. (Simon 1981, ch 5., p. 148)

Here the hierarchical structure of the system is a reflection of the process of design, rather than the process of construction, although the two aspects might well be related. In fact, design is a particular case of problem solving, and as Simon (1981, p. 206), points out "we can take over the watchmaker parable and apply it also to problem solving." The idea is that a problem is decomposed into subproblems such that, if each is solved then the original problem will be solved. Each subproblem can then be further decomposed, etc., resulting in a hierarchical structure of problems.

Simon's watchmaker parable involved the hierarchical organization of work by a single person. In what follows, it will be useful to have in mind another illustration, which involves the allocation of work among many persons. With apologies to Henry Mintzberg (1979, pp. 1–2), I shall call this the parable of the firm. This firm has two major divisions, Manufacturing and Marketing, plus a Corporate Office (I have deliberately refrained from calling this last the corporate headquarters). The Manufacturing Division contains many production units, some of them spatially separated, and not all of them producing the same product. The Corporate Office performs a number of "overhead-type" activities, including services that are used by both of the other divisions, like payroll and financial accounting, legal services, research and development, etc. As I have described it, one can define a hierarchy in the firm by the relation "is part of." Thus a production unit is part of the Manufacturing Division, which is in turn a part of the firm as a whole. But I have said nothing about any hierarchical organization of authority. In particular, one could imagine that the Corporate Office is simply a service organization, with the task of supplying certain services requested by the two main divisions, but without any authority over them.

One might expect that the organization of work in the firm—in Simon's sense—would have a powerful influence on the way the decentralization of managing is organized. In particular, would a hierarchical design of the processes of production lead to hierarchical management? Some aspects of this will be explored in the following sections, but I can reveal right now that, from a purely theoretical point of view, this is still an open question, and that a full exploration will require substantial additional research. For recent theoretical discussions of the organization of work in modern manufacturing see Masahiko Aoki (1990) and Paul Milgrom and John Roberts (1990).

4. *Decentralization of Information Processing*

Although managers in a firm have many different functions, one of their most important functions is that of processing information. We might think of the information processing part of the firm as one huge decision-making machine, which takes signals from the environment and transforms them into actions to be taken by the "real workers." Of course, as I already pointed out, every worker on an assembly-line or lathe, and every salesperson in the field, makes many decisions every day that are not precisely dictated by management and that has always been so. The point I am making here is that in the modern corporation a large part of the information processing activities are highly *decentralized,* i.e., assigned to a large number of persons in the corporation who specialize in these activities. This is so, even though corporations are thought to be highly centralized from the point of view of authority and supervision.

It is the main theme of this section that hierarchical structures, which are usually thought of as the epitome of the centralization of authority, are also remarkably effective in decentralizing the activities of information processing. As we have seen, the decentralization of information processing is dictated by the large scale of modern enterprises, which makes it impossible for any single person to do it all. Thus the limited capacity of individuals for information processing implies that this activity uses significant amounts of scarce resources, including people, and hence information processing would appear to be a natural object of economic study. On the other hand, it is the computer scientists, more than any others, who have specialized in studying how effectively to organize the resources used in information processing. The present section, then, is on the boundary between economics and computer science. (For bibliographic notes, see the end of this section.)

The economist may have noted that I used the word "effective" rather than "efficient." This is because computer scientists—who are more like engineers than economists—are generally concerned with improving things, rather than with proving that something is optimal under some unrealistic and restrictive assumptions. In this sense, the focus of the present section is more like that of computer science than that of mainstream economics. On the other hand, since my concern is with a human information processing organization rather than with a physical computer or network of computers, my focus is a little different than that found in the current literature on computer science.

Various aspects of the processing of information are costly, and therefore should be "economized:"

1. the observation of the data about the environment.
2. the capabilities and numbers of the individual processors (persons, equipment).
3. the communication network that transmits and switches the data (both original and partly processed) among the processors.
4. the delay between the observation of the data and the implementation of the decision(s).

The last aspect, delay, is costly to the extent that the delayed decisions are obsolete (e.g., not timely).

In fact, computer scientists have been largely concerned with delay, that is, the time it takes to compute a particular function. Here I shall give equal attention to economizing the number of processors. I do not here consider the cost of communication; this has, I think, some empirical justification in human organizations. The problem of "information overload" is apparently particularly acute in modern times, and is a reflection of the relative cheapness of communication compared to processing (digestion) of information.

Also, in the present section I take the amount of environmental data as given, but in fact it should be an endogenous variable, determined by the balance between its cost and its value. This consideration is, however, deferred to the next section.

I shall now consider explicitly two paradigms of the transformation of environmental signals into decisions: (1) linear decision rules, and (2) pattern matching.

In the first, the decision is a linear function of the environmental signals. This corresponds, in particular, to the typical processing of accounting information. Numerical data are rescaled to common units, like dollars or minute-miles, and then added up. Thus we may think of the calculation of a linear function as occurring in two stages: (1) each variable is multiplied by its respective coefficient (conversion to a common unit), and (2) the resulting products are added up (aggregation). The items to be aggregated might well be vectors, not just numbers, so that the coefficients are matrices. The decentralization of computing the linear function is dictated by the fact that the number of items to be added (numbers or vectors) is very large.

In the second paradigm, the decision maker, i.e., decision-making organization, compares the "pattern" of data about the environment with the members of a finite set of reference patterns, picking the one that is "closest" in some sense. To each reference pattern corresponds a decision, so that the problem of choosing a decision is reduced to the calculation of the closest reference pattern. For example, the data and reference patterns might be represented as vectors in a space of very large dimension, and the measure of closeness might be ordinary Euclidean distance. This dimension is so large that no single processor can handle an entire vector at once. Each huge vector will, therefore, have to be divided up into smaller component vectors for processing. In fact, we can imagine that the original observations themselves are on the component vectors, rather than on the entire vectors.

It is interesting that both "addition" and "finding a minimum" are associative operations,[2] and thus lend themselves naturally—as we shall see—to decentralization, or what the computer scientist would call *parallel computation* (Jacob Schwartz 1980). In fact, from a formal point of view these concepts are the same in the context of information processing.[3]

These considerations lead me to consider the following problem: Given N items to be added, and P processors, arrange and program the processors to add the N items in minimum time. Here, for "add" we can read any associative operation, and the "items" are anything amenable to the associative operation. As I have stated it, this problem is not well defined, because I have not been precise about what a processor is. In what follows, a *processor* is an object with an *in-box*, a *register*, and a *clock*. Time is measured in cycles; in one cycle a processor can take one item from its in-box and add it to its register. From each processor there may be one or more one-way communication links to other processors. At prescribed times, a processor can also send the contents of its register to the in-boxes of other processors to which it is directly linked, and then reinitialize its own register to "zero;" this can be done in any cycle without additional elapsed time. Finally, there is a particular processor that, at a designated time, sends out the contents of its register as the result of the computation, i.e., the grand total. The set of processors and links will be called a *network*. The *program* prescribes the original assignment of the items to the several processors' in-boxes, and the times of communication and final output. The number of cycles used to perform the computation will be called the *delay*.

I can now restate the problem: *arrange*

[2] A binary operation, say $*$, is associative if $(A * B) * C = A * (B * C)$. Addition, multiplication, minimum, maximum, set-union, and set-intersection are all associative. The operation "x to the power y" is not.

[3] Another computer science term for decentralized processing is "distributed computation." In the computer science literature, the terms "parallel" and "distributed" connote different sets of research problems (and usually different researchers), but both are concerned with what I have called decentralized information processing.

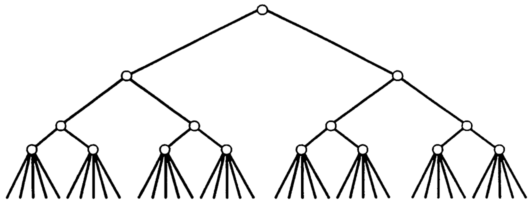

Figure 4. A hierarchical network with 15 processors, 40 items, and delay 11

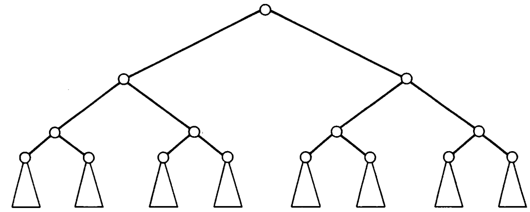

Figure 5a. Regular hierarchy before reduction

the given P processors in a network, and program the network, so as to add the given N items with a minimum delay,

Figure 4 illustrates a hierarchical network, with 15 processors indicated by circles, and 40 items. The links joining the processors are to be understood as pointing upward. One processor is the immediate superior of another if there is a direct link pointing upward from the second to the first. The successive levels in the diagram indicate the ranks, of which there are 4. Here is the way the program works. The 40 items are originally assigned equally to the processors of the lowest level (rank 1). This is indicated in figures by the 5 lines coming up into each of the 8 lowest-levels processors. The computation starts with each first-level processor adding its 5 items into its register. At the end of the 5th cycle each first-level processor then takes 2 cycles to add its items and send its partial sum to its third-level immediate superior, etc. At the end of 11 cycles the single fourth-level processor (the root of the tree) puts out the grand total of the 40 items; thus the delay is 11.

As we shall see, the minimum achievable delay is actually 8, not 11. Indeed, we shall also see that one can achieve the same delay, 8, with fewer than 15 processors, namely 8.

I shall say that a *network is efficient* for a given number of items if the number of processors cannot be decreased without increasing the delay, or vice versa. Thus the network in Figure 4 is not efficient. By extension, I shall say that *the pair (P,C) is efficient for N items* if there is a network with *P* processors that is efficient for *N* items and adds them with a delay equal to *C*.

Although the hierarchical network of Figure 4 is not efficient, one can show that hierarchical networks are sufficient to attain efficiency, in the following sense: *For any number of items, N, and any pair (P,C) that is efficient for N, there is a hierarchical network with P processors that can add the N items in C cycles.*

I shall show how to construct such efficient hierarchies, but first I need to introduce some terminology. The hierarchy in Figure 4 has a symmetric appearance. I shall call a hierarchy *regular* if (1) all the immediate subordinates of any processor are at the next lower level and (2) at each level above the first, all members of the same level have the same number of immediate subordinates. All of the processors at one level that are the immediate subordinates of the same processor at the next higher level will be called a *cadre*. The hierarchy in Figure 4 is not only regular, but each cadre has two members. (In general, however, in a regular hierarchy cadres at different levels need not be of the same size.)

I can now describe how a regular hierarchy can be "reduced" so as to decrease both the total number of processors (*P*) and the delay (*C*). I suppose we start with the *N* items allocated equally (or as equally as possible) among the lowest-level processors. Figure 5a shows the hierarchy of Figure 4, but with each group

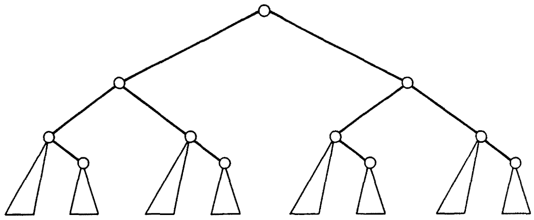

Figure 5b. Regular hierarchy after 1 stage of reduction

of 5 items at the bottom replaced by a triangle, or "fan." The reduction will be done in stages; at each stage we reduce the amount of idleness in the network. At stage 1, we eliminate one member of each cadre at level 1, and assign its items to its corresponding immediate superior. Figure 5b shows the result of applying the first stage of reduction. As there were originally 4 cadres at level 1, 4 processors have been eliminated at level 1, reducing the total number of processors from 15 to 11. Each second-level processor now has 5 items plus 1 first-level processor assigned to it.

I shall call the items and/or immediately subordinate processors assigned to a processor its *predecessors*. Let R denote the number of levels in the hierarchy. The reduction procedure is completed in stages as follows: at stage $r < R$, one processor is eliminated from each level-r cadre, and its predecessors are assigned to its immediate superior at level $r + 1$.

Figures 5c and 5d show the second and third stages of reduction for the hierarchy of Figure 5a. The number of pro-

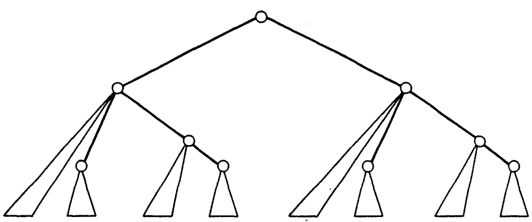

Figure 5c. Regular hierarchy after 2 stages of reduction

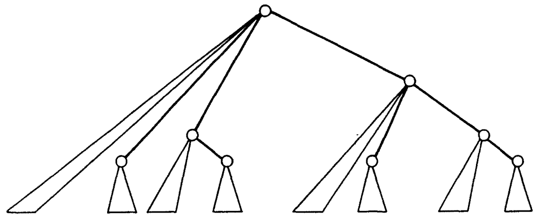

Figure 5d. Regular hierarchy after 3 stages of reduction (final)

cessors has been reduced from 15 to 18, and the number of cycles from 11 to 8. Although there is still some idleness in the network, no further increase is efficiency is possible.

There is, however, something odd about Figure 5d, at least as a picture of an organizational hierarchy. We see that the top ranking processor has immediate subordinates at all levels. In fact, a similar phenomenon is repeated at each lower level. Reporting through skipped levels is not unheard of in corporate hierarchies (in fact, at AT&T this is called "skip-level reporting"), but the practice does not seem to be as widespread as the above reduction process would suggest.[4]

I shall now give the solution to the problem previously posed. Recall that, for integers N and P, N mod P denotes the remainder after dividing N by P. To add N items with P processors, the minimum delay is given by the following formula:

$$\text{Min } C = \lfloor N/P \rfloor + \lceil \log(P + N \bmod P) \rceil,$$

where the brackets $\lfloor\ \rfloor$ denote rounding down to the nearest integer, the brackets $\lceil\ \rceil$ denote rounding up to the nearest integer, and the logarithm is taken to the base 2. Furthermore, this minimum de-

[4] Figure 5d is somewhat misleading because in the original hierarchy each cadre had only two members. If we had started with larger cadres, then the relative importance of skip-level reporting would have been smaller.

TABLE 7
MINIMUM DELAY (C) FOR ADDING 40 ITEMS

P	MIN C	
1	40	*
2	21	*
3	15	*
4	12	*
5	11	*
6	10	*
7	9	*
8	8	*
9	8	*
10	8	
11	8	
11	8	
12	7	*
13	7	
..	..	
20	7	
..	..	
40	7	

Source: see text.
P = number of processors.
* denotes an efficient network

lay is attained by a hierarchy of the type that one gets from reducing a regular hierarchy, as described above. Indeed, it is attained by reducing a regular hierarchy whose cadres originally all have size two! (Note that the hierarchy of Figure 5a does have this property.)

Table 7 illustrates this formula for $N = 40$ and P varying from 1 to 40. A one-processor hierarchy is the slowest, with a delay of 40 cycles. The minimum delay is 7 cycles, and can be attained with 12 processors; the use of more processors will not further reduce the delay. Thus a network with more than 12 processors is not efficient for 40 items. Similarly, we see from the table that networks with 9, 10, and 11 processors are not efficient, either, since a delay of 8 cycles can be attained with 8 processors. Thus, although the formula gives the minimum number of cycles for any number of processors, not every pair (P,C) generated by the formula is efficient. Figure 6 shows a graph of Table 7, and illustrates the same phenomenon. For very large numbers of items, however, these inefficiencies will not be very significant unless the number of processors is too large. For example, Figure 7 shows a plot of the minimum delay vs. the number of processors for $N = 10{,}000$; inefficiency does not become a significant problem until the number of processors exceeds 2000 (approximately). (Note that the scales on both axes of Figure 7 are logarithmic.)

Figures 6 and 7 illustrate the tradeoff between delay and number of processors. This tradeoff is a reflection of the tradeoff between serial and parallel processing. With few processors, there is little parallel processing of the items at the first level, which causes a large delay. Many processors permit much parallel processing, which reduces the delay.

I must now address a fact about real organizations I have thus far ignored, namely, that typically new data about the environment will be coming in periodically, with the consequence that new decisions must be calculated periodically. (Computer scientists sometimes call this the *systolic mode.*) This can cause a problem for our highly efficient reduced hierarchies if the new data arrive too frequently. The reason is that, in an efficient reduced hierarchy, *the highest ranking processor is busy all the time.* Thus, if the time between periodic arrivals of new data—or as I shall call them, new *cohorts* of data—is less than the delay in computing each sum, then the backlog of unprocessed cohorts, and the delays in the calculation of the corresponding sums, will also increase without bound. In this case, the successive decisions will become unboundedly obsolete! (Notice that this problem does not arise in a regular hier-

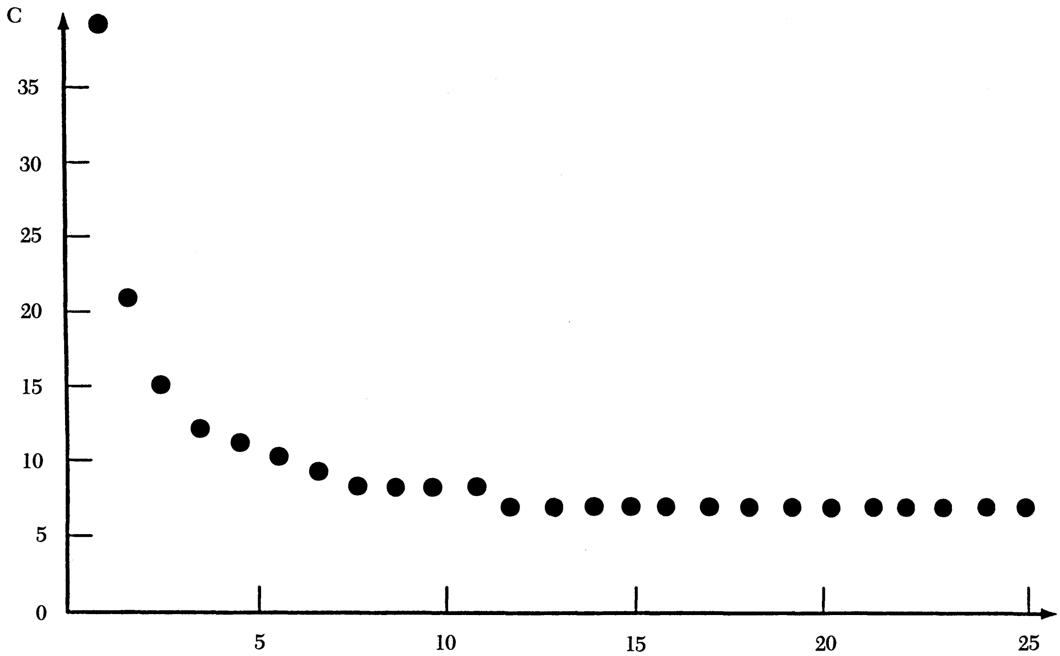

Figure 6. N = 40
Min C vs. P

archy unless the time taken by one *level* exceeds the time between cohorts.)

Timothy Van Zandt (1990) has recently solved the problem of constructing efficient networks for the systolic mode. Roughly speaking, one assigns an efficient "one-shot" tree to each incoming cohort; as processors are successively freed up from working on one cohort, they are assigned to the next available cohort. (This is not quite correct, however; paradoxically, one can usually further reduce idle time and average delay by adding some processors to the efficient "one-shot" tree!) Typically, the total number of processors, P, will be larger than the number of processors assigned to any one cohort, say Q. Although a simple formula for the minimum delay is not available for the systolic mode, the following pair of equations determines an approximation to the efficiency frontier:

$$C = (N/Q) + \log(Q),$$
$$P = (N + Q - 1)/T,$$

where a new cohort arrives every T cycles, and is processed by Q processors (on the average). (These equations are exact in those cases in which all idle time has been eliminated, which, however, is usually not possible. In any case, the equations provide a *lower bound* on all feasible pairs (P,C).)

As one varies Q between 1 and N, one traces out the $P - C$ efficiency frontier (approximately). In fact, one can get quite close to this frontier with a symmetrical hierarchy, as follows. The network is made up of one or more groups of "preprocessors" and a single "overhead" tree, which is symmetric. Each incoming cohort is assigned to one group of preprocessors; as each group finishes its task of preprocessing it sends its par-

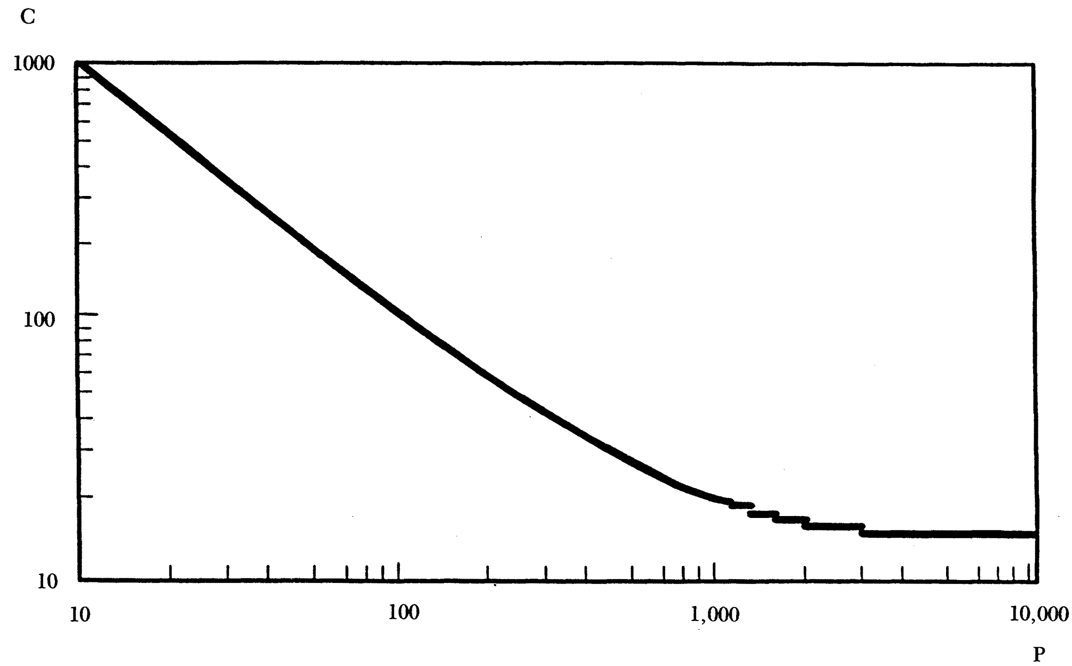

Figure 7. N = 10,000

Min C vs. P

tial sums to the overhead tree and then turns to the next available cohort. The overhead tree is designed so that it can handle the successive groups of partial sums without creating a backlog. I call the resulting network a "preprocessing overhead tree" (POT). Figure 8 shows the (P,C) points corresponding to selected POTs, superimposed on the graph of the approximate efficiency frontier. One sees that the POTs are close to efficient except when P is large.

I turn now to the question of returns to scale. What, exactly, should we mean by this? Although I have referred to the processing delay as "costly," it is probably more in conformity with the economic concept of a production function to regard delay as a "quality" of the output—the answer—with the processors and the items to be processed as the inputs. From this point of view, the above equations tell us two things. First, as the second question shows, the number of processors, P, must increase at least in proportion to N, the number of items to be processed. Second, and more striking, the first equation shows us that, as the number of items increases, the delay, C must also increase; even if the number of processors is unlimited, the minimum delay is $(1 + \log N)$. Thus we have decreasing returns to scale in a very strong sense.

Is there any escape from this dismal conclusion? If we recall that the purpose of the information processing is to provide information for decision making, then we might ask: what are the returns to scale from the use of the information, taking account of the fact that the information must be processed before it can be used? Unfortunately, the answer to this question will depend upon the par-

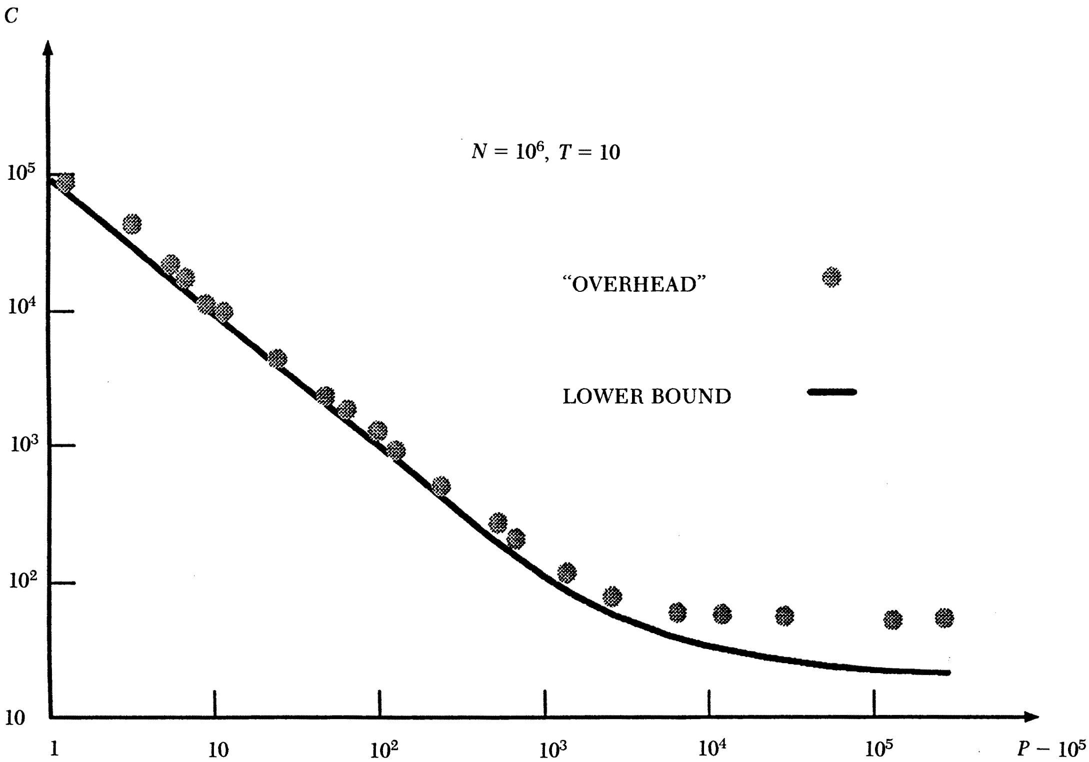

Figure 8. The Systolic Mode

ticular decision problem for which the information is used. In particular, it will depend upon: (1) the statistical properties of the environment and the observed data, and (2) the functional form that describes how the loss due to an incorrect decision depends on the "decision error." It remains to be seen whether something useful and general can be said about this problem; from a series of examples it would appear that one can get anything from decreasing to increasing returns to scale, depending on the assumptions that one makes about the decision problem. More detailed discussions are provided by Radner (1989) and Radner and Van Zandt (1992).

Now to summarize this section of my paper. In the context of a decision-theoretic model of the firm, I have represented managers and their helpers as individual information processors of limited capacity, and using ideas from computer science I have explored the efficiency of different "architectures" for the network of manager-processors. Here "efficiency" is measured in terms of (1) the number of processors and (2) the delay between the receipt of information by the organization and the implementation of decisions. The tradeoff between these two "costs" is achieved by varying the degree of *parallelism* in the network of processors relative to the amount of *serial* processing done by individual processors. The first important conclusion is that *efficiency can be achieved by hierarchical networks.* The second is that, if one regards the delay as the quality of the output, then there are decreasing returns to scale in a strong sense, namely, one cannot maintain constant quality indefinitely as one increases the number of items to be processed, even if the processors are free. However, if one considers the value of the information that is processed, in the context of a particular decision problem, then it appears that—depending on the decision problem—one can get anything from decreasing to increasing returns to scale,

I conclude this section with a few bibliographic notes. As I noted above, two strands of literature have influenced the ideas presented here, one from economics, and one from computer science. Thomas A. Marschak and C. Bartlett McGuire (1971) were probably the first to propose the model of a finite automaton as a formalism of the notion of a boundedly rational decision maker. The model of a decision-making organization as a network of information processors was explored by Jacob Marschak and myself (1972, ch. 9), but our analysis was concerned more with the decentralization of information than of information processing. In a similar spirit, Thomas A. Marschak and Stefan Reichelstein (1987) studied conditions under which a "hierarchical" structure of decision making would be efficient in a broader set of structures. In their model, every processor is also responsible for the final decision about some action variable, and the only cost of processing is that of communication. Their analysis derived some conditions under which hierarchy would be preferred, but I shall not attempt to summarize their results here. Two papers of Michael Keren and David Levhari (1983, 1989) provide an alternative way to look at the problem of minimizing the costs of information processing in an organization. In the "economics strand," the current research of Stanley Reiter and Kenneth Mount is most closely related to the present paper, and I have benefited from exposure to that research in conferences (Mount and Reiter 1982) and in their draft manuscript, "A Model of Computing with Human Agents" (1990). Their focus is primarily on the computation of equilibria of resource allocation mechanisms, and on the tradeoff between the delay and the amount of information required to be communicated. The latter is measured, roughly speaking, by the dimension of the space of messages utilized in the allocation mechanism.

Thomas A. Marschak (1986) and Jacob Marschak and Roy Radner (1972, ch. 7) studied the effect of delay on the value of decisions, particularly in the context of decentralization of information. As pointed out above, the cost of delayed decision may not be simply proportional to the delay; the functional form of the dependence of cost on delay will depend on the intertemporal statistical properties of the stochastic process of environmental data.

From the "computer science strand," the reader familiar with the article, "Ultracomputers," by Schwartz (1980), will recognize how heavily I have relied on the ideas in that paper. I have also benefited from the paper by Clyde Kruskal et al. (1988).

Recently, some game theorists have used the automaton model to explore how the boundedness of rationality might alter the predictions of Nash equilibrium theory, especially in sequential games (see Ariel Rubinstein 1986; Abraham Neyman 1985; Ehud Kalai and William Stanford 1988; Dilip Abreu and Rubinstein 1988; and the references cited there). In most of these explorations, however, the boundedness of a player's rationality is expressed only by a bound on the number of states of the automaton, a direction that is quite orthogonal to the one taken here.

5. *Decentralization of Information*

In the previous section, I focused on the process of producing a decision from a large amount of data, so large that the information processing task of producing the decision had to be divided up—decentralized—among a number of separate processors. The decision was the output of a single processor, for example, the top of the hierarchy.

In a firm, there are many, many different decisions to be made, and it is totally impractical for them to be put out by the same processor, and as a function of the same information. This situation, in which different decisions are based on different information, I shall call the *decentralization of information*.

In principle, the same network could be used to compute different decisions from different sets of incoming data (using, of course, different programs). But in practice this never happens for all of the decisions in a firm. We sometimes read of an executive who tries to decide everything, is unwilling to delegate any decisions to subordinates, but the point of such stories is that these executives get into trouble. Even such stories exaggerate; *no executive can truly decide everything in a timely fashion!*

The same limitations that prevent any one person from deciding everything also make it uneconomical for all decisions to be based on the same data. We observe that firms take in enormous amounts of data every day, but most decisions are based on a very small part of it, and the reason may seem obvious. For most decision problems, a relatively small amount of information enables one to make a fairly good decision, if the information is chosen wisely. On the other hand, the information required for a good decision is typically different for different decision problems.

Let me illustrate this in the context of the "parable of the firm' (Section 3). A small decision about the maintenance of a machine can be made quite well based on information local to the production unit; similarly, a good decision about the order in which to visit customers can be made on the basis of information local to the sales office. On the other hand, a decision to change the rate of production should be based, in part, on information about conditions at other production units, as well as on information about demand for the product. But even in this case, completely detailed information about all other production units and all sales units is not really needed; certain aggregate measures will be adequate.

The efficient use of information in an informationally decentralized organization is the subject of the *theory of teams*, an early development in the formal theory of organization. This theory, originally proposed by Jacob Marschak, has also been taken up by applied mathematicians in the area of decision and control theory (Jacob Marschak and Radner 1972; McGuire and Radner 1986; Ki H. Kim and Fred W. Roush 1987; for an elementary exposition, see Radner 1986). In the theory of teams, efficiency is evaluated in terms of an overall organizational goal or objective function (e.g., profit in the case of the firm). The focus is on (1) the incomplete and heterogeneous[5] dissemination of information among the several decision makers (i.e., informational

[5] A note on terminology: it has become common to use the term "asymmetric information" for what I have here called "heterogeneous information" or "informational decentralization." This is unfortunate, as the following example shows. Suppose that there are two decision makers, *A* and *B*, and that *A* costlessly observes information *X*, whereas *B* costlessly observes *Y*. Consider three different information structures: (1) *A's* decision is based on *X* alone, and *B's* on *Y* alone, i.e., there is no communication; (2) *B* communicates *Y* to *A* and then *A's* decision is based on *X* and *Y*, while *B's* is based on *Y* alone; (3) there is complete communication, so that *A* and *B* each base their decisions on *X* and *Y*. In ordinary as well as mathematical parlance, one would normally say that information structures 1 and 3 are symmetric, whereas 2 is asymmetric. On the other hand, in structures 1 and 2 information is decentralized or heterogeneous, whereas some current authors would call it "asymmetric."

decentralization), (2) the characterization of decision functions that are optimal, given that decentralization, and (3) the comparison of alternative (decentralized) information structures, under the assumption that each one will be used efficiently. In this theory, no attention is paid to the private incentives of the individual decision makers, who might as well be computers. In this sense, the theory of teams occupies a middle ground between the theory of decision for a single person and the theory of games.

This is not the place to review systematically research on the theory of teams. However, I want to call attention to two propositions that are relevant to my previous discussion of information processing and hierarchy.

First, under certain regularity conditions, for a given structure of information each decision maker's optimal decision function will be approximately linear in his information; to put it another way, linear decision rules will be approximately optimal.[6] This provides a justification of the analysis in Section 4 of the processing of information using associative operations. Recall the conclusion there that—under certain conditions—efficient processing could be achieved by hierarchical networks. In the situation I am now discussing, with decentralized information, the implication is this: for each decision maker, there is an efficient hierarchy to calculate his decision as a function of his peculiar information. *But these hierarchies will typically be different for different decision makers, since different decisions will be based on different sets of original data items.* When we think about this situation carefully, we see that the formulation of Section 4 was not really adequate to consider the problem of efficient information processing in a team. It rarely will be efficient to use a different network for *every* decision, or even for every decision maker. On the other hand, in all but the very smallest organizations, there will be different networks for different sets of decisions and decision makers. In fact, empirical studies of "informal organization" have shown this to be universally true (Arnold S. Tannenbaum 1966 and Mintzberg 1989 on "adhocracy"). As far as I know, there has been no formal analysis of the general question of how to organize processors of given capabilities to compute a *number of different decisions functions* in an economically efficient manner, i.e., the *team-theoretic analogue* of the problem discussed in Section 4.

The second proposition I shall discuss has to do with what is sometimes called "management by exception." By this I mean a behavior in which the value of an observed variable or pattern is reported to a superior only when it is "exceptional" or "unusual." For example, the maintenance status of equipment in a particular production unit would not be reported to the production vice president (except possibly at infrequent periodic intervals) unless there were a massive breakdown that threatened to upset the planned production schedule. Theoretical studies, as well as practical experience, suggest that management by exception is a powerful device for economizing on the use of information in an organization (Jacob Marschak and Radner 1972, ch. 6). On the other hand, those theoretical studies did not explicitly integrate team theory with the new model of decentralized information processing of Section 4. Notice that, with management by exception, the number of items to be processed in any cohort of data is not constant, but will fluctuate randomly

[6] Important exceptions to this proposition can arise when one person's information is about other persons' actions; this was first shown by Hans Witsenhausen (1968), who independently formulated a team-theoretic framework.

from one cohort to the next. The effect of this is that there will be *stochastic queues* of items in the in-boxes of the processors, rather than a regular flow, and hence the decision delays will also be stochastic. Similar analytical problems arise in the study of telecommunications networks, especially data networks, and one hopes that similar methods will be useful here. (For further recent research in the spirit of team theory see Jacques Crémer 1980; John Geanakoplos and Milgrom 1991; and Raaj K. Sah and Joseph E. Stiglitz 1986.)

6. *Decentralization of Incentives*

I have argued that, in any but the smallest firms, no one person has all of the information relevant to the firm's activities. It follows that no one person can completely control all of those activities. This is so even in firms that are described as highly "centralized." From this fundamental observation, it follows that individual members of the firm will have some freedom to choose their own actions. If in addition, there is some divergence among the members' goals or objectives, then one can expect some inefficiencies to arise in the firm's operations. The theoretical analysis of these inefficiencies, and the possible remedies by means of organizational design, are the subject of the present section.

In this discussion, I shall concentrate on the members of the firm who do have some freedom of action, and I shall therefore call them *agents*. If the behavior of the agents is "rational" in the sense typically used by economists and decision theorists, then the appropriate formal model would appear to be the theory of games, especially games of incomplete information, as developed in the past two decades (John Harsanyi 1967–68; Roger Myerson 1985). Furthermore, the relationships among members of a firm are typically long lived, calling for an analysis in terms of dynamic games.

Two special paradigmatic models have arisen in the game-theoretic analysis of the firm. In the first, which I have elsewhere called a *partnership*, the agents act together to produce a joint outcome (e.g., output or profit). This outcome can be observed by the several agents, but they cannot directly observe each others' actions, nor do they completely share each others' information. In the most general—and realistic—case of this theoretical model, the outcome is also influenced by random variables that are only partially observed, if at all. The incompleteness of the information leads to what the statisticians call a "confounding" of the sources of variations of the outcomes, making it difficult to assign responsibility to the individual agents for the occurrence of unsatisfactory outcomes. It is this confounding that leads to organizational inefficiency, if the goals of the agents are not identical. In particular, the agents can engage in what is colloquially called "free riding."

The so-called "principal-agent" model provides a second paradigm, which may be relevant to hierarchical organization. In the simplest principal-agent model, there are two players. The first, the *principal*, performs no immediately useful actions himself, but monitors the activities of the other player, the *agent*, whose actions, together with the (stochastic) environment, determine the outcome. In the standard model, this outcome is a number, like money. The principal can observe neither the agent's action nor the stochastic environment, but both players can observe the outcome. Thus we again have a situation in which the confounding of the sources of variation of the outcome makes it difficult to correctly assign to the agent responsibility for the occurrence of satisfactory outcomes. However, the phenomenon of free riding on others is absent. The principal is restricted to rewarding the agent according to the outcome, and then retains the residual for himself. In a more elaborated version of the model, the agent and the principal may each have some information about the environment, but their information is not the same. In addition, the principal may have partial, but not complete, information about the agent's action.

Thus there may be two different sources of inefficiency in the principal-agent relationship. The first, called *moral hazard*, refers to the fact that the principal cannot accurately monitor the agent's actions, and hence experiences a loss of control over the agent. The second source exists when the agent knows something the principal does not, leading to *misrepresentation* or *adverse selection*.

In fact, most organizations combine aspects of both the partnership and principal-agent models. A hierarchy of authority can be thought of as a cascade of principal-agent relationships, each supervisor acting as a principal in relation to his subordinates, and as an agent in relation to his own supervisor. On the other hand, in most cases the valued outcomes of organizational activity depend on the joint actions of several agents, as in the partnership model, so that the assignment of individual responsibility for specific outcomes—as required by the principal-agent model—may not be justified. Unfortunately, I am not aware of significant progress on more comprehensive theoretical models of the firm that combine these two submodels in a systematic way.

In a less formal way, economists have long recognized aspects of principal-agent relationships in the firm. Adam Smith (1776, vol. 2) took a dim view of the relationship between the shareholders and the board of directors and/or senior management. After describing the difference between a joint stock company and a "private copartnery," he goes on to offer this opinion of joint stock companies:

> The trade of a joint stock company is always managed by a court of directors. This court, indeed, is frequently subject, in many respects, to the control of a general court of proprietors. But the greater part of those proprietors seldom pretend to understand anything of the business of the company, and when the spirit of faction happens not to prevail among them, give themselves no trouble about it, but receive contentedly such half-yearly or yearly dividend as the directors think proper to make to them. The directors of such companies, however, being the managers of other people's money rather than of their own, it cannot well be expected that they should watch over it with the same anxious vigilance with which the partners in a private copartnery frequently watch over their own. Like the stewards of a rich man, they are apt to consider attention to small matters as not for their master's honor, and very easily give themselves a dispensation from having it. Negligence and profusion, therefore, must always prevail, more or less, in the management of the affairs of such a company. It is upon this account that joint stock companies for foreign trade have seldom been able to maintain competition against private adventurers. (pp. 264–65)

Smith certainly understood the incentive problems inherent in the principal-agent relationship. On the other hand, it appears that this judgment of joint stock companies was too pessimistic. What would he have thought of the "Fortune 500?"

Frank Knight (1921) called attention to the principal-agent relationship between the entrepreneur and his workers. Here he focused on the allocation of risk between the entrepreneur and the worker. He argued that important characteristics of social organization can be traced to

> the system under which the confident and the venturesome assume the risk or insure the doubtful and timid by guaranteeing to the latter a specified income in return for an assignment of the actual results . . . With human nature

> as we know it, it would be impractical or very unusual for one man to guarantee to another a definite result of the latter's actions without being given the power to direct his work. And on the other hand the second party would not place himself under the direction of the first without such a guarantee. . . . The result of this manifold specialization of function is the enterprise and wage system of industry. Its existence in the world is the direct result of the fact of uncertainty. (pp. 269–70)

The allocation of risk in the presence of moral hazard is, in fact, one of the dominant themes of the contemporary principal-agent literature. The term "moral hazard" itself arose in the insurance industry, and the first formal economic analysis of moral hazard was probably given by Kenneth J. Arrow (1963) in his paper, "Uncertainty and the Welfare Economics of Medical Care."

The standard principal-agent model typically *assumes* that the principal is neutral towards risk and the agent is averse to risk. In such a context it is easy to see the conflict between insurance and incentives. Since the outcome of the agent's action is also influenced by random factors beyond his control, and he is averse to risk, he would like to be insured against this uncertainty. Since the principal is risk-neutral, an optimal allocation of risk requires that the principal bear all of it, i.e., that the agent's compensation be independent of the outcome. In this case, the agent would get a fixed compensation (wage), and the principal would get the outcome minus this fixed payment. However, with a payment independent of the outcome, the agent would have no incentive to try to make the outcome higher than lower. For example, if the outcome depends on the agent's "effort," and the agent is lazy, then the agent will not be likely to put in an optimal amount of effort; to induce the agent to put in the optimal effort, his compensation would have to depend on the outcome. Thus there can be no scheme for sharing the output between the principal and the agent that simultaneously induces the optimal effort and fully insures the agent against risk.

Although this argument may sound plausible, I have not defined what I mean by "optimal," nor have I been precise about my theory of the behavior of the principal and agent in this situation. Because some readers may not be familiar with game theory, I shall take a little space to be more precise. I shall describe the principal-agent situation as a two-move game. The principal moves first, announcing a *compensation function*, namely, a schedule that determines the agent's compensation for each possible outcome. The agent moves second, choosing his action. The outcome is then determined as a function of the agent's action and some unobserved random variable. (Put another way, the probability distribution of the outcome depends on the agent's action.) The outcome is then observed by both players, and the agent is compensated by the principal according to the previously announced compensation function. The resulting utility to the principal is the difference between the actual outcome and the compensation that he pays the agent. This expresses the assumption that the principal is neutral towards risk. The resulting utility to the agent depends both on the action he has chosen and on the compensation that he receives, in a way that represents his aversion to risk. Each player is interested in maximizing his own *expected* utility. In this game, the principal's *strategy* is the same as his move, namely the announced compensation function, but the agent's strategy is a decision-rule that determines his action corresponding to each alternative compensation function that the principal could announce.

An *equilibrium* is a pair of strategies, one for the principal and one for the agent, such that:

1. Given the announced compensation function, the agent chooses an action that maximizes his own expected utility.
2. Given the optimizing behavior of the agent, the principal chooses a compensation function that maximizes his own expected utility.

In the formulation of a principal-agent model one typically adds one or both of the following constraints on the compensation function that the principal may announce. First, the compensation function must enable the agent to attain (ex ante) an "acceptable" expected utility. This constraint can be interpreted as requiring that the principal must offer the agent an expected utility at least as large as what the agent could obtain in other employment. Second, the agent's compensation is bounded below by some exogenously given bound. This second constraint recognizes that the agent's wealth is finite, and so the agent cannot be compelled to pay the principal arbitrarily large amounts of money (negative compensations).

A pair of strategies is defined to be *Pareto optimal* or *efficient* if no other strategy-pair yields one of the players more expected utility and yields the other no less. A basic proposition of principal-agent theory is that, with the above assumptions and some additional "reasonable" conditions, *an equilibrium is not efficient*. Here is a sketch of the argument. First, I shall argue that, since the principal is neutral towards risk, and the agent is averse to risk, in an efficient strategy pair the agent's compensation must be independent of the outcome. Suppose, to the contrary, that different outcomes led to different compensations, and let w be the *expected compensation.* Since the agent is averse to risk, he would be better off if he used the same action but received a fixed compensation equal to w. The principal, on the other hand, would be no worse off in this new situation, since he is neutral towards risk. (Indeed, if one wanted to make both players strictly better off, the principal could pay the agent a fixed compensation slightly less than w.)

On the other hand, a strategy-pair in which the agent's compensation does not depend on the outcome typically cannot be an equilibrium, unless by coincidence the action that the agent most prefers is also part of an efficient strategy-pair. For example, if increasing the probability of higher outcomes requires more "effort" by the agent, and the agent prefers less effort to more, then if the compensation is independent of the outcome the agent will have no incentive to exert any effort at all! *The incentive requirements for equilibrium, therefore, will typically be incompatible with the conditions for efficiency.*

An exception to the basic proposition occurs if the agent is neutral towards risk and is sufficiently wealthy. (The principal may be risk-neutral or risk-averse.) In this case, an efficient equilibrium is obtained if the principal sells the agent a "franchise" to the enterprise, that is, the agent pays the principal a fixed fee, and then keeps the entire outcome.[7]

Are there any remedies for the inefficiency of equilibrium in the principal-agent relationship? One possible remedy is for the principal to expend resources to monitor the agent's action (and, more generally, his information and environment). Whether this will improve *net* efficiency will depend, of course, on the cost of monitoring. The prevalence of de facto decentralization suggests that accurate and complete monitoring of agents' actions and information in all but the

[7] Up to this point I have been discussing what is called the "static" or "one-period" model. For a thorough treatment of this model with moral hazard, see Sanford S. Grossman and Oliver D. Hart (1983).

smallest firms is too costly to be efficient, or even practicable.

Another remedy for the inefficiency of equilibrium may be available if the principal-agent relationship is a long-term one. The long-term relationship is usually modeled by game theorists as a situation in which the one-period situation is repeated over and over again.[8] These repetitions give the principal an opportunity to observe the results of the agent's actions over a number of periods, and to use some statistical test to infer whether or not the agent was choosing the appropriate action. The repetitions also provide the principal with opportunities to "punish" the agent for apparent departures from the appropriate action. Finally, the fact that the agent's compensation in any one period can be made to depend on the outcomes in a number of previous periods (for example, on the average outcome over a number of periods) provides the principal with an indirect means of insuring the agent, at least partially, against random fluctuations in the outcomes that are not due to fluctuations in the agent's actions. Thus, the repetitions provide an opportunity to reduce the agent's risk without reducing his incentive to perform well.

Although this remedy is in some sense available, game theory does not unequivocally predict that the players will adopt it. I want to explain this statement more carefully, because it is related to a basic problem that game theory has in dealing with long-term relationships. In the repeated game, each player's strategy is a complex object. Thus the principal's strategy is a sequence of decision rules that he uses to determine the agent's compensation in any one period as a function of all of the previous outcomes. Likewise, the agent's strategy is a sequence of decision rules that determines his action in each period as a function of his past observations of the process (outcomes and compensations). In principle, the players' strategy spaces are very large and include very complex strategies.

It may come as no surprise, then, that a *repeated principal-agent game typically has infinitely many equilibria*. One of these equilibria consists of the simple repetition of the (inefficient) one-period equilibrium. Others may be more efficient. Indeed, one can show that, under reasonable conditions, *the less the players discount future utility, the closer the most efficient equilibria will be to full efficiency; in the limit, when they do not discount the future at all, there will be fully efficient equilibria*. On the other hand, game theory itself gives no reason why the players would end up in one equilibrium rather than another.

Unfortunately, I have no space to discuss the game-theoretic treatments of partnerships, except to say that the results are qualitatively similar (although there are some interesting differences): (1) equilibria of the one-period game are typically inefficient, (2) there are many equilibria of the repeated game, some of them more efficient than the one-period equilibrium (or equilibria), and (3) with lower discount rates, more efficient equilibria become available. I should point out that, in addition to moral hazard, another potential source of inefficiency in a partnership is the phenomenon of *free riding* (Radner 1991).

I have thus far said little about misrepresentation, except to mention it as another source of inefficiency. Unfortunately, I shall have no space to discuss this important topic, either. (For an elementary exposition, see Radner 1987; for recent references, see Nahum D. Melumad and Reichelstein 1989.)

It is time to ask: What has game theory

[8] For a rigorous treatment of repeated principal-agent games see Radner (1985, 1986b), and the references cited there.

contributed to our understanding of the economics of managing, and especially the decentralization of incentives? On the positive side, it has illuminated in a rigorous way the sources of the inefficiencies that we can expect to arise with the decentralization of incentives. It also suggests how long-term relationships can be exploited to improve efficiency, especially if the agents in the firm do not discount the future too heavily. Although these results have been derived in very special models, similar results appear in other branches of game theory, and so we could reasonably expect that they are fairly robust.

If we look at the matter in more detail, however, the theoretical analysis remains somewhat disappointing. Principal-agent-like relationships certainly obtain between the shareholders and the board of directors, between the board and the management, and between management and the workers. But each of these groups is not a single player, and the strategic interactions are potentially much richer and more complex than a two person principal agent model can represent.

Moreover, if we look at individuals in the firm, especially in the managing sector, it is rare that we find a person whose output can be realistically measured in money or any other one-dimensional variable. Indeed, it is well known that within the firm it is usually difficult to attribute a definite output to any single member. On an informal level, we can certainly expect to find the pernicious effects of moral hazard, misrepresentation, and free riding in these more complex situations, but we look to game theory to take us beyond the level of informal and intuitive understanding.

I have already alluded to the infinite multiplicity of equilibria that one typically finds in repeated-game models of long-term relationships; one says in these cases that *equilibrium is indeterminate*. This means that game theory does not provide sharp predictions of behavior in these situations. One can imagine elaborating the game theory models to describe processes of learning, adaptation, and evolution, and the corresponding influences of "history," but research of this kind is still in its infancy (David Canning 1989, and Drew Fudenberg and David Kreps 1989).

Finally, we lack a comprehensive model in which we can explore the relative efficiency of different organizational structures. For example, it is not yet possible to compare the efficiency of partnerships and principal-agent relationships because the two models make completely different assumptions about the possibility of imputing outcomes to individuals.

Although, as an economist, I am disappointed in the contribution that game theory has made thus far to the economics of managing, as a practicing game theorist I am not discouraged by the challenge of what remains to be done. But for the time being, game theory has not provided us with an "economic" explanation of the conditions under which we could expect the hierarchical organization of authority and incentives to occur.

7. *Loose Ends*

In this section I will take up some topics that do not yet fit neatly into any economic theory of managing, but probably ought to be taken seriously as we go about developing that theory. The first of these is "coordination," a term so common in writings about organization that you might well wonder why I have waited so long to mention it. According to the *Shorter Oxford English Dictionary*, (p. 1855) one definition of coordination is:

> Harmonious combination of agents or functions towards the production of a result; said espe-

cially in Physiology of the combined action of a number of muscles in the production of certain complex movements.

I suppose "harmonious" in our context would mean "good," or even "optimal," so that coordination would mean "making several decisions that are jointly optimal."

If that were all there were to it—which may well be the case for many authors—then I need not have mentioned the word "coordination" at all. But let me propose two more special, and related, meanings. In a team, where we have a well-defined notion of optimality, there may be several different optimal combinations of decisions. In particular, different assignments of persons to jobs may be equally good. For example, in repairing a leaky dike, it will be important to get the right number of sandbags in the right places, but it may not matter just who fills bags with sand and who carries which bags. As volunteers arrive at the dike to offer their services, there will be a coordinator who assigns them to their respective tasks and locations. A second—and related—interpretation would be applicable to a (noncooperative) game in which there are several equilibria. A coordinator might play the role of persuading the players to focus on a better equilibrium rather than on a worse one. However, here the situation is more complicated. If there are several essentially different Pareto optimal equilibria, then there will be a conflict of interest among the players concerning which equilibrium should be implemented (this is, of course, what I mean by "essentially different"). The job of a coordinator now must include, not only the identification of Pareto optimal equilibria, but also the resolution of the conflict inherent in the choice among them.

I have written as if coordination requires a single person to do it, but that is not generally so. Indeed, under the rubric "mechanism design," game theorists have studied how to design metagames, or how to redesign the original game, so that the equilibria will be Pareto optimal. An example is the Groves-Vickery-Clarke mechanism, which has in particular been applied to the problem of allocating resources to the production of overhead-type activities within the firm.[9]

Mintzberg (1989, p. 101) lists three ways in which work is coordinated in organizations, the third constituting in itself four different methods:

Mutual adjustment.
Direct supervision and authority.
Standardization
  of work processes,
  of outputs,
  of skills and knowledge,
  of norms.

Only the second, direct supervision, would seem to imply hierarchy. The first, mutual adjustment, suggests an equilibrium of a game. The standardization of work, outputs, and skills, suggests decision rules (or strategies or behavior) that are generally expected and accepted, which again has an equilibrium flavor.

In the standardization of norms, we seem to be leaving the confines of economic theory and game theory, in which the tastes and beliefs of the agents are taken as given. In their popular book, *In Search of Excellence*, Thomas J. Peters and Robert H. Waterman (1982) stress the importance of values and "corporate culture." One of the most important tasks of leadership (another popular organization term I have neglected) is said to be the formation and maintenance

[9] See (Theodore Groves and Martin Loeb 1979). For recent contributions to the study of "incentive-compatible" mechanisms, see (Groves, Radner, and Reiter 1987, especially ch. 2). For an elementary exposition in the context of the firm, see (Radner 1986a).

of values in the firm. At the beginning of Chapter 9 they write:

> Let us suppose that we were asked for the one all-purpose bit of advice for management, one truth that we were able to distill from the excellent-companies research. We might be tempted to reply, "Figure out your value system." Decide what your company *stands for*. What does your enterprise do that gives everyone the most pride? Put yourself out ten or twenty years in the future: what would you look back on with greatest satisfaction? (p. 279)

Far from assuming that employees come to the firm with given goals, Peters and Waterman devote their whole Chapter 3 to the heading, "Man Waiting for Motivation;" the "leader" (boss, management) is going to play a crucial role in supplying that motivation. (Lest the reader think that this is just an isolated example of "pop management wisdom," I want to assure you that this theme appears in many management texts and monographs on the sociology of organizations.) However, to pursue this topic in any depth would require a serious look at what has been written about the psychology—and social psychology—of human motivation, and I have neither the space nor the expertise to do that here.

8. *Conclusion*

In summary, I hope that I have persuaded the reader that managing has become a significant activity in the economies of industrial nations. Not only is the activity significant, but large numbers of persons are specialists in managing or in activities that support managing.

The large size of many modern firms, and the limited capacities of individuals for observation, communication, information processing, and decision making are factors that have contributed to this phenomenon. These limitations lead to the decentralization of information processing, and of information, and hence to the de facto decentralization of power and incentives. This is so even in firms that are said to be highly "centralized."

I have sketched some of the contributions of recent research in economic theory to our understanding of the economics of managing, by which I mean the consideration of the resources that go into the activities of managing—especially the human resources—and the ways in which different organizations of managing do a better or worse job of economizing those resources and produce better or worse results.

I described a model of decentralized information processing that incorporates in a limited way some bounds on individual processing capacities. Using this model, I could assess the relative efficiencies of different network structures, including hierarchical ones. Also, within this model, decentralized information processing exhibits decreasing returns to scale.

I also described Herbert Simon's suggestive discussion of the hierarchical organization of work, and I sketched some implications of such an organization for the efficient decentralization of information. Here further research is needed to integrate the analyses of the decentralization of information processing and the decentralization of information.

The theoretical analyses of principal-agent and partnership models, especially in their dynamic versions, have given us some rigorous insights into the loss of control and efficiency due to the decentralization of incentives, but I had to admit to disappointment in the progress that has been made thus far. In particular, game-theoretic treatments typically leave us with a serious multiplicity of equilibria, and hence with indeterminate predictions of behavior. Also, there seem to be no substantial analyses of comprehensive models in which we can explore—from the incentive point of view—the relative efficiency of different structures of managing.

With regard to hierarchy itself, what have we learned from economic theory? The most positive and definite result was the one about the efficiency of hierarchies for information processing (Section 4). But this is precisely the area in which we know, empirically, from the study of "informal organization," that firms are *not* predominantly hierarchical. At least, if there are hierarchies embedded in the patterns of information processing, there are many overlapping ones; *in fact, this is predicted by the theory of the decentralization of information* (Section 5).

On the other hand, in the area of incentives where hierarchies of authority and supervision seem most evident in reality, economic theorists—and their colleagues, the game theorists—have not provided an incisive comparative analysis of both hierarchical and nonhierarchical forms of organization.

Now the absence of a comparative analysis is not the same as a negative analysis that shows that there is no economic justification for hierarchies of authority. However, it does tempt one to speculate about noneconomic—or partly noneconomic—explanations of such hierarchies.

One possible explanation is both economic and historical. Small owner-managed firms were authoritarian because the owners wanted to maintain control over their assets. As the firm grew, the authoritarian pattern was repeated at successive levels as managers were added who were not themselves owners, perhaps just in imitation of the pattern that was already there. However, if nonhierarchical management were in fact more efficient than hierarchical management, then why weren't such structures eventually adopted *within* the management sector, while maintaining the authority of the owners as a group over the management as a group?

In fact, there has been growing criticism of purely hierarchical structures of authority in firms, not only from the left, but also mainstream analysts, consultants, and managers themselves. One example is the "matrix management" movement, in which a single hierarchy of authority is replaced by two or more overlapping—and sometimes conflicting—hierarchies (Stanley M. Davis and Paul R. Lawrence 1977). A second example is the recent spurt of interest in "flat organizations," which—although formally hierarchical—have very few levels, and give each supervisor so many subordinates that effective authoritarian control becomes impossible (Peter Drucker 1988; Allan Cox 1989). A third line of thought asserts that activities in large modern firms have become so interdependent—whatever that means precisely—that de facto hierarchies of power no longer exist, and that power is exercised by mutual adjustment, persuasion, exhortation, and leadership, rather than by virtue of office alone. For an exciting discussion of organization structures that are neither hierarchies nor neoclassical markets—and the challenge to economic theory that they represent—I recommend two recent papers by Michael Piore (1989a, 1989b) on research that he has been doing for the International Labor Organization.

Going farther afield, one might find the roots of authoritarian hierarchies in human personality itself, as described by Theodor Adorno et al (1950) in *The Authoritarian Personality*, and by Erich Fromm (1941) in *Escape from Freedom.* After all, we grow up in authoritarian families, and many aspects of our society recreate and perpetuate these hierarchical structures. From this point of view, it should not be surprising that most employees feel most comfortable in hierarchical firms.

All of this takes me far from my own field of expertise, and also far from the comfortable haunts of mainstream economic analysis and game theory. But rather than withdraw to safer ground, I am going to leave us in these dangerous precincts, in the hope that some of the more daring—and probably younger—theorists will be tempted to venture out from the citadel of general market equilibrium, and try to build some solid structures here, too.

References

In addition to works cited in the text, the following list of references includes items that may be useful for further reading.

Abreu, Dilip and Ariel, Rubinstein. "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, Nov. 1988, *56*(6), pp. 1259–82.

Adorno, Theodor W. et al, *The authoritarian personality.* NY: Harper and Row, 1950.

Alchian, Armen A. and Harold, Demsetz. "Production, Information Costs, and Economic Organization," *Amer. Econ. Rev.*, Dec. 1972, *62*(5), pp. 777–95.

Aoki, Masahiko. "Toward an Economic Model of the Japanese Firm," *J. Econ. Lit.*, Mar. 1990, *28*(1), pp. 1–17.

Arrow, Kenneth J. "Uncertainty and the Welfare Economics of Medical Care," *Amer. Econ. Rev.*, Dec. 1963, *53*(5), pp. 941–73.

———. *The limits of organization.* NY: Norton, 1974.

Barnard, Chester I. *The functions of the executive.* Cambridge, MA: Harvard U. Press, 1938.

Beckmann, Martin J. *Tinbergen lectures on organization theory.* Berlin: Springer-Verlag, 1983.

Buttrick, John. "The Inside Contract System," *J. Econ. Hist.*, 1952, *12*(3), pp. 205–21.

Canning, David. "Social Equilibrium." Unpub. ms. Pembroke College, Cambridge, 1989.

Chandler, Alfred D., Jr. *The visible hand.* Cambridge, MA: Harvard U. Press, 1977.

Clark, John M. "Economics and Modern Psychology," *J. Polit. Econ.*, Jan. 1918, *26*(1), pp. 1–30.

Cox, Allan. "Managing without Hierarchy: Even 'Flat' Companies Need Leaders," *New York Times*, Aug. 20, 1989, *138*, Sect. 3. p. F3.

Crémer, Jacques. "A Partial Theory of the Optimal Organization of a Bureaucracy," *Bell. J. Econ.*, Autumn 1980, *11*(2), pp. 683–93.

Davis, Stanley M. and Lawrence, Paul R. *Matrix.* Reading, MA: Addison-Wesley, 1977.

Drucker, Peter E. "The Coming of the New Organization," *Harvard Bus. Rev.*, Jan./Feb. 1988, *66*(1), pp. 45–53.

"The Fortune 500," *Fortune*, Apr. 24, 1989, *119*(9), pp. 354–76.

"The Fortune International 500," *Fortune,* July 31, 1989, *120*(3), pp. 279–318.

Fromm, Erich. *Escape from freedom.* NY: Rinehart, 1941.

Fudenberg, Drew and Kreps, David. "A Theory of Learning, Experimentation, and Equilibrium in Games." Unpub. ms. Economics Dept., Massachusetts Institute of Technology, Cambridge, MA, 1989.

Galbraith, Jay. *Designing complex organizations.* Reading, MA: Addison-Wesley, 1973.

Geanakoplos, John and Milgrom, Paul. "A Theory of Hierarchies Based on Limited Managerial Attention," *J. Japan. Int. Economies*, Sept. 1991, *5*(3), pp. 205–25.

Gilboa, Itzhak and Samet, Dov. "Bounded Versus Unbounded Rationality: The Tyranny of the Weak," *Games and Economic Beahvior,* Sept. 1989, *1*(30, pp. 213–21.

Grossman, Sanford J. and Hart, Oliver D. "An Analysis of the Principal-Agent Problem," *Econometrica*, Jan. 1983, *51*(1), pp. 7–45.

Groves, Theodore and Loeb, Martin. "Incentives in a Divisionalized Firm," *Manage. Sci.*, Mar. 1979, *25*(3), pp. 221–30.

Groves, Theodore; Radner, Roy and Reiter, Stanley, *Information incentives, and economic mechanisms: Essays in honor of Lenoid Hurwicz.* Minneapolis, MN: U. of Minnesota Press, 1987.

Harsanyi, John C. "Games with Incomplete Information Played by Bayesian Players," *Manage. Sci.*, 1967–68, *14*, pp. 159–82, 320–34, 486–502.

Hart, Oliver. "An Economist's Perspective on the Theory of the Firm," *Columbia Law Rev.*, Nov. 1989, *89*(7), pp. 1757–74.

Hess, James D. *The economics of organization.* Amsterdam: North-Holland, 1983.

Kalai, Ehud and Stanford, William. "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, Mar. 1988, *56*(2), pp. 397–410.

Keren, Michael and Levhari, David. "The Internal Organization of the Firm and the Shape of Average Costs," *Bell. J. Econ.*, Autumn 1983, *14*(2), pp. 474–86.

———. "Decentralization, Aggregation, Control Loss, and Costs in a Hierarchical Model of the Firm," *J. Econ. Behav. Org.*, Mar. 1989, *11*(2), pp. 213–36.

Kim, Ki H. and Roush, Fred W. *Team theory.* NY: Halsted Press, 1987.

Knight, Frank H. *Risk, uncertainty, and profits.* Cambridge, MA: Houghton Mifflin, 1921; reprinted by Kelley and Millman, New York, 1957.

Kruskal, Clyde P. et al. "A Complexity Theory of Efficient Parallel Algorithms," *Theoretical Computer Science*, Mar. 1990, *71*(1), pp. 95–132.

Linhart, Peter B.; Radner, Roy and Sattertwaite, Mark A. "Introduction: Symposium on Noncooperative Bargaining," *J. Econ. Theory*, June 1989, *48*(1), pp. 1–17.

Linhart, Peter B. et al. *Bargaining under incomplete information.* NY: Academic Press, 1992.

Marris, Robin. *The economic theory of managerial capitalism.* NY: Macmillan, 1964.

Marris, Robin and Wood, Adrian. *The corporate* *economy*. Cambridge, MA: Harvard U. Press, 1971.

MARSCHAK, JACOB AND RADNER, ROY. *Economic theory of teams*. New Haven, CT: Yale U. Press, 1972.

MARSCHAK, THOMAS A. "Computation in Organizations," in MCGUIRE AND RADNER 1986, ch. 12. pp. 237–82.

MARSCHAK, THOMAS A. AND MCGUIRE, C. BARTLETT. "Lecture Notes on Economic Models for Organization Design." Unpub. ms. Graduate School of Business Administration, U. of California, Berkeley, 1971.

MARSCHAK, THOMAS A. AND REICHELSTEIN, STEFAN. "Network Mechanisms, Informational Efficiencies and the Role of Hierarchies." Unpub. ms. Graduate School of Business Administration, Stanford U., Dec. 1987.

MARSHALL, ALFRED, *Principles of economics*. London: Macmillan, [1890] 1930.

MCCRAW, THOMAS K. *Prophets of regulation*. Cambridge, MA: Harvard U. Press, 1984.

MCGUIRE, C. BARTLETT AND RADNER, ROY. *Decision and organization*. 2nd. ed. Minneapolis, MN: U. of Minnesota Press, 1986.

MELUMAD, NAHUM D. AND REICHELSTEIN, STEFAN. "Value of Communication in Agencies," *J. Econ. Theory*, Apr. 1989, *47*(2), pp. 334–68.

MILGROM, PAUL AND ROBERTS, JOHN. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *Amer. Econ. Rev.*, June 1990, *80*(3), pp. 511–28.

MINTZBERG, HENRY. *The structuring of organizations*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

———. *Mintzberg on management*. NY: Free Press, 1989.

MIRRLEES, JAMES. "The Optimal Structure of Incentives and Authority within an Organization," *Bell J. Econ.*, Spring 1976, *7*(1), pp. 105–31.

MOUNT, KENNETH R. AND REITER, STANLEY. "Computation, Communciation, and Performance in Resource Allocation." Paper presented at the CEME-NBER Decentralization Seminar, U. of Minnesota, Minneapolis, May 21–23, 1982.

———. " A Model of Computing with Human Agents." Unpub. ms. Center for Mathematical Studies in Economics and Management Science, Northwestern U., Evanston, IL, June 1990.

MYERSON, ROGER B. "Bayesian Equilibrium and Incentive Compatibility: An Introduction," in *Social goals and social organization*. Eds.: LEONID HURWICZ, DAVID SCHMEIDLER, AND HUGO SONNENSCHEIN. Cambridge: Cambridge U. Press, 1985, pp. 229–59.

NEYMAN, ABRAHAM. "Bounded Complexity Justifies Cooperation in the Finitely-Repeated Prisoners' Dilemma," *Economics Letters*, 1985, *19*, pp. 227–29.

PELIKAN, PAVEL. "Evolution, Economic Competence, and the Market for Corporate Control," *J. Econ. Behav. Organ.*, Dec. 1989, *12*(3), pp. 279–303.

PENZIAS, ARNO A. *Ideas and information*. NY: Norton, 1989.

PETERS, THOMAS J. AND WATERMAN, ROBERT H. JR. *In search of excellence*. NY: Harper & Row, 1982.

PIORE, MICHAEL J. "Corporate Reform in American Manufacturing and the Challenge to Economic Theory." Unpub. ms. Massachusetts Institute of Technology, Cambridge, MA, 1989b.

———. "Work, Labor, and Action: Work Experience in a System of Flexible Production." Unpub. ms. Massachusetts Institute of Technology, Cambridge, MA, 1989b.

PORTER, GLENN. *The rise of big business, 1860–1910*. NY: Crowell, 1973.

PRATT, JOHN W. AND ZECKHAUSER, RICHARD J., eds. *Principals and agents: The structure of business*. Boston: Harvard Business School Press, 1985.

RADNER, ROY. "Repeated Principal-Agent Games with Discounting," *Econometrica*, Sept. 1985, *53*(5), pp. 1173–97.

———. "Teams," in McGuire and Radner 1986, ch. 10, pp. 189–215.

———. "The Internal Economy of Large Firms," *Econ. J.*, 1986a, *96* (Supplement), pp. 1–22.

———. "Repeated Moral Hazard with Low Discount Rates," in *Uncertainty, information, and communication: Essays in honor of Kenneth J. Arrow*. Vol. 3. Eds: WALTER P. HELLER, ROSS M. STARR, AND DAVID A. STARRETT. Cambridge, MA: Cambridge U. Press, 1986b, pp. 25–63.

———. "Decentralization and Incentives," in GROVES, RADNER, AND REITER 1987, ch. 1, pp. 3–47.

———. "The Organization of Decentralized Information Processing." Unpub. ms. AT&T Bell Laboratories, Murray Hill, NJ, 1989.

———. "Dynamic Games in Organization Theory," *J. Econ. Behav. Organ.*, July 1991, *16*(1–2), pp. 217–60.

RADNER, ROY AND VAN ZANDT, TIMOTHY. "Information Processing in Firms and Returns to Scale," *Annales d'Economie et de Statistique*, June 1992, 25/26, pp. 265–98.

ROBINSON, JOAN. "What Are the Questions?," *J. Econ. Lit.*, Dec. 1977, *15*(4), pp. 1318–39.

RUBINSTEIN, ARIEL. "Finite Automata Play the Repeated Prisoners' Dilemma," *J. Econ. Theory*, June 1986, *39*(1) pp. 83–96.

SAH, RAAJ K. AND STIGLITZ, JOSEPH E., "The Architecture of Economic Systems: Hierarchies and Polyarchies," *Amer. Econ. Rev.*, Sept. 1986, *76*(4), 716–27.

SAVAGE, LEONARD J. *The foundations of statistics*. NY: Wiley, 1954.

SCHWARTZ, JACOB T. "Ultracomputers," *ACM Transactions on Programming Languages and Systems*, Oct. 1980, *2*(4), pp. 484–521.

SIMON, HERBERT A. "The Architecture of Complexity," *Proc. Amer. Philosophical Soc.*, Dec. 1962, *106*(6), pp. 467–82; reprinted as Ch. 7 of SIMON 1981, pp. 192–229.

———. *The sciences of the artificial*. 2nd ed. Cambridge, MA: MIT Press, 1981.

SMITH, ADAM. *The wealth of nations*. 1776; reprinted by the U. of Chicago Press, Chicago, IL, 1976 (Cannan Edition).

SOLOMONS, DAVID. *Divisional performance: Mea-*

surement and control. NY: Markus Wiener, 1965.

TANNENBAUM, ARNOLD S. *Social psychology of the work organization.* Belmont, CA: Wadsworth, 1966.

U.S. BUREAU OF THE CENSUS. *Sixteenth census of the United States: 1940. Population. Comparative occupation statistics for the United States, 1870–1940. A comparison of the 1930 and the 1940 census occupation and industry classifications and statistics; a comparable series of occupation statistics, 1870 to 1930; and a social-economic grouping of the labor force, 1910 to 1940 . . .* Washington, DC: U.S. Dept. of Commerce, 1943.

———. *Historical statistics of the United States: Colonial times to 1970, bicentennial edition, Part 2,* Washington, DC: U.S. Dept. of Commerce, 1975.

U.S. BUREAU OF THE CENSUS. *1980 census of population.* Vol. 1. *Characteristics of the population.* Washington, DC: U.S. Dept. of Commerce, 1981.

U.S. DEPT. OF COMMERCE. *Statistical abstracts of the U.S.* Washington, DC, 1989.

VAN ZANDT, TIMOTHY. "Efficient Parallel Addition." Unpub. ms. AT&T Bell Laboratories, Murray Hill, NJ, 1990.

WILLIAMSON, OLIVER E. *Markets and hierarchies.* NY: Free Press, 1975.

WITSENHAUSEN, HANS S. "A Counterexample in Stochastic Optimum Control," *SIAM J. Control,* Feb. 1968, *6*(1), pp. 131–47.

Hierarchy: The Economics of Managing
Roy Radner
Journal of Economic Literature, Vol. 30, No. 3. (Sep., 1992), pp. 1382-1415.
Stable URL:
http://links.jstor.org/sici?sici=0022-0515%28199209%2930%3A3%3C1382%3AHTEOM%3E2.0.CO%3B2-L

[Footnotes]

[7] **An Analysis of the Principal-Agent Problem**
Sanford J. Grossman; Oliver D. Hart
Econometrica, Vol. 51, No. 1. (Jan., 1983), pp. 7-45.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198301%2951%3A1%3C7%3AAAOTPP%3E2.0.CO%3B2-A

[8] **Repeated Principal-Agent Games with Discounting**
Roy Radner
Econometrica, Vol. 53, No. 5. (Sep., 1985), pp. 1173-1198.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198509%2953%3A5%3C1173%3ARPGWD%3E2.0.CO%3B2-M

[9] **Incentives in a Divisionalized Firm**
Theodore Groves; Martin Loeb
Management Science, Vol. 25, No. 3. (Mar., 1979), pp. 221-230.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28197903%2925%3A3%3C221%3AIIADF%3E2.0.CO%3B2-D

[9] **The Internal Economy of Large Firms**
Roy Radner
The Economic Journal, Vol. 96, Supplement: Conference Papers. (1986), pp. 1-22.
Stable URL:
http://links.jstor.org/sici?sici=0013-0133%281986%2996%3C1%3ATIEOLF%3E2.0.CO%3B2-D

LINKED CITATIONS

References

The Structure of Nash Equilibrium in Repeated Games with Finite Automata
Dilip Abreu; Ariel Rubinstein
Econometrica, Vol. 56, No. 6. (Nov., 1988), pp. 1259-1281.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198811%2956%3A6%3C1259%3ATSONEI%3E2.0.CO%3B2-X

Production, Information Costs, and Economic Organization
Armen A. Alchian; Harold Demsetz
The American Economic Review, Vol. 62, No. 5. (Dec., 1972), pp. 777-795.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28197212%2962%3A5%3C777%3APICAEO%3E2.0.CO%3B2-Q

Toward an Economic Model of the Japanese Firm
Masahiko Aoki
Journal of Economic Literature, Vol. 28, No. 1. (Mar., 1990), pp. 1-27.
Stable URL:
http://links.jstor.org/sici?sici=0022-0515%28199003%2928%3A1%3C1%3ATAEMOT%3E2.0.CO%3B2-T

Uncertainty and the Welfare Economics of Medical Care
Kenneth J. Arrow
The American Economic Review, Vol. 53, No. 5. (Dec., 1963), pp. 941-973.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28196312%2953%3A5%3C941%3AUATWEO%3E2.0.CO%3B2-C

Economics and Modern Psycholoy: I
J. M. Clark
The Journal of Political Economy, Vol. 26, No. 1. (Jan., 1918), pp. 1-30.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28191801%2926%3A1%3C1%3AEAMPI%3E2.0.CO%3B2-F

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

An Analysis of the Principal-Agent Problem
Sanford J. Grossman; Oliver D. Hart
Econometrica, Vol. 51, No. 1. (Jan., 1983), pp. 7-45.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198301%2951%3A1%3C7%3AAAOTPP%3E2.0.CO%3B2-A

Incentives in a Divisionalized Firm
Theodore Groves; Martin Loeb
Management Science, Vol. 25, No. 3. (Mar., 1979), pp. 221-230.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28197903%2925%3A3%3C221%3AIIADF%3E2.0.CO%3B2-D

Games with Incomplete Information Played by "Bayesian" Players, I-III. Part I. The Basic Model
John C. Harsanyi
Management Science, Vol. 14, No. 3, Theory Series. (Nov., 1967), pp. 159-182.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28196711%2914%3A3%3C159%3AGWIIPB%3E2.0.CO%3B2-P

Finite Rationality and Interpersonal Complexity in Repeated Games
Ehud Kalai; William Stanford
Econometrica, Vol. 56, No. 2. (Mar., 1988), pp. 397-410.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198803%2956%3A2%3C397%3AFRAICI%3E2.0.CO%3B2-F

The Economics of Modern Manufacturing: Technology, Strategy, and Organization
Paul Milgrom; John Roberts
The American Economic Review, Vol. 80, No. 3. (Jun., 1990), pp. 511-528.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28199006%2980%3A3%3C511%3ATEOMMT%3E2.0.CO%3B2-H

Repeated Principal-Agent Games with Discounting
Roy Radner
Econometrica, Vol. 53, No. 5. (Sep., 1985), pp. 1173-1198.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198509%2953%3A5%3C1173%3ARPGWD%3E2.0.CO%3B2-M

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

The Internal Economy of Large Firms
Roy Radner
The Economic Journal, Vol. 96, Supplement: Conference Papers. (1986), pp. 1-22.
Stable URL:
http://links.jstor.org/sici?sici=0013-0133%281986%2996%3C1%3ATIEOLF%3E2.0.CO%3B2-D

What Are the Questions?
Joan Robinson
Journal of Economic Literature, Vol. 15, No. 4. (Dec., 1977), pp. 1318-1339.
Stable URL:
http://links.jstor.org/sici?sici=0022-0515%28197712%2915%3A4%3C1318%3AWATQ%3E2.0.CO%3B2-P

The Architecture of Economic Systems: Hierarchies and Polyarchies
Raaj Kumar Sah; Joseph E. Stiglitz
The American Economic Review, Vol. 76, No. 4. (Sep., 1986), pp. 716-727.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28198609%2976%3A4%3C716%3ATAOESH%3E2.0.CO%3B2-9

NOTE: *The reference numbering from the original has been maintained in this citation list.*