Production, Information Costs, and Economic Organization

By ARMEN A. ALCHIAN AND HAROLD DEMSETZ*

The mark of a capitalistic society is that resources are owned and allocated by such nongovernmental organizations as firms, households, and markets. Resource owners increase productivity through cooperative specialization and this leads to the demand for economic organizations which facilitate cooperation. When a lumber mill employs a cabinetmaker, cooperation between specialists is achieved within a firm, and when a cabinetmaker purchases wood from a lumberman, the cooperation takes place across markets (or between firms). Two important problems face a theory of economic organization—to explain the conditions that determine whether the gains from specialization and cooperative production can better be obtained within an organization like the firm, or across markets, and to explain the structure of the organization.

It is common to see the firm characterized by the power to settle issues by fiat, by authority, or by disciplinary action superior to that available in the conventional market. This is delusion. The firm does not own all its inputs. It has no power of fiat, no authority, no disciplinary action any different in the slightest degree from ordinary market contracting between any two people. I can "punish" you only by withholding future business or by seeking redress in the courts for any failure to honor our exchange agreement. That is exactly all that any employer can do. He can fire or sue, just as I can fire my grocer by stopping purchases from him or sue him for delivering faulty products. What then is the content of the presumed power to manage and assign workers to various tasks? Exactly the same as one little consumer's power to manage and assign his grocer to various tasks. The single consumer can assign his grocer to the task of obtaining whatever the customer can induce the grocer to provide at a price acceptable to both parties. That is precisely all that an employer can do to an employee. To speak of managing, directing, or assigning workers to various tasks is a deceptive way of noting that the employer continually is involved in renegotiation of contracts on terms that must be acceptable to both parties. Telling an employee to type this letter rather than to file that document is like my telling a grocer to sell me this brand of tuna rather than that brand of bread. I have no contract to continue to purchase from the grocer and neither the employer nor the employee is bound by any contractual obligations to continue their relationship. Long-term contracts between employer and employee are not the essence of the organization we call a firm. My grocer can count on my returning day after day and purchasing his services and goods even with the prices not always marked on the goods —because I know what they are—and he adapts his activity to conform to my directions to him as to what I want each day . . . he is not my employee.

Wherein then is the relationship between a grocer and his employee different from that between a grocer and his cus-

* Professors of economics at the University of California, Los Angeles. Acknowledgment is made for financial aid from the E. Lilly Endowment, Inc. grant to UCLA for research in the behavioral effects of property rights.

tomers? It is in a *team* use of inputs and a centralized position of some party in the contractual arrangements of *all* other inputs. It is the *centralized contractual agent in a team productive process*—not some superior authoritarian directive or disciplinary power. Exactly what is a team process and why does it induce the contractual form, called the firm? These problems motivate the inquiry of this paper.

I. The Metering Problem

The economic organization through which input owners cooperate will make better use of their comparative advantages to the extent that it facilitates the payment of rewards in accord with productivity. If rewards were random, and without regard to productive effort, no incentive to productive effort would be provided by the organization; and if rewards were negatively correlated with productivity the organization would be subject to sabotage. Two key demands are placed on an economic organization—metering input productivity and metering rewards.[1]

Metering problems sometimes can be resolved well through the exchange of products across competitive markets, because in many situations markets yield a high correlation between rewards and productivity. If a farmer increases his output of wheat by 10 percent at the prevailing market price, his receipts also increase by 10 percent. This method of organizing economic activity meters the *output directly*, reveals the marginal product and apportions the *rewards* to resource owners in accord with that direct measurement of their outputs. The success of this decentralized, market exchange in promoting productive specialization requires that changes in market rewards fall on those responsible for changes in *output*.[2]

The classic relationship in economics that runs from marginal productivity to the distribution of income implicitly *assumes* the existence of an organization, be it the market or the firm, that allocates rewards to resources in accord with their productivity. The problem of economic organization, the economical means of metering productivity and rewards, is not confronted directly in the classical analysis of production and distribution. Instead, that analysis tends to assume sufficiently economic—or zero cost—means, as if productivity automatically created its reward. We conjecture the direction of causation is the reverse—the specific sys-

[1] Meter means to measure and also to apportion. One can meter (measure) output and one can also meter (control) the output. We use the word to denote both; the context should indicate which.

[2] A producer's wealth would be reduced by the present capitalized value of the future income lost by loss of reputation. Reputation, i.e., credibility, is an asset, which is another way of saying that reliable information about expected performance is both a costly and a valuable good. For acts of God that interfere with contract performance, both parties have incentives to reach a settlement akin to that which would have been reached if such events had been covered by specific contingency clauses. The reason, again, is that a reputation for "honest" dealings—i.e., for actions similar to those that would probably have been reached had the contract provided this contingency—is wealth.

Almost every contract is open-ended in that many contingencies are uncovered. For example, if a fire delays production of a promised product by A to B, and if B contends that A has not fulfilled the contract, how is the dispute settled and what recompense, if any, does A grant to B? A person uninitiated in such questions may be surprised by the extent to which contracts permit either party to escape performance or to nullify the contract. In fact, it is hard to imagine any contract, which, when taken solely in terms of its stipulations, could not be evaded by one of the parties. Yet that is the ruling, viable type of contract. Why? Undoubtedly the best discussion that we have seen on this question is by Stewart Macaulay.

There are means not only of detecting or preventing cheating, but also for deciding how to allocate the losses or gains of unpredictable events or quality of items exchanged. Sales contracts contain warranties, guarantees, collateral, return privileges and penalty clauses for specific nonperformance. These are means of assignment of *risks* of losses of cheating. A lower price without warranty—an "as is" purchase—places more of the risk on the buyer while the seller buys insurance against losses of his "cheating." On the other hand, a warranty or return privilege or service contract places more risk on the seller with insurance being bought by the buyer.

tem of rewarding which is relied upon stimulates a particular productivity response. If the economic organization meters poorly, with rewards and productivity only loosely correlated, then productivity will be smaller; but if the economic organization meters well productivity will be greater. What makes metering difficult and hence induces means of economizing on metering costs?

II. Team Production

Two men jointly lift heavy cargo into trucks. Solely by observing the total weight loaded per day, it is impossible to determine each person's marginal productivity. With team production it is difficult, solely by observing total output, to either define or determine *each* individual's contribution to this output of the cooperating inputs. The output is yielded by a team, by definition, and it is not a *sum* of separable outputs of each of its members. Team production of Z involves at least two inputs, X_i and X_j, with $\partial^2 Z/\partial X_i \partial X_j \neq 0$.[3] The production function is *not* separable into two functions each involving only inputs X_i or only inputs X_j. Consequently there is no *sum* of Z of two separable functions to treat as the Z of the team production function. (An example of a *separable* case is $Z = aX_i^2 + bX_j^2$ which is separable into $Z_i = aX_i^2$ and $Z_j = bX_j^2$, and $Z = Z_i + Z_j$. This is not team production.) There exist production techniques in which the Z obtained is greater than if X_i and X_j had produced separable Z. Team production will be used if it yields an output enough larger than the sum of separable production of Z to cover the costs of organizing and disciplining team members—the topics of this paper.[4]

[3] The function is separable into additive functions if the cross partial derivative is zero, i.e., if $\partial^2 Z/\partial X_i \partial X_j = 0$.

[4] With sufficient generality of notation and conception this team production function could be formulated as a case of the generalized production function interpretation given by our colleague, E. A. Thompson.

Usual explanations of the gains from cooperative behavior rely on exchange and production in accord with the comparative advantage specialization principle with separable additive production. However, as suggested above there is a source of gain from cooperative activity involving working as a *team*, wherein individual cooperating inputs do not yield identifiable, separate products which can be *summed* to measure the total output. For this cooperative productive activity, here called "team" production, measuring *marginal* productivity and making payments in accord therewith is more expensive by an order of magnitude than for separable production functions.

Team production, to repeat, is production in which 1) several types of resources are used and 2) the product is not a sum of separable outputs of each cooperating resource. An additional factor creates a team organization problem—3) not all resources used in team production belong to one person.

We do not inquire into why all the jointly used resources are not owned by one person, but instead into the types of organization, contracts, and informational and payment procedures used among owners of teamed inputs. With respect to the one-owner case, perhaps it is sufficient merely to note that (a) slavery is prohibited, (b) one might assume risk aversion as a reason for one person's not borrowing enough to purchase all the assets or sources of services rather than renting them, and (c) the purchase-resale spread may be so large that costs of short-term ownership exceed rental costs. Our problem is viewed basically as one of organization among different people, not of the physical goods or services, however much there must be selection and choice of combination of the latter.

How can the members of a team be rewarded and induced to work efficiently?

In team production, marginal products of cooperative team members are not so directly and separably (i.e., cheaply) observable. What a team offers to the market can be taken as the marginal product of the team but not of the team members. The costs of metering or ascertaining the marginal products of the team's members is what calls forth new organizations and procedures. Clues to each input's productivity can be secured by observing *behavior* of individual inputs. When lifting cargo into the truck, how rapidly does a man move to the next piece to be loaded, how many cigarette breaks does he take, does the item being lifted tilt downward toward his side?

If detecting such behavior were costless, neither party would have an incentive to shirk, because neither could impose the cost of his shirking on the other (if their cooperation was agreed to voluntarily). But since costs must be incurred to monitor each other, each input owner will have more incentive to shirk when he works as part of a team, than if his performance could be monitored easily or if he did not work as a team. If there is a net increase in productivity available by team production, net of the metering cost associated with disciplining the team, then team production will be relied upon rather than a multitude of bilateral exchange of separable individual outputs.

Both leisure and higher income enter a person's utility function.[5] Hence, each person should adjust his work and realized reward so as to equate the marginal rate of substitution between leisure and production of real output to his marginal rate of substitution in consumption. That is, he would adjust his rate of work to bring his demand prices of leisure and output to equality with their true costs. However, with detection, policing, monitoring, measuring or metering costs, each person will be induced to take more leisure, because the effect of relaxing on *his realized* (reward) rate of substitution between output and leisure will be less than the effect on the *true* rate of substitution. His realized cost of leisure will fall more than the true cost of leisure, so he "buys" more leisure (i.e., more nonpecuniary reward).

If his relaxation cannot be detected perfectly at zero cost, part of its effects will be borne by others in the team, thus making *his* realized cost of relaxation less than the true total cost to the team. The difficulty of detecting such actions permits the private costs of his actions to be less than their full costs. Since each person responds to his private realizable rate of substitution (in production) rather than the true total (i.e., social) rate, and so long as there are costs for other people to detect his shift toward relaxation, it will not pay (them) to force him to readjust completely by making him realize the true cost. Only enough efforts will be made to equate the marginal gains of detection activity with the marginal costs of detection; and that implies a lower rate of productive effort and more shirking than in a costless monitoring, or measuring, world.

In a university, the faculty use office telephones, paper, and mail for personal uses beyond strict university productivity. The university administrators could stop such practices by identifying *the* responsible person in each case, but they can do so only at higher costs than administrators are willing to incur. The extra costs of identifying each party (rather than merely identifying the presence of such activity) would exceed the savings from diminished faculty "turpitudinal peccadilloes." So the faculty is allowed some degree of "privileges, perquisites, or fringe benefits." And the total of the pecuniary wages paid

[5] More precisely: "if anything other than pecuniary income enters his utility function." Leisure stands for all nonpecuniary income for simplicity of exposition.

is lower because of this irreducible (at acceptable costs) degree of amenity-seizing activity. Pay is lower in pecuniary terms and higher in leisure, conveniences, and ease of work. But still every person would prefer to see detection made more effective (if it were somehow possible to monitor costlessly) so that he, as part of the now more effectively producing team, could thereby realize a higher pecuniary pay and less leisure. If everyone could, at zero cost, have his reward-realized rate brought to the true production possibility real rate, all could achieve a more preferred position. But detection of the responsible parties is costly; that cost acts like a tax on work rewards.[6] Viable shirking is the result.

What forms of organizing team production will lower the cost of detecting "performance" (i.e., marginal productivity) and bring personally realized rates of substitution closer to true rates of substitution? Market competition, in principle, could monitor some team production. (It already *organizes* teams.) Input owners who are not team members can offer, in return for a smaller share of the team's rewards, to replace excessively (i.e., overpaid) shirking members. Market competition among potential team members would determine team membership and individual rewards. There would be no team leader, manager, organizer, owner, or employer. For such decentralized organizational control to work, outsiders, possibly after observing each team's total output, can speculate about their capabilities as team members and, by a market competitive process, revised teams with greater productive ability will be formed and sustained. Incumbent members will be constrained by threats of replacement by outsiders offering services for lower reward shares or offering greater rewards to the other members of the team. Any team member who shirked in the expectation that the reduced output effect would not be attributed to him will be displaced if his activity is detected. Teams of productive inputs, like business units, would evolve in apparent spontaneity in the market—without any central organizing agent, team manager, or boss.

But completely effective control cannot be expected from individualized market competition for two reasons. First, for this competition to be completely effective, new challengers for team membership must know where, and to what extent, shirking is a serious problem, i.e., know they can increase net output as compared with the inputs they replace. To the extent that this is true it is probably possible for existing fellow team members to recognize the shirking. But, by definition, the detection of shirking by observing team output is costly for team production. Secondly, assume the presence of detection costs, and assume that in order to secure a place on the team a new input owner must accept a smaller share of rewards (or a promise to produce more). Then his incentive to shirk would still be at least as great as the incentives of the inputs replaced, because he still bears less than the entire reduction in team output for which he is responsible.

III. The Classical Firm

One method of reducing shirking is for someone to specialize as a monitor to check the input performance of team members.[7]

[6] Do not assume that the sole result of the cost of detecting shirking is one form of payment (more leisure and less take home money). With several members of the team, each has an incentive to cheat against each other by engaging in more than the average amount of such leisure if the employer can not tell at zero cost which employee is taking more than average. As a result the total productivity of the team is lowered. Shirking detection costs thus change the form of payment and also result in lower total rewards. Because the cross partial derivatives are positive, shirking reduces other people's marginal products.

[7] What is meant by performance? Input energy, initiative, work attitude, perspiration, rate of exhaustion?

(*Continued*)

But who will monitor the monitor? One constraint on the monitor is the aforesaid market competition offered by other monitors, but for reasons already given, that is not perfectly effective. Another constraint can be imposed on the monitor: give him title to the net earnings of the team, net of payments to other inputs. If owners of cooperating inputs agree with the monitor that he is to receive any residual product above prescribed amounts (hopefully, the marginal value products of the other inputs), the monitor will have an added incentive not to shirk as a monitor. Specialization in monitoring plus reliance on a residual claimant status will reduce shirking; but additional links are needed to forge the firm of classical economic theory. How will the residual claimant monitor the other inputs?

We use the term monitor to connote several activities in addition to its disciplinary connotation. It connotes measuring output performance, apportioning rewards, observing the input behavior of inputs as means of detecting or estimating their marginal productivity and giving assignments or instructions in what to do and how to do it. (It also includes, as we shall show later, authority to terminate or revise contracts.) Perhaps the contrast between a football coach and team captain is helpful. The coach selects strategies and tactics and sends in instructions about what plays to utilize. The captain is essentially an observer and reporter of the performance at close hand of the members. The latter is an inspector-steward and the former a supervisor manager. For the present all these activities are included in the rubric "monitoring." All these tasks are, in principle, negotiable across markets, but we are presuming that such market measurement of marginal productivities and job reassignments are not so cheaply performed for team production. And in particular our analysis suggests that it is not so much the costs of spontaneously negotiating contracts in the markets among groups for team production as it is the detection of the performance of individual members of the team that calls for the organization noted here.

The specialist *who receives the residual rewards* will be the monitor of the members of the team (i.e., will manage the use of cooperative inputs). The monitor earns his residual through the reduction in shirking that he brings about, not only by the prices that he agrees to pay the owners of the inputs, but also by observing and directing the actions or uses of these inputs. *Managing or examining the ways to which inputs are used in team production is a method of metering the marginal productivity of individual inputs to the team's output.*

To discipline team members and reduce shirking, the residual claimant must have power to revise the contract terms and incentives of *individual* members without having to terminate or alter every other input's contract. Hence, team members who seek to increase their productivity will assign to the monitor not only the residual claimant right but also the right to alter individual membership and performance on the team. Each team member, of course, can terminate his own membership (i.e., quit the team), but only the monitor may unilaterally terminate the membership of any of the

Or output? It is the latter that is sought—the *effect* or output. But performance is nicely ambiguous because it suggests both input and output. It is *nicely* ambiguous because as we shall see, sometimes by inspecting a team member's input activity we can better judge his output effect, perhaps not with complete accuracy but better than by watching the output of the *team*. It is not always the case that watching input activity is the only or best means of detecting, measuring or monitoring output effects of each team member, but in some cases it is a useful way. For the moment the word performance glosses over these aspects and facilitates concentration on other issues.

other members without necessarily terminating the team itself or his association with the team; and he alone can expand or reduce membership, alter the mix of membership, or sell the right to be the residual claimant-monitor of the team. It is this entire bundle of rights: 1) to be a residual claimant; 2) to observe input behavior; 3) to be the central party common to all contracts with inputs; 4) to alter the membership of the team; and 5) to sell these rights, that defines the *ownership* (or the employer) of the *classical* (capitalist, free-enterprise) firm. The coalescing of these rights has arisen, our analysis asserts, because it resolves the shirking-information problem of team production better than does the noncentralized contractual arrangement.

The relationship of each team member to the *owner* of the firm (i.e., the party common to all input contracts *and* the residual claimant) is simply a "quid pro quo" contract. Each makes a purchase and sale. The employee "orders" the owner of the team to pay him money in the same sense that the employer directs the team member to perform certain acts. The employee can terminate the contract as readily as can the employer, and long-term contracts, therefore, are not an essential attribute of the firm. Nor are "authoritarian," "dictational," or "fiat" attributes relevant to the conception of the firm or its efficiency.

In summary, two necessary conditions exist for the emergence of the firm on the prior assumption that more than pecuniary wealth enter utility functions: 1) It is possible to increase productivity through team-oriented production, a production technique for which it is costly to directly measure the marginal outputs of the cooperating inputs. This makes it more difficult to restrict shirking through simple market exchange between cooperating inputs. 2) It is economical to estimate marginal productivity by observing or specifying input behavior. The simultaneous occurrence of both these preconditions leads to the contractual organization of inputs, known as the *classical capitalist firms* with (a) joint input production, (b) several input owners, (c) one party who is common to all the contracts of the joint inputs, (d) who has rights to renegotiate any input's contract independently of contracts with other input owners, (e) who holds the residual claim, and (f) who has the right to sell his central contractual residual status.[8]

Other Theories of the Firm

At this juncture, as an aside, we briefly place this theory of the firm in the contexts of those offered by Ronald Coase and Frank Knight.[9] Our view of the firm is not necessarily inconsistent with Coase's; we attempt to go further and identify refutable implications. Coase's penetrating insight is to make more of the fact that markets do not operate costlessly, and he relies on the cost of using markets to *form* contracts as his basic explanation for the existence of firms. We do not disagree with the proposition that, *ceteris paribus*, the higher is the cost of transacting across markets the greater will be the comparative advantage of organizing resources within the firm; it is a difficult proposition to disagree with or to refute. We could with equal ease subscribe to a theory of the firm based on the cost of managing, for surely it is true that, *ceteris paribus*, the lower is the cost of managing the greater will be the comparative advantage of organizing resources within the firm. To move the theory forward, it is necessary to know what is meant by a firm and to

[8] Removal of (b) converts a capitalist proprietary firm to a socialist firm.

[9] Recognition must also be made to the seminal inquiries by Morris Silver and Richard Auster, and by H. B. Malmgren.

explain the circumstances under which the cost of "managing" resources is low relative to the cost of allocating resources through market transaction. The conception of and rationale for the classical firm that we propose takes a step down the path pointed out by Coase toward that goal. Consideration of team production, team organization, difficulty in metering outputs, and the problem of shirking are important to our explanation but, so far as we can ascertain, not in Coase's. Coase's analysis insofar as it had heretofore been developed would suggest open-ended contracts but does not appear to imply anything more—neither the residual claimant status nor the distinction between employee and subcontractor status (nor any of the implications indicated below). And it is not true that employees are generally employed on the basis of long-term contractual arrangements any more than on a series of short-term or indefinite length contracts.

The importance of our proposed additional elements is revealed, for example, by the explanation of why the person to whom the control monitor is responsible receives the residual, and also by our later discussion of the implications about the corporation, partnerships, and profit sharing. These alternative forms for organization of the firm are difficult to resolve on the basis of market transaction costs only. Our exposition also suggests a definition of the classical firm—something crucial that was heretofore absent.

In addition, sometimes a technological development will lower the cost of market transactions while, at the same time, it expands the role of the firm. When the "putting out" system was used for weaving, inputs were organized largely through market negotiations. With the development of efficient central sources of power, it became economical to perform weaving in proximity to the power source and to engage in team production. The bringing in of weavers surely must have resulted in a reduction in the cost of negotiating (forming) contracts. Yet, what we observe is the beginning of the factory system in which inputs are organized within a firm. Why? The weavers did not simply move to a common source of power that they could tap like an electric line, purchasing power while they used their own equipment. Now team production in the joint use of equipment became more important. The measurement of marginal productivity, which now involved interactions between workers, especially through their joint use of machines, became more difficult though contract negotiating cost was reduced, while managing the *behavior* of inputs became easier because of the increased centralization of activity. The firm as an organization expanded even though the cost of transactions was reduced by the advent of centralized power. The same could be said for modern assembly lines. Hence the emergence of central power sources expanded the scope of productive activity in which the firm enjoyed a comparative advantage as an organizational form.

Some economists, following Knight, have identified the bearing of risks of wealth changes with the director or central employer without explaining why that is a viable arrangement. Presumably, the more risk-averse inputs become employees rather than owners of the classical firm. Risk averseness and uncertainty *with regard to the firm's fortunes* have little, if anything, to do with our explanation although it helps to explain why all resources in a team are not owned by one person. That is, the role of risk taken in the sense of absorbing the windfalls that buffet the firm because of unforeseen competition, technological change, or fluctuations in demand are not central to our theory, although it is true that imperfect knowledge and, therefore, risk, in *this* sense of risk, underlie the problem of

monitoring team behavior. We deduce the system of paying the manager with a residual claim (the equity) from the desire to have efficient means to reduce shirking so as to make team production economical and not from the smaller aversion to the risks of enterprise in a dynamic economy. We conjecture that "distribution-of-risk" is not a valid rationale for the *existence* and organization of the *classical* firm.

Although we have emphasized team production as creating a costly metering task and have treated team production as an essential (necessary?) condition for the firm, would not other obstacles to cheap metering also call forth the same kind of contractual arrangement here denoted as a firm? For example, suppose a farmer produces wheat in an easily ascertained quantity but with subtle and difficult to detect quality variations determined by how the farmer grew the wheat. A vertical integration could allow a purchaser to control the farmer's behavior in order to more economically estimate productivity. But this is not a case of joint or team production, unless "information" can be considered part of the product. (While a good case could be made for that broader conception of production, we shall ignore it here.) Instead of forming a firm, a buyer can contract to have his inspector on the site of production, just as home builders contract with architects to supervise building contracts; that arrangement is not a firm. Still, a firm might be organized in the production of many products wherein no team production or jointness of use of separately owned resources is involved.

This possibility rather clearly indicates a broader, or complementary, approach to that which we have chosen. 1) As we do in this paper, it can be argued that the firm is the particular policing device utilized when joint team production is present. If other sources of high policing costs arise, as in the wheat case just indicated, some other form of contractual arrangement will be used. Thus to each source of informational cost there may be a different type of policing and contractual arrangement. 2) On the other hand, one can say that where policing is difficult across markets, various forms of contractual arrangements are devised, but there is no reason for that known as the firm to be uniquely related or even highly correlated with team production, as defined here. It might be used equally probably and viably for other sources of high policing cost. We have not intensively analyzed other sources, and we can only note that our current and readily revisable conjecture is that 1) is valid, and has motivated us in our current endeavor. In any event, the test of the theory advanced here is to see whether the conditions we have identified are necessary for firms to have long-run viability rather than merely births with high infant mortality. Conglomerate firms or collections of separate production agencies into one owning organization can be interpreted as an investment trust or investment diversification device—probably along the lines that motivated Knight's interpretation. A holding company can be called a firm, because of the common association of the word firm with any ownership unit that owns income sources. The term firm as commonly used is so turgid of meaning that we can not hope to explain every entity to which the name is attached in common or even technical literature. Instead, we seek to identify and explain a particular contractual arrangement induced by the cost of information factors analyzed in this paper.

IV. Types of Firms

A. Profit-Sharing Firms

Explicit in our explanation of the capitalist firm is the assumption that the cost of *managing* the team's inputs by a central monitor, who disciplines himself because he is a residual claimant, is low

relative to the cost of metering the marginal outputs of team members.

If we look within a firm to see who monitors—hires, fires, changes, promotes, and renegotiates—we should find him being a residual claimant or, at least, one whose pay or reward is more than any others correlated with fluctuations in the residual value of the firm. They more likely will have options or rights or bonuses than will inputs with other tasks.

An implicit "auxiliary" assumption of our explanation of the firm is that the cost of team production is increased if the residual claim is not held entirely by the central monitor. That is, we assume that if profit sharing had to be relied upon for *all* team members, losses from the resulting increase in central monitor shirking would exceed the output gains from the increased incentives of other team members not to shirk. If the optimal team size is only two owners of inputs, then an equal division of profits and losses between them will leave each with stronger incentives to reduce shirking than if the optimal team size is large, for in the latter case only a smaller percentage of the losses occasioned by the shirker will be borne by him. Incentives to shirk are positively related to the optimal size of the team under an equal profit-sharing scheme.[10]

The preceding does not imply that profit sharing is never viable. Profit sharing to encourage self-policing is more appropriate for small teams. And, indeed, where input owners are free to make whatever contractual arrangements suit them, as generally is true in capitalist economies, profit sharing seems largely limited to partnerships with a relatively small number of *active*[11] partners. Another advantage of such arrangements for smaller teams is that it permits more effective reciprocal monitoring among inputs. Monitoring need not be entirely specialized.

Profit sharing is more viable if small team size is associated with situations where the cost of specialized management of inputs is large relative to the increased productivity potential in team effort. We conjecture that the cost of managing team inputs increases if the productivity of a team member is difficult to correlate with his behavior. In "artistic" or "professional" work, watching a man's activities is not a good clue to what he is actually thinking or doing with his mind. While it is relatively easy to manage or direct the loading of trucks by a team of dock workers where input activity is so highly related in an obvious way to output, it is more difficult to manage and direct a lawyer in the preparation and presentation of a case. Dock workers can be directed in detail without the monitor himself loading the truck, and assembly line workers can be monitored by varying the speed of the assembly line, but detailed direction in the preparation of a law case would require in much greater degree that the monitor prepare the case himself. As a result, artistic or professional inputs, such as lawyers, advertising specialists, and doctors, will be given relatively freer reign with regard to individual behavior. If the management of inputs is relatively costly, or ineffective, as it would seem to be in these cases, but, nonetheless if team effort is more productive than separable production with exchange across markets, then there will develop a tendency to use profit-sharing schemes to provide incentives to avoid shirking.[12]

[10] While the degree to which residual claims are centralized will affect the size of the team, this will be only one of many factors that determine team size, so as an approximation, we can treat team size as exogenously determined. Under certain assumptions about the shape of the "typical" utility function, the incentive to avoid shirking with unequal profit-sharing can be measured by the Herfindahl index.

[11] The use of the word active will be clarified in our discussion of the corporation, which follows below.

[12] Some sharing contracts, like crop sharing, or rental

B. Socialist Firms

We have analyzed the classical proprietorship and the profit-sharing firms in the context of free association and choice of economic organization. Such organizations need not be the most viable when political constraints limit the forms of organization that can be chosen. It is one thing to have profit sharing when professional or artistic talents are used by small teams. But if political or tax or subsidy considerations induce profit-sharing techniques when these are not otherwise economically justified, then additional management techniques will be developed to help reduce the degree of shirking.

For example, most, if not all, firms in Jugoslavia are owned by the employees in the restricted sense that all share in the residual. This is true for large firms and for firms which employ nonartistic, or nonprofessional, workers as well. With a decay of political constraints, most of these firms could be expected to rely on paid wages rather than shares in the residual. This rests on our auxiliary assumption that general sharing in the residual results in losses from enhanced shirking by the monitor that exceed the gains from reduced shirking by residual-sharing employees. If this were not so, profit sharing with employees should have occurred more frequently in Western societies where such organizations are neither banned nor preferred politically. Where residual sharing by employees is politically imposed, as in Jugoslavia, we are led to expect that some management technique will arise to reduce the shirking by the central monitor, a technique that will not be found frequently in Western societies since the monitor retains all (or much) of the residual in the West and profit sharing is largely confined to small, professional-artistic team production situations. We do find in the larger scale residual-sharing firms in Jugoslavia that there are employee committees that can recommend (to the state) the termination of a manager's contract (veto his continuance) with the enterprise. We conjecture that the workers' committee is given the right to recommend the termination of the manager's contract precisely because the general sharing of the residual increases "excessively" the manager's incentive to shirk.[13]

C. The Corporation

All firms must initially acquire command over some resources. The corporation does so primarily by selling promises of future returns to those who (as creditors or owners) provide financial capital. In some situations resources can be acquired in advance from consumers by promises of future delivery (for example, advance sale of a proposed book). Or where the firm is a few artistic or professional persons, each can "chip in" with time and talent until the sale of services brings in revenues. For the most part, capital can be acquired more cheaply if many (risk-averse) investors contribute small portions to a large investment. The economies of raising large sums of equity capital in this way suggest that modifications in the relationship among corporate inputs are required to cope with the shirking problem

payments based on gross sales in retail stores, come close to profit sharing. However, it is gross output sharing rather than profit sharing. We are unable to specify the implications of the difference. We refer the reader to S. N. Cheung.

[13] Incidentally, investment activity will be changed. The inability to capitalize the investment value as "take-home" proviate property *wealth* of the members of the firm means that the benefits of the investment must be taken as annual income by those who are employed at the time of the income. Investment will be confined more to those with shorter life and with higher rates or pay-offs if the alternative of investing is paying out the firm's income to its employees to take home and use as private property. For a development of this proposition, see the papers by Eirik Furobotn and Svetozar Pejovich, and by Pejovich.

that arises with profit sharing among large numbers of corporate stockholders. One modification is limited liability, especially for firms that are large relative to a stockholder's wealth. It serves to protect stockholders from large losses no matter how they are caused.

If every stock owner participated in each decision in a corporation, not only would large bureaucratic costs be incurred, but many would shirk the task of becoming well informed on the issue to be decided, since the losses associated with unexpectedly bad decisions will be borne in large part by the many other corporate shareholders. More effective control of corporate activity is achieved for most purposes by transferring decision authority to a smaller group, whose main function is to negotiate with and manage (renegotiate with) the other inputs of the team. The corporate stockholders retain the authority to revise the membership of the management group and over major decisions that affect the structure of the corporation or its dissolution.

As a result a new modification of partnerships is induced—the right to sale of corporate shares without approval of any other stockholders. Any shareholder can remove his wealth from control by those with whom he has differences of opinion. Rather than try to control the decisions of the management, which is harder to do with many stockholders than with only a few, unrestricted salability provides a more acceptable escape to each stockholder from continued policies with which he disagrees.

Indeed, the policing of managerial shirking relies on across-market competition from new groups of would-be managers as well as competition from members within the firm who seek to displace existing management. In addition to competition from outside and inside managers, control is facilitated by the temporary congealing of share votes into voting blocs owned by one or a few contenders. Proxy battles or stock-purchases concentrate the votes required to displace the existing management or modify managerial policies. But it is more than a change in policy that is sought by the newly formed financial interests, whether of new stockholders or not. It is the capitalization of expected future benefits into stock prices that concentrates on the innovators the wealth gains of their actions if they own large numbers of shares. Without capitalization of future benefits, there would be less incentive to incur the costs required to exert informed decisive influence on the corporation's policies and managing personnel. Temporarily, the structure of ownership is reformed, moving away from diffused ownership into decisive power blocs, and this is a transient resurgence of the classical firm with power again concentrated in those who have title to the residual.

In assessing the significance of stockholders' power it is not the usual diffusion of voting power that is significant but instead the frequency with which voting congeals into decisive changes. Even a one-man owned company may have a long term with just one manager—continuously being approved by the owner. Similarly a dispersed voting power corporation may be also characterized by a long-lived management. The question is the probability of replacement of the management if it behaves in ways not acceptable to a majority of the stockholders. The unrestricted salability of stock and the transfer of proxies enhances the probability of decisive action in the event current stockholders or any outsider believes that management is not doing a good job with the corporation. We are not comparing the corporate responsiveness to that of a single proprietorship; instead, we are indicating features of the corporate structure that are induced by the problem of

delegated authority to manager-monitors.[14]

D. *Mutual and Nonprofit Firms*

The benefits obtained by the new management are greater if the stock can be purchased and sold, because this enables *capitalization* of anticipated future improvements into present *wealth* of new managers who bought stock and created a larger capital by their management changes. But in nonprofit corporations, colleges, churches, country clubs, mutual savings banks, mutual insurance companies, and "coops," the future consequences of improved management are not

[14] Instead of thinking of shareholders as joint *owners*, we can think of them as investors, like bondholders, except that the stockholders are more optimistic than bondholders about the enterprise prospects. Instead of buying bonds in the corporation, thus enjoying smaller risks, shareholders prefer to invest funds with a greater realizable return if the firm prospers as expected, but with smaller (possibly negative) returns if the firm performs in a manner closer to that expected by the more pessimistic investors. The pessimistic investors, in turn, regard only the bonds as likely to pay off.

If the entrepreneur-organizer is to raise capital on the best terms to him, it is to his advantage, as well as that of prospective investors, to recognize these differences in expectations. The residual claim on earnings enjoyed by shareholders does not serve the function of enhancing their efficiency as monitors in the general situation. The stockholders are "merely" the less risk-averse or the more optimistic member of the group that finances the firm. Being more optimistic than the average and seeing a higher mean value future return, they are willing to pay more for a certificate that allows them to realize gain on their expectations. One method of doing so is to buy claims to the distribution of returns that "they see" while bondholders, who are more pessimistic, purchase a claim to the distribution that they see as more likely to emerge. Stockholders are then comparable to warrant holders. They care not about the voting rights (usually not attached to warrants); they are in the same position in so far as voting rights are concerned as are bondholders. The only difference is in the probability distribution of rewards and the terms on which they can place their bets.

If we treat bondholders, preferred and convertible preferred stockholders, and common stockholders and warrant holders as simply different classes of investors—differing not only in their risk averseness but in their beliefs about the probability distribution of the firm's future earnings, why should stockholders be regarded as "owners" in any sense distinct from the other financial investors? The entrepreneur-organizer, who let us assume is the chief operating officer and sole repository of control of the corporation, does not find his authority residing in common stockholders (except in the case of a take over). Does this type of control make any difference in the way the firm is conducted? Would it make any difference in the kinds of behavior that would be tolerated by competing managers and investors (and we here deliberately refrain from thinking of them as owner-stockholders in the traditional sense)?

Investment old timers recall a significant incidence of nonvoting common stock, now prohibited in corporations whose stock is traded on listed exchanges. (Why prohibited?) The entrepreneur in those days could hold voting shares while investors held nonvoting shares, which in every other respect were identical. Nonvoting share holders were simply investors devoid of ownership connotations. The control and behavior of inside owners in such corporations has never, so far as we have ascertained, been carefully studied. For example, at the simplest level of interest, does the evidence indicate that nonvoting shareholders fared any worse because of not having voting rights? Did owners permit the nonvoting holders the normal return available to voting shareholders? Though evidence is prohibitively expensive to obtain, it is remarkable that voting and nonvoting shares sold for essentially identical prices, even during some proxy battles. However, our casual evidence deserves no more than interest-initiating weight.

One more point. The facade is deceptive. Instead of nonvoting shares, today we have warrants, convertible preferred stocks all of which are solely or partly "equity" claims without voting rights, though they could be converted into voting shares.

In sum, is it the case that the stockholder-investor relationship is one emanating from the *division* of *ownership* among several people, or is it that the collection of investment funds from people of varying anticipations is the underlying factor? If the latter, why should any of them be thought of as the owners in whom voting rights, whatever they may signify or however exercisable, should reside in order to enhance efficiency? Why voting rights in any of the outside, participating investors?

Our initial perception of this possibly significant difference in interpretation was precipitated by Henry Manne. A reading of his paper makes it clear that it is hard to understand why an investor who wishes to back and "share" in the consequences of some new business should necessarily have to acquire voting power (i.e., power to change the manager-operator) in order to invest in the venture. In fact, we invest in some ventures in the hope that no other stockholders will be so "foolish" as to try to toss out the incumbent management. We want him to have the power to stay in office, and for the prospect of sharing in his fortunes we buy nonvoting common stock. Our willingness to invest is enhanced by the knowledge that we can act legally via fraud, embezzlement and other laws to help assure that we outside investors will not be "milked" beyond our initial discounted anticipations.

capitalized into present wealth of stockholders. (As if to make more difficult that competition by new would-be monitors, mutiple shares of ownership in those enterprises cannot be bought by one person.) One should, therefore, find greater shirking in nonprofit, mutually owned enterprises. (This suggests that nonprofit enterprises are especially appropriate in realms of endeavor where more shirking is desired and where redirected uses of the enterprise in response to market-revealed values is less desired.)

E. Partnerships

Team production in artistic or professional intellectual skills will more likely be by partnerships than other types of team production. This amounts to market-organized team activity and to a non-employer status. Self-monitoring partnerships, therefore, will be used rather than employer-employee contracts, and these organizations will be small to prevent an excessive dilution of efforts through shirking. Also, partnerships are more likely to occur among relatives or long-standing acquaintances, not necessarily because they share a common utility function, but also because each knows better the other's work characteristics and tendencies to shirk.

F. Employee Unions

Employee unions, whatever else they do, perform as monitors for employees. Employers monitor employees and similarly employees monitor an employer's performance. Are correct wages paid on time and in good currency? Usually, this is extremely easy to check. But some forms of employer performance are less easy to meter and are more subject to employer shirking. Fringe benefits often are in nonpecuniary, contingent form; medical, hospital, and accident insurance, and retirement pensions are contingent payments or performances partly in *kind* by employers to employees. Each employee cannot judge the character of such payments as easily as money wages. Insurance is a contingent payment—what the employee will get upon the contingent event may come as a disappointment. If he could easily determine what other employees had gotten upon such contingent events he could judge more accurately the performance by the employer. He could "trust" the employer not to shirk in such fringe contingent payments, but he would prefer an effective and economic monitor of those payments. We see a specialist monitor—the union employees' agent—hired by them and monitoring those aspects of employer payment most difficult for the employees to monitor. Employees should be willing to employ a specialist monitor to administer such hard-to-detect employer performance, even though their monitor has incentives to use pension and retirement funds not entirely for the benefit of employees.

V. Team Spirit and Loyalty

Every team member would prefer a team in which no one, not even himself, shirked. Then the true marginal costs and values could be equated to achieve more preferred positions. If one could enhance a common interest in nonshirking in the guise of a team loyalty or team spirit, the team would be more efficient. In those sports where team activity is most clearly exemplified, the sense of loyalty and team spirit is most strongly urged. Obviously the team is better, with team spirit and loyalty, because of the reduced shirking—not because of some other feature inherent in loyalty or spirit as such.[15]

[15] *Sports Leagues:* Professional sports contests among teams is typically conducted by a *league* of teams. We assume that sports consumers are interested not only in absolute sporting skill but also in skills *relative* to other teams. Being slightly better than opposing teams enables one to claim a major portion of the receipts; the

Corporations and business firms try to instill a spirit of loyalty. This should not be viewed simply as a device to increase profits by *over*-working or misleading the employees, nor as an adolescent urge for belonging. It promotes a closer approximation to the employees' potentially available true rates of substitution between production and leisure and enables each team member to achieve a more preferred situation. The difficulty, of course, is to create economically that team spirit and loyalty. It can be preached with an aura of moral code of conduct—a morality with literally the same basis as the ten commandments—to restrict our conduct toward what we would choose if we bore our full costs.

inferior team does not release resources and reduce costs, since they were expected in the play of contest. Hence, absolute skill is developed beyond the equality of marginal investment in sporting skill with its true social marginal value product. It follows there will be a tendency to overinvest in training athletes and developing teams. "Reverse shirking" arises, as budding players are induced to overpractice hyperactively relative to the social marginal value of their enhanced skills. To prevent overinvestment, the teams seek an agreement with each other to restrict practice, size of teams, and even pay of the team members (which reduces incentives of young people to overinvest in developing skills). Ideally, if all the contestant teams were owned by one owner, overinvestment in sports would be avoided, much as ownership of common fisheries or underground oil or water reserve would prevent overinvestment. This hyperactivity (to suggest the opposite of shirking) is controlled by the league of teams, wherein the league adopts a common set of constraints on each team's behavior. In effect, the teams are no longer really owned by the team owners but are supervised by them, much as the franchisers of some product. They are not full-fledged owners of their business, including the brand name, and can not "do what they wish" as franchises. Comparable to the franchiser, is the league commissioner or conference president, who seeks to restrain hyperactivity, as individual team supervisors compete with each other and cause external diseconomies. Such restraints are usually regarded as anticompetitive, antisocial, collusive-cartel devices to restrain free open competition, and reduce players' salaries. However, the interpretation presented here is premised on an attempt to avoid hyperinvestment in team sports production. Of course, the team operators have an incentive, once the league is formed and restraints are placed on hyperinvestment activity, to go further and obtain the private benefits of monopoly restriction. To what extent overinvestment is replaced by monopoly restriction is not yet determinable; nor have we seen an empirical test of these two competing, but mutually consistent interpretations. (This interpretation of league-sports activity was proposed by Earl Thompson and formulated by Michael Canes.) Again, athletic teams clearly exemplify the specialization of monitoring with captains and coaches; a captain detects shirkers while the coach trains and selects strategies and tactics. Both functions may be centralized in one person.

VI. Kinds of Inputs Owned by the Firm

To this point the discussion has examined why firms, as we have defined them, exist? That is, why is there an owner-employer who is the common party to contracts with other owners of inputs in team activity? The answer to that question should also indicate the kind of the jointly used resources likely to be owned by the central-owner-monitor and the kind likely to be hired from people who are not team-owners. Can we identify characteristics or features of various inputs that lead to their being hired or to their being owned by the firm?

How can residual-claimant, central-employer-owner demonstrate ability to pay the other hired inputs the promised amount in the event of a loss? He can pay in advance or he can commit wealth sufficient to cover negative residuals. The latter will take the form of machines, land, buildings, or raw materials committed to the firm. Commitments of labor-wealth (i.e., human wealth) given the property rights in people, is less feasible. These considerations suggest that residual claimants—owners of the firm—will be investors of resalable capital equipment in the firm. The goods or inputs more likely to be invested, than rented, by the owners of the enterprise, will have higher resale values relative to the initial cost and will have longer expected use in a firm relative to the economic life of the good.

But beyond these factors are those developed above to explain the existence of

the institution known as the firm—the costs of detecting output performance. When a durable resource is used it will have a marginal product and a depreciation. Its use requires payment to cover at least use-induced depreciation; unless that user cost is specifically detectable, payment for it will be demanded in accord with *expected* depreciation. And we can ascertain circumstances for each. An indestructible hammer with a readily detectable marginal product has zero user cost. But suppose the hammer were destructible and that careless (which is easier than careful) use is more abusive and causes greater depreciation of the hammer. Suppose in addition the abuse is easier to detect by observing the way it is used than by observing only the hammer after its use, or by measuring the output scored from a hammer by a laborer. If the hammer were rented and used in the absence of the owner, the depreciation would be greater than if the use were observed by the owner and the user charged in accord with the imposed depreciation. (Careless use is more likely than careful use—if one does not pay for the greater depreciation.) An absentee owner would therefore ask for a higher rental price because of the higher *expected* user cost than if the item were used by the owner. The expectation is higher because of the greater difficulty of observing specific user cost, by inspection of the hammer after use. Renting is therefore in this case more costly than owner use. This is the valid content of the misleading expressions about ownership being more economical than renting—ignoring all other factors that may work in the opposite direction, like tax provision, short-term occupancy and capital risk avoidance.

Better examples are tools of the trade. Watch repairers, engineers, and carpenters tend to own their own tools especially if they are portable. Trucks are more likely to be employee owned rather than other equally expensive team inputs because it is relatively cheap for the driver to police the care taken in using a truck. Policing the use of trucks by a nondriver owner is more likely to occur for trucks that are not specialized to one driver, like public transit busses.

The factor with which we are concerned here is one related to the costs of monitoring not only the gross product performance of an input but also the abuse or depreciation inflicted on the input in the course of its use. If depreciation or user cost is more cheaply detected when the owner can see its use than by only seeing the input before and after, there is a force toward owner use rather than renting. Resources whose user cost is harder to detect when used by someone else, tend on this count to be owner-used. Absentee ownership, in the lay language, will be less likely. Assume momentarily that labor service cannot be performed in the absence of its owner. The labor owner can more cheaply monitor any abuse of himself than if somehow labor-services could be provided without the labor owner observing its mode of use or knowing what was happening. Also his incentive to abuse himself is increased if he does not own himself.[16]

[16] Professional athletes in baseball, football, and basketball, where athletes having sold their source of service to the team owners upon entering into sports activity, are owned by team owners. Here the team owners must monitor the athletes' physical condition and behavior to protect the team owners' wealth. The athlete has *less* (not, *no*) incentive to protect or enhance his athletic prowess since capital value changes have less impact on his own wealth and more on the team owners. Thus, some athletes sign up for big initial bonuses (representing present capital value of future services). Future salaries are lower by the annuity value of the prepaid "bonus" and hence the athlete has *less* to lose by subsequent abuse of his athletic prowess. Any decline in his subsequent service value would in part be borne by the team owner who owns the players' future service. This does not say these losses of future salaries have no effect on preservation of athletic talent (we are not making a "sunk cost" error). Instead, we assert that the

The similarity between the preceding analysis and the question of absentee landlordism and of sharecropping arrangements is no accident. The same factors which explain the contractual arrangements known as a firm help to explain the incidence of tenancy, labor hiring or sharecropping.[17]

VII. Firms as a Specialized Market Institution for Collecting, Collating, and Selling Input Information

The firm serves as a highly specialized surrogate market. Any person contemplating a joint-input activity must search and detect the qualities of available joint inputs. He could contact an employment agency, but that agency in a small town would have little advantage over a large firm with many inputs. The employer, by virtue of monitoring many inputs, acquires special superior information about their productive talents. This aids his *directive* (i.e., market hiring) efficiency. He "sells" his information to employee-inputs as he aids them in ascertaining good input combinations for team activity. Those who work as employees or who rent services to him are using him to discern superior combinations of inputs. Not only does the director-employer "decide" what each input will produce, he also estimates which heterogeneous inputs will work together jointly more efficiently, and he does this in the context of a privately owned market for forming teams. The department store is a firm and is a superior private market. People who shop and work in one town can as well shop and work in a privately owned firm.

This marketing function is obscured in the theoretical literature by the assumption of homogeneous factors. Or it is tacitly left for individuals to do themselves via personal market search, much as if a person had to search without benefit of specialist retailers. Whether or not the firm arose because of this efficient information service, it gives the director-employer more knowledge about the productive talents of the team's inputs, and a basis for superior decisions about efficient or profitable combinations of those heterogeneous resources.

In other words, opportunities for profitable team production by inputs already within the firm may be ascertained more economically and accurately than for resources outside the firm. Superior combinations of inputs can be more economically identified and formed from resources already used in the organization than by obtaining new resources (and knowledge of them) from the outside. Promotion and revision of employee assignments (contracts) will be preferred by a firm to the hiring of new inputs. To the extent that this occurs there is reason to expect the firm to be able to operate as a conglomerate rather than persist in producing a single product. Efficient production with heterogeneous resources is a result not of having *better* resources but in *knowing more accurately* the relative productive performances of those resources. Poorer resources can be paid less in accord with their inferiority; greater accuracy of

preservation is reduced, not eliminated, because the amount of loss of wealth suffered is smaller. The athlete will spend less to maintain or enhance his prowess thereafter. The effect of this revised incentive system is evidenced in comparisons of the kinds of attention and care imposed on the athletes at the "expense of the team owner" in the case where atheletes' future servies are owned by the team owner with that where future labor service values are owned by the athlete himself. Why athletes' future athletic services are owned by the team owners rather than being hired is a question we should be able to answer. One presumption is cartelization and monopsony gains to team owners. Another is exactly the theory being expounded in this paper—costs of monitoring production of athletes; we know not on which to rely.

[17] The analysis used by Cheung in explaining the prevalence of sharecropping and land tenancy arrangements is built squarely on the same factors—the costs of detecting output performance of jointly used inputs in team production and the costs of detecting user costs imposed on the various inputs if owner used or if rented.

knowledge of the potential and actual productive actions of inputs rather than having high productivity resources makes a firm (or an assignment of inputs) profitable.[18]

VIII. Summary

While ordinary contracts facilitate efficient specialization according to comparative advantage, a special class of contracts among a group of joint inputs to a team production process is commonly used for team production. Instead of multilateral contracts among all the joint inputs' owners, a central common party to a set of bilateral contracts facilitates efficient organization of the joint inguts in team production. The terms of the contracts form the basis of the entity called the firm—especially appropriate for organizing team production processes.

Team productive activity is that in which a union, or joint use, of inputs yields a larger output than the sum of the products of the separately used inputs. This team production requires—like all other production processes—an assessment of marginal productivities if efficient production is to be achieved. Nonseparability of the products of several differently owned joint inputs raises the cost of assessing the marginal productivities of those resources or services of each input owner. Monitoring or metering the productivities to match marginal productivities to costs of inputs and thereby to reduce shirking can be achieved more economically (than by across market bilateral negotiations among inputs) in a firm.

The essence of the classical firm is identified here as a contractual structure with: 1) joint input production; 2) several input owners; 3) one party who is common to all the contracts of the joint inputs; 4) who has rights to renegotiate any input's contract independently of contracts with other input owners; 5) who holds the residual claim; and 6) who has the right to sell his central contractual residual status. The central agent is called the firm's owner and the employer. No authoritarian control is involved; the arrangement is simply a contractual structure subject to continuous renegotiation with the central agent. The contractual structure arises as a means of enhancing efficient organization of team production. In particular, the ability to detect shirking among owners of jointly used inputs in team production is enhanced (detection costs are reduced) by this arrangement and the discipline (by revision of contracts) of input owners is made more economic.

Testable implications are suggested by the analysis of different types of organizations—nonprofit, proprietary for profit, unions, cooperatives, partnerships, and by the kinds of inputs that tend to be owned by the firm in contrast to those employed by the firm.

We conclude with a highly conjectural

[18] According to our interpretation, the firm is a specialized surrogate for a market for team use of inputs; it provides superior (i.e., cheaper) collection and collation of knowledge about heterogeneous resources. The greater the set of inputs about which knowledge of performance is being collated within a firm the greater are the present costs of the collation activity. Then, the larger the firm (market) the greater the attenuation of monitor control. To counter this force, the firm will be divisionalized in ways that economize on those costs—just as will the market be specialized. So far as we can ascertain, other theories of the reasons for firms have no such implications.

In Japan, employees by custom work nearly their entire lives with one firm, and the firm agrees to that expectation. Firms will tend to be large and conglomerate to enable a broader scope of input revision. Each firm is, in effect, a small economy engaging in "intranational and international" trade. Analogously, Americans expect to spend their whole lives in the United States, and the bigger the country, in terms of variety of resources, the easier it is to adjust to changing tastes and circumstances. Japan, with its lifetime employees, should be characterized more by large, conglomerate firms. Presumably, at some size of the firm, specialized knowledge about inputs becomes as expensive to transmit across divisions of the firms as it does across markets to other firms.

but possibly significant interpretation. As a consequence of the flow of information to the central party (employer), the firm takes on the characteristic of an efficient market in that information about the productive characteristics of a large set of specific inputs is now more cheaply available. Better recombinations or new uses of resources can be more efficiently ascertained than by the conventional search through the general market. In this sense inputs compete with each other within and via a firm rather than solely across markets as conventionally conceived. Emphasis on interfirm competition obscures intrafirm competition among inputs. Conceiving competition as the *revelation and exchange* of knowledge or information about qualities, potential uses of different inputs in different potential applications indicates that the firm is a device for enchancing competition among sets of input resources as well as a device for more efficiently rewarding the inputs. In contrast to markets and cities which can be viewed as publicly or nonowned market places, the firm can be considered a privately owned market; if so, we could consider the firm and the ordinary market as competing types of markets, competition between private proprietary markets and public or communal markets. Could it be that the market suffers from the defects of communal property rights in organizing and influencing uses of valuable resources?

REFERENCES

M. Canes, "A Model of a Sports League," unpublished doctoral dissertation, UCLA 1970.

S. N. Cheung, *The Theory of Share Tenancy,* Chicago 1969.

R. H. Coase, "The Nature of the Firm," *Economica,* Nov. 1937, *4,* 386–405; reprinted in G. J. Stigler and K. Boulding, eds., *Readings in Price Theory,* Homewood 1952, 331–51.

E. Furobotn and S. Pejovich, "Property Rights and the Behavior of the Firm in a Socialist State," *Zeitschrift für Nationalökonomie,* 1970, *30,* 431–454.

F. H. Knight, *Risk, Uncertainty and Profit,* New York 1965.

S. Macaulay, "Non-Contractual Relations in Business: A Preliminary Study," *Amer. Sociological Rev.,* 1968, *28,* 55–69.

H. B. Malmgren, "Information, Expectations and the Theory of the Firm," *Quart J. Econ.,* Aug. 1961, *75,* 399–421.

H. Manne, "Our Two Corporation Systems: Law and Economics," *Virginia Law Rev.,* Mar. 1967, *53,* No. 2, 259–84.

S. Pejovich, "The Firm, Monetary Policy and Property Rights in a Planned Economy," *Western Econ. J.,* Sept. 1969, *7,* 193–200.

M. Silver and R. Auster, "Entrepreneurship, Profit, and the Limits on Firm Size," *J. Bus. Univ. Chicago,* Apr. 1969, *42,* 277–81.

E. A. Thompson, "Nonpecuniary Rewards and the Aggregate Production Function," *Rev. Econ. Statist.,* Nov. 1970, *52,* 395–404.