1978

A Generalization of the "Property Rights" Theory of the Firm

Ronald Wintrobe

RESEARCH REPORT 7807

A GENERALIZATION OF THE "PROPERTY RIGHTS" THEORY OF THE FIRM

by

Ronald Wintrobe

March, 1978

A Generalization of the "Property Rights" Theory of the Firm

Ronald Wintrobe

University of Western Ontario

This paper is based on chapter 2 of my doctoral dissertation at the University of Toronto. An earlier version was presented at the March, 1977 meetings of the Public Choice Society. I am indebted to Albert Breton, Charles K. Harley, Allan Hynes, Yehuda Kotowitz and Louis Wilde for Helpful comments and suggestions. Any errors remain my responsibility.

1. Introduction:

The purpose of this paper is to provide a theory of internal organization, i.e., to show the conditions under which the organization of production within firms or government bureaus is superior to market exchange. The argument to be presented owes much to that expounded by Alchian and Demsetz[1], in that one basic condition for hierarchy is input complementarity, but it both differs from and goes beyond their paper in a number of respects. The analysis is also related to the important work of Oliver Williamson on market vs hierarchical organization[2]; apart from detailed similarities and differences, this paper does attempt to place the theory of internal organization squarely within contemporary economic theory, rather than to depart from it, as Williamson believes is necessary to understand the operation of hierarchies[3].

Moreover, the present argument incorporates the valid insights of the other major previous theories of the firm-- those of Coase, Knight, Stigler and Robinson[4] and thus attempts to provide a single integrating perspective for demonstrating the gain from hierarchical as opposed to market organization.

Alchian and Demsetz argued that the market metering system of payments to factors according to their output (marginal products) tends to be inefficient under the following conditions:

(1) Factors care about non-pecuniary aspects of their jobs, and, in particular, they prefer on-the-job leisure to working.

(2) It is possible to increase productivity through "team-oriented production", in which factors are complements, and in which the marginal products of different factors are non-separable.

Alchian and Demsetz showed that, in "team" production it is difficult to measure each factor's marginal productivity by looking at changes in output. Non-separability implies that it is impossible to isolate the individual contribution of each factor to output, and to remunerate factors of this basis. If factors are paid on the basis of the total output of the team each factor will have an incentive to shirk, since the loss of income from shirking will be partly borne by other members of the team. However, it may be possible, by observing the behaviour of factors (input) to detect such shirking more easily, and if this third condition is satisfied, a firm emerges since the factors will jointly agree to hire a monitor to supervise them to reduce shirking. In order to provide the proper incentive to the monitor, he must receive the residual claim to the sum of the factor's earnings.

Now, although this argument establishes a case for hierarchy in the organization of team production, it is weak as a rationale for the existence of firms. Indeed, there is no place in the theory for the classic function of the firm --the coordination of production by hierarchical direction, rather than by the market. The major previous theorists of the firm --Coase and Knight-- used this function of direction to account for the special contractual relation between employer and employee. Coase defined this contract as

> ....one whereby the factor, for a certain remuneration (which may be fixed or fluctuating) agrees to obey the directions of an entrepreneur within certain limits.[5]

Coase's definition clearly accords with common usage.

Frank Knight differed from Coase in attributing the origin of firms to uncertainty rather than to transactions costs. Knight's "firm" is a

> ...system under which the confident and 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...[6]

But, he also emphasized the fact of direction:

> With human nature as we know it would be unpracticable or very unusual for one man to guarantee to another a definite result of the latter's actions without being given 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.[7]

Alchian and Demsetz' theory of the firm, on the other hand, does not account for the phenomenon of hierarchical direction. The possibility of shirking explains why the factors would agree to appoint a common monitor, but it does not explain why they would agree to submit their resources to common direction as well as monitoring.

A second difficulty with the theory is that the costs of metering shirking cannot be very large, and hence using differences in these costs to explain variations in organizational structure and reward systems-- presumably, one of the basic purposes of a theory of the firm-- does not appear very promising.

Finally, the case for team production requires that the marginal products of the different inputs be non-separable. If, as has been pointed out elsewhere,[8] most integrated activities are separable, then a large number of activities which are in fact vertically integrated do not fit Alchian and Demsetz' theory.

An explanation of the vertical integration of separable activities has been provided by Stigler,[9] based on Adam Smith's famous principle that the division of labour is limited by the size of the market. Now, although complementarity is a broader concept than the division of labour, it is clearly intrinsic to the process of specialization: when a production

process is divided into sub-processes, the intermediate outputs of each of the component processes are obviously complements.

The assumption that factor complementarity is essential to hierarchical organization is also, we believe, the key insight of the work of Alchian and Demsetz. This paper attempts to build further on the work of Stigler and Alchian and Demsetz. The latter's assumption of non-separability and the associated notion of "team" production is discarded. We try to show that the basic rationale for hierarchical organization lies instead in the problem of coordination, which arises when factors are complements, and where the number of factor suppliers is small, and that the methods by which the firm executes this task require that it have precisely the characteristics attributed to it by Coase and Knight.

The "firm" is defined by the structure of property rights implied in the contract between employer and employee. Thus, in a system of <u>private</u> property rights (1) the owners of resource inputs retain the exclusive right to use them as they see fit; (2) the ownership of resources is freely transferable.[10] In the employer-employee contract, employees give up (1) but not (2): employees do not "sell" their resources to an organization; they do give up the right of exclusive control over the use to which their time is put to officers of the organization who are their hierarchical superiors. During the specified hours of employment, the allocation of the employee's time may or may not at all times be directly regulated by his superiors; but the exclusive <u>right</u> of control is theirs; that is what has been "purchased" by the salary offered in the employment contract.

Given this conception of the firm, when will it exist, i.e., under what conditions is this contractual relation an efficient organizational form?

To outline the argument, we will suggest first that the following four conditions are sufficient to predict the emergence of hierarchical organization in production. They are:

(1) The division of labour increase productivity (complementarity)
(2) There are a relatively small number of factor suppliers
(3) Uncertainty
(4) The technology of production is relatively "routine".

Section 2 demonstrates that, if the first two conditions are present, production by different inputs must be coordinated in order to maximize final output. And we show there that if the third condition -- uncertainty-- is present, this problem of coordination is greatly exacerbated. Section 3 contrasts the respective efficiency of the firm and the market at this task of coordination, and shows how the addition of the fourth condition ensures that hierarchical organization will displace market relations in the organization of production. The characteristic disadvantage of hierarchy when condition (4) does not hold is pointed out, and the choice between market and hierarchy is discussed under more general conditions. Section 4 concludes the paper.

2. The Division of Labour and the Problem of Coordination

In this section, we demonstrate that problems of coordination arise whenever two conditions hold: factors are complements, and their number is small. The first of these conditions is due to the division of labour, and the second arises because of Adam Smith's theorem - the division of labour is limited to the extent of the market.

That the division of labour makes for more efficient production, and that these technical efficiencies are often only achievable at relatively large volumes of production --so that they are the principle basis of economies of scale-- is well known. That the process of specialization also gives rise to problems of coordination has been neglected in theories of the firm,[11] although it was emphasized by Robinson.[12] When the process of production is factored into sub-processes, the sub-processes must be coordinated, i.e. the different sub-processes must be timed so that their outputs are made available simultaneously in the correct proportions, and their outputs must be made to "fit" together in various ways: in terms of their physical dimensions, in terms of the respective qualities of the various components, etc. Now, if the market for an intermediate output is sufficiently large to permit a number of firms to specialize in its production, the problem of lack of "fit" along any dimension such as availability or quality vanishes. Where interchangeability is not present to that degree, so that the number of producers of a given sub-process is small, maximization of final output for given inputs requires coordination of the production of the intermediate output with the intermediate outputs of complementary sub-processes.

To illustrate the problem of coordination mathematically, assume for simplicity that there are only two relevant dimensions to the outputs of the different sub-processes-- their average level, and their dispersion or variability. Factors choose among a range of different methods of production, which differ from each other only in these two characteristics. Methods with greater dispersion might be chosen for at least two reasons: firstly, uncertain methods might yield a higher average level of intermediate

output, i.e. $\frac{\partial \bar{x}}{\partial \sigma^2_x} > 0$ where $\bar{x}$ is average intermediate output and σ^2_x is its variance. Secondly, methods with some variance would in effect be "chosen" if factors simply have an aversion to discipline: if factors are free to produce whenever they choose, there is no reason to believe that they would choose perfectly regular rates of production.[13]

Assume further that only two methods are available, one stochastic method and one which is non-stochastic. It can then be shown (see the appendix):(1) that if the intermediate outputs of the two factors are perfect substitutes, then as long as the expected intermediate outputs from the use of the two methods is the same, expected final output is the same no matter which method is used; (2) that this is <u>not</u> the case if the intermediate outputs are complements. Even if expected intermediate outputs are identical under either procedure, when inputs are complements, expected final output will be less when the stochastic method is used. The greater the degree of complementarity between intermediate outputs, the more the level of expected final outputs declines when factors use the stochastic method.

The reason is obvious: when factors use the stochastic method, dispersion in their rates of output implies that the two intermediate factor outputs will not be made available simultaneously in the correct proportions in all states of the world; as is well known, complementarity implies that they must be made available in the correct proportions if expected final output is to be maximized.

The implication of this result is that production will be organized inefficiently if the factors were to contract to meter and pay each solely on the basis of the quantities of their intermediate outputs. We

have identified two dimensions of output-- its level, and its dispersion-- and if payment to factors is made on the basis of one of these dimensions only, say, an index of the level of intermediate output-- factors will neglect the second-- the dispersion of output. In equilibrium, factors will not exercise enough discipline, and they will take too much risk, since this mode of payment provides no incentive for them to regularize their production; accordingly, they will not take into account the damanges to complementary factors from insufficient availability of their resources at some times and over-availability at others.

Analytically, the damages imposed on complementary factors are "external" to the payment system used, and each factor will not equate, at the margin, the damages to complementary factors from dispersion in his rate of output with the costs to him of reducing that dispersion.

Such external effects do not arise if factors are perfect substitutes because, in that case, factors may be combined in any proportion, and changes in the variance of their intermediate outputs has no effect on the level of final output. Similarly, if there were a relatively large number of suppliers, variations in the rates of production of individual supplies would wash out by the law of large numbers,[14] and again, dispersion in intermediate outputs is irrelevant to the maximization of expected final output.

Note that these external effects do arise even if no shirking takes place, even if factors do not care about non-pecuniary aspects of their jobs and even if the intermediate outputs of the different factors are separable. The sole conditions required are that factors be complements, and that there

are a relatively small number of suppliers. Under these conditions, production by different factors inputs must be coordinated to minimize costs.

The argument is the same with respect to the other dimensions of factor output. Suppose there are N such dimensions. If any one of these dimensions in any one factor's output is not priced, that factor will ignore it in his production decisions. His output will not be coordinated with the outputs of complementary suppliers along the unpriced dimension, and, at the margin, the damages to complementary factors will be larger than the costs to that factor of taking that dimension into account.

3. Markets vs. Firms as Coordinators

Will complementary sub-processes therefore be integrated, and coordinated within firms, or can the necessary coordination take place via market relations between the different suppliers? In this section we examine the respective efficiency of the firm and the market at the task of coordination. We continue to simplify the analysis by assuming the complementary factors need to be coordinated along a single dimension (timing) only.

The solution to the problem of coordination via the market is that of pricing the timing of production. If the market could correctly approximate the required timing price, production would be appropriately coordinated without the emergence of a hierarchical structure. Indeed,

market prices for timing are not unusual, as exemplified by contracts, e.g. in construction, which stipulate penalties for late delivery of goods.

But what is the appropriate penalty? The proper incentive to each factor can only be provided by a penalty which reflects the total loss in the value of output from late delivery, and this loss is equal to the sum of the loss in the marginal productivities of all the complementary factors affected. It follows that the appropriate timing penalty for each complementary factor cannot be determined independently from that for the other factors with which it is to be combined. The timing price must therefore be negotiated, and the negotiations must take place multilaterally between all of the complementary factors. The larger the number of affected parties, and the greater the interdependence between them, the larger the required timing penalty.

If such a penalty is successfully negotiated, then each of the suppliers is effectively insured against losses in the value of their own intermediate output due to shortfalls in each others' rates of production. Moreover, the frequency of these shortfalls will be reduced in that the incentive to each supplier for discretionary timing of his own rate of production is eliminated. There remains, however, the problem of random shortfalls in production due to circumstances not under the suppliers' control. Each supplier would like to purchase insurance against the possibility of having to pay the timing penalty due to unforeseen circumstances which might cause delays in delivery. But no private firm would offer such insurance, since this would amount to insuring the supplier against possible adverse consequences of his own production decisions, and the usual limit on insurance due to the problem of "moral hazard" is obviously especially acute in this context.

Where feasible, however, suppliers can insure themselves against losses due to random variations in production by holding buffer inventories. Where this is not feasible, or where it is expensive, either the buyer of the intermediate outputs or the suppliers may provide insurance by agreeing to a lower timing penalty than the appropriate one, since the "appropriate" one may be too high for any of them to bear. The problem of moral hazard remains, and it increases as the timing penalty is lowered: "random" deviations in production become increasingly confounded with shortfalls due to supplier negligence and discretionary timing. The lower the penalty, the greater the lack of coordination, i.e., the greater the frequency of shortfalls in production, and the more the burden of these shortfalls falls either collectively on the suppliers or on the buyer of the complementary factors, depending on which party provides the insurance.

Exactly the same analysis applies to the problem of coordinating other dimensions of output. The different intermediate outputs not only have to be timed together, they have to fit together, and to be of the same quality. In addition to a timing price, the market would have to determine prices for "fitting" and for "quality". Again, multilateral negotiations are required, and penalties for fitting and for quality would have to be set by the affected complementary factors together. And again, uncertainty-- and the difficulty of insuring suppliers against deviations not under their control militates against agreement on the appropriate penalties.

Coordination through market negotiations therefore involves

two sets of "extra" costs: the first is the costs of multilateral negotiations, and the second is the costs of either self-insurance, of assuming the risks of paying the timing, fitting and quality penalties, or the increased risks to all the suppliers due to the moral hazard problem if the suppliers mutually insure themselves by agreeing to lower penalties. Each of these sets of costs will act like a tax on competitive factors allocated to complementary uses. The equilibrium allocation of factors among different uses will therefore reflect this tax, and as a consequence the market will under-allocate factors to uses where the degree of interdependence is large, relative to the allocation which would prevail if timing could be priced costlessly.

Now, each of these sets of costs are in the nature of "transactions" costs; as such, they are not inherent in the process of production but will vary, depending on the way in which that process is organized. Our next purpose is to show how these particular transactions costs may be avoided if the buyer of the complementary factors "integrates" them into a firm, i.e., becomes their employer.

Recall our definition of the firm in terms of the particular contractual relation between employer and employee. Under that contract, the employer obtains the right to direct production. To do so, he must guarantee the employee's wages, in return for which he obtains the assignment of the firm's profits. This amounts to the offer of a comprehensive insurance policy to suppliers (potential employees) and there is some fixed wage sufficiently high which will persuade any number of them to accept.

Suppliers agree to become subordinates in a hierarchy. Their superiors prescribe the procedures to be followed, and the employees are paid a fixed

wage provided they adhere to the specified procedures. Clearly, in order to make the system effective, employees must be supervised to ensure that the prescribed procedures are carried out, and there must be penalties which are sufficiently high to deter disobedience.

It is easy to see how such firms can coordinate production better than the market. First, within the firm, the problem of multilateral negotiations is lessened, both by the presence of a central agent-- the employer-- to the bargaining process, and, more importantly, by the non-specific nature of the contract[15] made with that central agent: the right of direction (within limits, of course) is vested in the employer. And since the employee's wage is not tied to the outcome of production, he is indifferent on this score-- and therefore, less likely to bargain-- as to which procedures he is instructed to follow.[16]

Second, the contractual relation is one where the employer provides insurance to the employee, since the employer assumes the risk of quantity or quality variations in the employees' production. Employees are paid a fixed wage, and are not responsible for the outcomes of their decisions. They are responsible only for adhering to specified procedures. Recall that the obstacle to market provision of this insurance is the moral hazard factor. This problem can be overcome whenever the insurer can specify a set of procedures which minimizes the probability of "hazard", and can monitor the actions of the insured to check that these procedures are followed. This is exactly what is achieved through the hierarchical organization of production. The employer will take external effects into account in prescribing the procedures to be followed by his employees. For example, with respect to the problem of timing, rules are prescribed

governing hours of work, and which limit the discretion of employees with respect to their choice of procedures, so that complementary intermediate inputs are produced at regular rates. Thus problems of "hazard" --in this case, damages to complementary inputs from irregularities in supplies-- are minimized. And hierarchical organization, unlike that of the market, facilitates the monitoring of the actions of employees rather than their output (the consequences of their actions) to ensure that specified procedures are in fact carried out.

What is the characteristic disadvantage of hierarchy? It follows from the fact that the incentives facing employees are not tied to their output. Being paid a fixed wage, they have no incentive except to follow the rules. How serious a limitation this is depends, roughly speaking, on how comprehensively the rules may be specified by the manager (employer), i.e., on how routine the tasks of employees may be made to be. To make this notion more precise, note first that in general, any technology may be defined as a set of procedures for getting useful output from inputs. A task is a subset of these, assigned to an individual or a department (the user of the technology). A task may be made relatively routine if the productivity of user discretion may be made to be relatively low, and this will be the case whenever the appropriate procedures may be specified so that relatively few "exceptions" arise in the course of production, i.e. relatively few situations arise where the user could obtain more output by independently choosing or modifying the procedures. The technique of auto assembly is an obvious example of a routine technology. Other technologies (medicine, social work) are much less routine: in these cases, procedures for getting the maximum output from inputs may not be specified without

reference to the actual circumstances of production, and the productivity of user discretion is relatively high.

What permits the technology of production to be "routinized" within the firm? Clearly, the division of labour itself facilitates it: the subdivision of tasks into a series of simple, narrow, and repetitive sub-tasks obviously permits a more complete specification of the procedures to be followed in the performance of each sub-task than would be possible for the the whole operation. Adam Smith's pinmaking factory exemplifies this process of routinization through the division of labour.

On the other hand, specialization according to the principle of comparative advantage, which plays no role in Smith's famout discussion, could diminish the efficiency of routinization to the extent that it implies that subordinates would be more expert at what they do than their superiors, and therefore that procedures should be less well-specified in a hierarchy. These two processes of the division of labour are not at all the same. Routinization is facilitated by, and indeed in one sense is just another way of describing the process of the division of labour in Smith's sense. As discussed in more detail elsewhere,[17] the extent to which the division of labour may be further increased, and tasks made more routine is also related to the level of cumulative experience with a technology, and to factor skills (human capital). Routinization is facilitated by experience with a technology, and (since the productivity of discretion depends on the skill with which it is exercised) when factor skills are relatively low.

Now, where the employer can formulate sets of procedures for his subordinates so that their jobs are perfectly routine (there are no

exceptional or ambiguous cases) hierarchy is perfectly efficient, and will surely displace market reltions in the organization of production. The problem of coordination is solved because hierarchical rules are more efficient for this purpose than contractual stipulations; and the lack of incentives for employee discretion under this metering system is, in this case, an unimportant defect since discretion is unnecessary (its productivity is low) when tasks are routine.

It might be thought that the appropriate formulation of routine procedures would also lower the costs of market organization, since the routine procedures may be alternatively used as contractual stipulations. Moreover, if routine procedures yield a stable level of output, market coordination would also be facilitated by the institution of these procedures as contractual stipulations. However, there is no means by which the observance of these procedures can be enforced, except by direct observation of the inputs, and the comparative advantage of hierarchy over market is obvious in this respect. The enforcement of procedures for inputs simultaneously yields a more stable level of output, and provides insurance to the employees, and neither of these functions can be fulfilled in the absence of direct input observation, i.e. of internal integration. Consequently, an increase in the division of labour, i.e., an increase in the simplicity, repetitiveness and narrowness of tasks not only facilitates routinization, but increases the likelihood that the tasks will be internally integrated.

Where the technology of work cannot be made perfectly routine, however, efficient production requires that employees exercise their discretion appropriately, modifying, adapting, or simply disregarding "the rules" as this is warranted by the changing circumstances of production. But no incentive is provided by the firm's metering system for employees to

perform in this way. To be sure, such incentives can be incorporated into hierarchical incentive systems, and rewards to employees for the productive exercise of discretion are a feature of most real-world hierarchies. But the characteristic advantages of hierarchical organization disappear precisely to the extent that the jobs of employees are allowed to become less routinized. Those hierarchical organizations which are the least "bureaucratized" are also those where management cannot keep track of the activities of its employees-- because they are diverse-- and which are the most plagued by external effects and lack of coordination between departments.

To illustrate, one well-known hierarchical means of coordinating discretionary activity is the following: employees are rewarded for discretion (by bonuses, promotion, etc.) but direct authorization and approval is required from superiors before action is taken. The authorization insures the employee and effects central coordination. However, compared with the market, the price of this coordination is inevitable delays and red tape; compared with the routine hierarchical system, the price is the sacrifice of the scale economies of the "automatic" coordination achieved by rules, since each discretionary action must be monitored and considered individually.

At higher levels in the hierarchy, the productivity of discretion increases progressively, and the firm's incentive system inevitably comes to more closely resemble that of the market (e.g. divisional responsibilities, profit sharing) as the risks of decisions are progressively shifted to employees. The problem of coordination becomes progressively more important, but the possibilities are still superior to those of the market

through the use of "coinsurance" institutions such as "limited liability" (managers are ordinarily not held liable for outcomes deemed beyond their control),[18] joint authorization, and committees.

Where activities are separable, the choice of integrating another activity or purchasing its intermediate output may therefore be characterized as follows. The firm chooses an incentive system for employees in that activity which yields the efficient level of discretion, one which minimizes the total costs of production, including the losses from excessive routinization of subordinates' time on the one hand and lack of coordination on the other, and including the direct costs (monitoring, evaluating, etc.) of operating the control system.[19] These costs are compared to the costs of purchasing and, if necessary, modifying inputs from independent suppliers. The costs of integrating additional activities must ultimately increase.[20] The firm will therefore continue to integrate activities until the marginal costs of doing so are equal to those of purchasing the input.

Having chosen the efficient internal incentive system for an activity, the firm compares those costs with the costs of market purchase. For any activity, these costs will depend on the relative magnitude of the comparative advantage of market vs hierarchy in organizing its production. The comparative advantage of hierarchy is superior coordination; that of the market is the incentive provided by payment according to output to each supplier to use discretion to choose procedures which minimize costs. The relative growth of markets or hierarchical relations will be facilitated, depending on the costs of devices or institutions which alleviate the characteristic disadvantage of one or the other organizational form. Thus, market transactions will be relatively cheap when market institutions arise

which alleviate the coordination problem. One such device, used extensively in the English metal trades during the nineteenth century, is the employment of skilled "fitters", who adjusted the different parts of guns and other implements until they fitted together as required. This allowed firms to acquire the semi-finished parts from independent craftsmen, since it obviated the need for direct coordination between the specialist fabricator of each part.[21] The subsequent development of inter-changeable parts made the fitters obsolete, and further facilitated market transactions rather than hierarchy.

The contemporary equivalent of the "fitter" is the hierarchical "coordinator" who directly monitors the production of complementary outputs simultaneously checking for quality and fit, and redirecting factors as needed so that rates of production may be timed appropriately. The use of coordinators obviously requires and facilitates the integration of activities within firms, rather than facilitating market transactions, as did the fitters.

Market solutions exist for the coordination of fit and quality, especially in the case of raw materials, through standardized weights and measures, and in other cases, through guarantees of quality,[22] and "client" relationships with customers.[23] Where these coordination problems can be solved through the market, and the timing problem can be alleviated through the holding of buffer stocks, firms will be able to purchase intermediate inputs on the market more cheaply than they can produce them themselves. On the other hand, for relatively specialized inputs, or where guarantees are costly, e.g. because of customer moral hazard[24] buffer stocks can facilitate hierarchical integration. Thus, where the only type of discretion of importance is effort (as measured by the number of units

produced per unit time), buffer stocks can be combined with piece rates within the firm, and coordination is achieved while providing the appropriate incentive to employees to maximize effort.[25]

The primitive but general model presented here therefore does yield a number of implications with respect to the relative incidence of hierarchical vs. market relations, including many of those which have been presented elsewhere on the basis of more elaborate, but less general analyses. For example, firms will be more likely to integrate activities, ceteris paribus, the larger the costs of measuring outputs relative to the internal costs of monitoring inputs,[26] the smaller the productivity of discretion,[27] and the larger the costs of storing buffer stocks,[28] and, more generally, of ensuring a steady supply of inputs.[29] Firms will also tend to integrate activities more, the greater the risk aversion of suppliers (and hence will be more likely to integrate personnel whose capital is specific to an activity rather than general,[30]) or, alternatively, the smaller the wage differential required to compensate for increased routinization of work.

4. Conclusion

Economists are typically impressed by the "automatic" functioning of the market mechanism. The ability of firms and other hierarchical organizations to "automatically" elicit correct and predictable responses from their employees, if considered, is usually attributed to the stake of employees in the profits of the enterprise, yet to most employees in

organizations of any size, the profits of the firm are a pure public good, and cannot be a motivation in their decisions. I have tried to show that the central motivation in hierarchies is provided by rules, and appropriate incentives to ensure their use. These enable firms to coordinate the activities of their employees as "automatically" as the market.

Moreover, market organization is inefficient under certain conditions, namely where factors are complements, and the number of suppliers is small. To summarize that argument, efficient production requires that all of the relevant dimensions of each supplier's output be internalized into the incentive system by which suppliers are paid, so that (1) the costs of deviations are taken into account in each supplier's production decisions, but (2) suppliers are at the same time held not liable for deviations which arise due to circumstances beyond their control. The complexity of the contractual stipulations required to internalize all the relevant dimensions into a payment system based on an index of suppliers' output, the need to do this by means of multilateral negotiations among all the complementary suppliers, and the problem of insuring suppliers against variations in the quality and quantity of output not under their control all militate against the organization of production through market relations under these conditions.

Corresponding to each of these defects of the market is an advantage of hierarchy. Hierarchical rules substitute for contractual stipulations: the rules may be set unilaterally by hierarchical authority, rather than negotiated; the payment to suppliers of a fixed wage given their obedience to hierarchical rules provides insurance to the employees; and the specification of standard procedures and the accompanying low

monitoring costs solves the moral hazard problem associated with the provision of insurance.

Where jobs cannot be made routine through the specification of standard procedures, the characteristic defect of hierarchy assumes importance: no incentive can be provided for employees to use their discretion in cases where standard procedures are clearly inappropriate, without reducing the coordinating advantage of the hierarchical form. I have tried to show that the relative efficiency of market or hierarchy under these conditions depends on the respective sizes of the comparative advantage of hierachy (coordination) vs that of the market ("automatic" incentives to use discretion) and on the relative costs of institutions or practices which alleviate the characteristic disadvantage of each organizational form. A number of plausible implications, including many derived elsewhere, are consistent with this simple model of the choice between market and hierarchy.

FOOTNOTES

1. A. Alchian and H. Demsetz, "Production, Information Costs and Economic Organization", American Economic Review, (Dec. 1972).

2. O. Williamson, Markets and Hierarchies, New York Free Press, 1975.

3. O. Williamson, op. cit.

4. R. H. Coase, "The Nature of the Firm", Economica, (Nov. 1937). Reprinted in AEA, Readings in Price Theory, eds. Stigler and Boulding. All references are to the AEA reprint. F.H. Knight, Risk, Uncertainty and Profit (New York, 1965). G.J. Stigler, "The Division of Labour is Limited by the Extent of the Market" Journal of Political Economy, June, 1951. E. A. G. Robinson, The Structure of Competitive Industry, Cambridge: Cambridge University Press, 1958.

5. R. H. Coase, op. cit., p. 337.

6. F. H. Knight, op. cit., pp. 269-70.

7. Loc. cit.

8. O. E. Williamson, M. L. Wachter, and J. E. Harris, "Understanding The Employment Relation: The Analysis of Idiosyncratic Exchange." The Bell Journal of Economics, Spring, 1975.

9. G. J. Stigler, op. cit.

10. Alchian, A., and Allen, W. R.,University Economics, Belmont, Cal. Wadsworth, 1967

11. Recent theorists neglect it altogether. Coase discusses the possibility of grounding the theory of the firm in the division of labour, but then dismisses it.

12. E. A. G. Robinson, op. cit., especially p. 40.

13. If it were otherwise, firms would not have to enforce discipline. For a discussion of the managerial problem of enforcing discipline in the era of the Industrial Revolution, see S. Pollard, The Genesis

13. of Modern Management London: George Allen and Unwin, 1965

14. The requirement that the number of supplies be small does not necessarily imply that competition is imperfect. This requirement is perfectly consistent with the competitive condition that the existing suppliers are price-takers, so that, were any of them to raise the price of his intermediate output above the competitive level, new suppliers would enter the industry in the long run. We rule out monopolistic pricing behaviour-- which it is well known, can motivate vertical integration-- in order to focus on other forces which lead to it.

15. Williamson has emphasized this argument. See O. E. Williamson, op. cit.

16. Bargaining need not disappear, of course, and it will not if employees, for whatever reason, find some procedures more distasteful than others.

17. See R. Wintrobe, The Economics of Bureaucracy, University of Western Ontario Research Report #7714 (1977). For a contemporary model of the division of labour along Smithian lines, see R. A. Ippolito, "The Division of Labour in the Firm", Economic Inquiry, 1977

18. This point is made by Arrow. See K. J. Arrow, "Control in Large Organizations", Management Science, Vol. 10 1963-64.

19. This model is presented in detail in Wintrobe, op. cit.

20. A proof of this proposition may be found in Wintrobe, op. cit.

21. I am indebted to C. K. Harley for this for this example.

22. On guarantees, see G. Akerlof, "The Market for Lemons" Quarterly Journal of Economics, v. 84 (1970) and M. Darby and E. Karni, "Free Competition and the Optimal Amount of Fraud", Journal of Law and Economics, v. 67 (1973).

23. Darby and Karni, op. cit.

24. Loc. cit.

25. Stiglitz discusses the choice between piece rates and time rates, and some of his results parallel those here. See J. E. Stiglitz, "Incentives, Risk and Information: Notes Towards a Theory of Hierarchy" Bell Journal of Economics, v. 6 (1975)

26. Darby and Karni, op. cit. suggest that measurement problems can motivate internal integration. See also J. McManus, The Cost of Alternative Economic Organization", Canadian Journal of Economics, 1976.

27. Some empirical proxies for this variable are suggested in Wintrobe, op. cit. References to the use of similar concepts by organization theorists may also be found there.

28. See Williamson, op. cit.

29. Models of vertical integration based on this problem are K. J. Arrow, "Vertical Integration and Communication", Bell Journal of Economics, v. 6, 1975. and W. Oi, and A. P. Hunter, jr., Economics of Private Truck Transportation, William C. Brown Co., 1965.

30. On specific capital, see G. S. Becker, Human Capital, New York: National Bureau of Economic Research, 1964. For the argument that risk aversion is implied by specific capital investments, see D. Gordon, "A Neoclassical Theory of Keynesian Unemployment", Economic Inquiry, v. 12, 1974.

APPENDIX

A-1

Assume for simplicity that factors produce no useful output separately, but that each factor yields some intermediate output, X (which may or may not be observable). The intermediate outputs together yield useful final output when combined via a production function.

The quantity of intermediate output, X, produced by each factor is determined by his method of production. Assume that only two methods are available, one where output is stochastic, and one where it is certain. Use of the first method (P1) yields X with relative frequency p and 0 with relative frequency (1-p). The average output is

$$E(P1) = p(X) + (1-p)0 = pX$$

Method (P2) is provisionally defined to have the same expectation as P1. The use of P2 therefore yields pX units of intermediate output in all states of the world.

To show how complementary among the intermediate outputs of the factors can affect the optimal choice of method, assume that the production function is of the C.E.S. variety. The parameter β in that function is an indicator of the degree of complementarity, as measured by the elasticity of substitution σ, since $\sigma = \frac{1}{1+\beta}$. If both factors use method P2, the expected output in all states of the world is

$$\begin{aligned} E(Q_2^{-\beta}) &= [a(pX)^{-\beta} + (1-a)(pX)^{-\beta}] \\ &= a(pX)^{-\beta} + (pX)^{-\beta} - a(pX)^{-\beta} \\ &= (pX)^{-\beta} \end{aligned}$$

Hence, since $Q_2^{-\beta} = (pX)^{-\beta}$, $Q_2 = pX$, and

(3) $E(Q_2) = pX$ if both factors use P2

If both factors use method P1 , then the outcomes and their respective relative frequencies are

$Q_1^{-\beta}$	Q_1	Relative Frequencies
$a(X)^{-\beta}+(1-a)(X)^{-\beta}$	$[a(X)^{-\beta}+(1-a)(X)^{-\beta}]^{-\frac{1}{\beta}}$	p^2
$a(X)^{-\beta}$	$a^{-\frac{1}{\beta}} X$	$p(1-p)$
$(1-a)(X)^{-\beta}$	$(1-a)^{-\frac{1}{\beta}} (X)$	$p(1-p)$
0	0	$(1-p)^2$

Hence $E(Q_1) = p^2[a(X)^{-\beta} + (1-a)(X)^{-\beta}]^{-\frac{1}{\beta}}$

$$+ p(1-p)[a^{-\frac{1}{\beta}} X + (1-a)^{-\frac{1}{\beta}} X]$$

$$= p^2 X + p(1-p)[a^{-\frac{1}{\beta}} X + (1-a)^{-\frac{1}{\beta}} X]$$

Now examine what happens to the ratio $\frac{E(Q_1)}{E(Q_2)}$ as the parameter β is changed.

$$(4) \quad \frac{E(Q_1)}{E(Q_2)} = p + (1-p)[a^{-\frac{1}{\beta}} + (1-a)^{-\frac{1}{\beta}}]$$

If $\beta = -1$, $\frac{E(Q_1)}{E(Q_2)} = 1$

i.e. $E(Q_1) = E(Q_2)$. But as $\beta \to -1$, $\sigma \to \infty$, i.e., the intermediate outputs of the two factors are perfect substitutes. If $\beta > -1$, the two factors are complements, and, from equation (4), $E(Q_1) < E(Q_2)$. As the degree of complementarity increases equation (4) shows that the expected output using method P1 falls continuously while the expected output using method P2 is unchanged.