ou Louis PUTTERMAN (ed)

The Economic Nature of The Firm : A Reader

(Cambridge Univ. Press, 1986)

CHAPTER 1

From *The Wealth of Nations*

ADAM SMITH

Adam Smith (1723–90) was born in Kirkcaldy, Scotland. He received the Master of Arts degree from the University of Glasgow in 1740, and subsequently was Professor of Moral Philosophy at that University from 1752 to 1763, and Commissioner of Customs for Scotland from 1778 to 1790. *The Wealth of Nations* established Smith as the founding figure in classical political economy.

Of the division of labour

(From book I, chapter 1)

The greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity, and judgment with which it is any where directed, or applied, seem to have been the effects of the division of labour.

The effects of the division of labour, in the general business of society, will be more easily understood, by considering in what manner it operates in some particular manufactures. It is commonly supposed to be carried furthest in some very trifling ones; not perhaps that it really is carried further in them than in others of more importance: but in those trifling manufactures which are destined to supply the small wants of but a small number of people, the whole number of workmen must necessarily be small; and those employed in every different branch of the work can often be collected into the same workhouse, and placed at once under the view of the spectator. In those great manufactures, on the contrary, which are destined to supply the great wants of the great body of the people, every different branch of the work employs so great a number of workmen, that it is impossible to collect them all into the same workhouse. We can seldom see more, at one time, than those employed in one single branch. Though in such manufactures, therefore, the work may really be divided into a much greater number of parts, than in those of a more trifling nature, the division is not near so obvious, and has accordingly been much less observed.

From Adam Smith, *An Inquiry into the Nature and Causes of The Wealth of Nations*, originally published 1776. Excerpted from the Modern Library Edition, edited by Edwin Cannan. New York, 1937.

To take an example, therefore,[1] from a very trifling manufacture; but one in which the division of labour has been very often taken notice of, the trade of the pin-maker; a workman not educated to this business (which the division of labour has rendered a distinct trade), nor acquainted with the use of the machinery employed in it (to the invention of which the same division of labour has probably given occasion), could scarce, perhaps, with his utmost industry, make one pin in a day, and certainly could not make twenty. But in the way in which this business is now carried on, not only the whole work is a peculiar trade, but it is divided into a number of branches, of which the greater part are likewise peculiar trades. One man draws out the wire, another straights it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on, is a peculiar business, to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations, which, in some manufactories, are all performed by distinct hands, though in others the same man will sometimes perform two or three of them. I have seen a small manufactory of this kind where ten men only were employed, and where some of them consequently performed two or three distinct operations. But though they were very poor, and therefore but indifferently accommodated with the necessary machinery, they could, when they exerted themselves, make among them about twelve pounds of pins in a day. There are in a pound upwards of four thousand pins of a middling size. Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of forty-eight thousand pins, might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently, and without any of them having been educated to this particular business, they certainly could not each of them have made twenty, perhaps not one pin in a day; that is, certainly, not the two hundred and fortieth, perhaps not the four thousand eight hundredth part of what they are at present capable of performing, in consequence of a proper division and combination of their different operations.

In every other art and manufacture, the effects of the division of labour are similar to what they are in this very trifling one; though, in many of them, the labour can neither be so much subdivided, nor reduced to so great a simplicity of operation. The division of labour, however, so far as it can be introduced, occasions, in every art, a proportionable increase of the productive powers of labour. The separation of different trades and employments from one another,

[1] Another and perhaps more important reason for taking an example like that which follows is the possibility of exhibiting the advantages of division of labour in statistical form.

seems to have taken place, in consequence of this advantage. This separation too is generally carried furthest in those countries which enjoy the highest degree of industry and improvement; what is the work of one man in a rude state of society, being generally that of several in an improved one.

. . .

This great increase of the quantity of work, which, in consequence of the division of labour, the same number of people are capable of performing, is owing to three different circumstances; first, to the increase of dexterity in every particular workman; secondly, to the saving of the time which is commonly lost in passing from one species of work to another; and lastly, to the invention of a great number of machines which facilitate and abridge labour, and enable one man to do the work of many.

First, the improvement of the dexterity of the workman necessarily increases the quantity of the work he can perform; and the division of labour, by reducing every man's business to some one simple operation, and by making this operation the sole employment of his life, necessarily increases very much the dexterity of the workman. A common smith, who, though accustomed to handle the hammer, has never been used to make nails, if upon some particular occasion he is obliged to attempt it, will scarce, I am assured, be able to make above two or three hundred nails in a day, and those too very bad ones. A smith who has been accustomed to make nails, but whose sole or principal business has not been that of a nailer, can seldom with his utmost diligence make more than eight hundred or a thousand nails in a day. I have seen several boys under twenty years of age who had never exercised any other trade but that of making nails, and who, when they exerted themselves, could make, each of them, upwards of two thousand three hundred nails in a day. The making of a nail, however, is by no means one of the simplest operations. The same person blows the bellows, stirs or mends the fire as there is occasion, heats the iron, and forges every part of the nail: In forging the head too he is obliged to change his tools. The different operations into which the making of a pin, or of a metal button, is subdivided, are all of them much more simple, and the dexterity of the person, of whose life it has been the sole business to perform them, is usually much greater. The rapidity with which some of the operations of those manufacturers are performed, exceeds what the human hand could, by those who had never seen them, be supposed capable of acquiring.

Secondly, the advantage which is gained by saving the time commonly lost in passing from one sort of work to another, is much greater than we should at first view be apt to imagine it. It is impossible to pass very quickly from one kind of work to another, that is carried on in a different place, and with quite different tools. A country weaver, who cultivates a small farm, must

lose a good deal of time in passing from his loom to the field, and from the field to his loom. When the two trades can be carried on in the same workhouse, the loss of time is no doubt much less. It is even in this case, however, very considerable. A man commonly saunters a little in turning his hand from one sort of employment to another. When he first begins the new work he is seldom very keen and hearty; his mind, as they say, does not go to it, and for some time he rather trifles than applies to good purpose. The habit of sauntering and of indolent careless application, which is naturally, or rather necessarily acquired by every country workman who is obliged to change his work and his tools every half hour, and to apply his hand in twenty different ways almost every day of his life; renders him almost always slothful and lazy, and incapable of any vigorous application even on the most pressing occasions. Independent, therefore, of his deficiency in point of dexterity, this cause alone must always reduce considerable the quantity of work which he is capable of performing.

Thirdly, and lastly, every body must be sensible how much labour is facilitated and abridged by the application of proper machinery. It is unnecessary to give any example. I shall only observe, therefore, that the invention of all those machines by which labour is so much facilitated and abridged, seems to have been originally owing to the division of labour. Men are much more likely to discover easier and readier methods of attaining any object, when the whole attention of their minds is directed towards that single object, than when it is dissipated among a great variety of things. But in consequence of the division of labour, the whole of every man's attention comes naturally to be directed towards some one very simple object. It is naturally to be expected, therefore, that some one or other of those who are employed in each particular branch of labour should soon find out easier and readier methods of performing their own particular work, wherever the nature of it admits of such improvement. A great part of the machines made use of in those manufactures in which labour is most subdivided, were originally the inventions of common workmen, who, being each of them employed in some very simple operation, naturally turned their thoughts towards finding out easier and readier methods of performing it. Whoever has been much accustomed to visit such manufactures, must frequently have been shewn very pretty machines, which were the inventions of such workmen, in order to facilitate and quicken their own particular part of the work. In the first fire-engines, a boy was constantly employed to open and shut alternately the communication between the boiler and the cylinder, according as the piston either ascended or descended. One of those boys, who loved to play with his companions, observed that, by tying a string from the handle of the valve which opened this communication to another part of the machine, the valve would open and shut without his assis-

tance, and leave him at liberty to divert himself with his play-fellows. One of the greatest improvements that has been made upon this machine, since it was first invented, was in this manner the discovery of a boy who wanted to save his own labour.

All the improvements in machinery, however, have by no means been the inventions of those who had occasion to use the machines. Many improvements have been made by the ingenuity of the makers of the machines, when to make them became the business of a peculiar trade; and some by that of those who are called philosophers or men of speculation, whose trade it is not to do any thing, but to observe every thing; and who, upon that account, are often capable of combining together the powers of the most distant and dissimilar objects. In the progress of society, philosophy or speculation becomes, like every other employment, the principal or sole trade and occupation of a particular class of citizens. Like every other employment too, it is subdivided into a great number of different branches, each of which affords occupation to a peculiar tribe or class of philosophers; and this subdivision of employment in philosophy, as well as in every other business, improves dexterity, and saves time. Each individual becomes more expert in his own peculiar branch, more work is done upon the whole, and the quantity of science is considerably increased by it.

It is the great multiplication of the productions of all the different arts, in consequence of the division of labour, which occasions, in a well-governed society, that universal opulence which extends itself to the lowest ranks of the people. Every workman has a great quantity of his own work to dispose of beyond what he himself has occasion for; and every other workman being exactly in the same situation, he is enabled to exchange a great quantity of his own goods for a great quantity, or, what comes to the same thing, for the price of a great quantity of theirs. He supplies them abundantly with what they have occasion for, and they accommodate him as amply with what he has occasion for, and a general plenty diffuses itself through all the different ranks of the society.

Observe the accommodation of the most common artificer or day labourer in a civilized and thriving country, and you will perceive that the number of people of whose industry a part, though but a small part, has been employed in procuring him this accommodation, exceeds all computation. The woollen coat, for example, which covers the day-labourer, as coarse and rough as it may appear, is the produce of the joint labour of a great multitude of workmen. The shepherd, the sorter of wool, the wool-comber or carder, the dyer, the scribbler, the spinner, the weaver, the fuller, the dresser, with many others, must all join their different arts in order to complete even this homely production. How many merchants and carriers, besides, must have been em-

ployed in transporting the materials from some of those workmen to others who often live in a very distant part of the country! how much commerce and navigation in particular, how many ship-builders, sailors, sail-makers, rope-makers, must have been employed in order to bring together the different drugs made use of by the dyer, which often come from the remotest corners of the world! What a variety of labour too is necessary in order to produce the tools of the meanest of those workmen! To say nothing of such complicated machines as the ship of the sailor, the mill of the fuller, or even the loom of the weaver, let us consider only what a variety of labour is requisite in order to form that very simple machine, the shears with which the shepherd clips the wool. The miner, the builder of the furnace for smelting the ore, the feller of the timber, the burner of the charcoal to be made use of in the smelting-house, the brick-maker, the brick-layer, the workmen who attend the furnace, the mill-wright, the forger, the smith, must all of them join their different arts in order to produce them. Were we to examine, in the same manner, all the different parts of his dress and household furniture, the coarse linen shirt which he wears next his skin, the shoes which cover his feet, the bed which he lies on, and all the different parts which compose it, the kitchengrate at which he prepares his victuals, the coals which he makes use of for that purpose, dug from the bowels of the earth, and brought him to perhaps by a long sea and a long land carriage, all the other utensils of his kitchen, all the furniture of his table, the knives and forks, the earthen or pewter plates upon which he serves up and divides his victuals, the different hands employed in preparing his bread and his beer, the glass window which lets in the heat and the light, and keeps out the wind and the rain, with all the knowledge and art requisite for preparing that beautiful and happy invention, without which these northern parts of the world could scarce have afforded a very comfortable habitation, together with the tools of all the different workmen employed in producing those different conveniences; if we examine, I say, all these things, and consider what a variety of labour is employed about each of them, we shall be sensible that without the assistance and co-operation of many thousands, the very meanest person in a civilized country could not be provided, even according to, what we very falsely imagine, the easy and simple manner in which he is commonly accommodated. Compared, indeed, with the more extravagant luxury of the great, his accommodation must no doubt appear extremely simple and easy; and yet it may be true, perhaps that the accommodation of an European prince does not always so much exceed that of an industrious and frugal peasant, as the accommodation of the latter exceeds that of many an African king, the absolute master of the lives and liberties of ten thousand naked savages.

. . .

From The Wealth of Nations

That the division of labour is limited by the extent of the market

(From book I, chapter 3)

As it is the power of exchanging that gives occasion to the division of labour, so the extent of this division must always be limited by the extent of that power, or in other words, by the extent of the market. When the market is very small, no person can have any encouragement to dedicate himself entirely to one employment, for want of the power to exchange all that surplus part of the produce of his own labour, which is over and above his own consumption, for such parts of the produce of other men's labour as he has occasion for.

There are some sorts of industry, even of the lowest kind, which can be carried on no where but in a great town. A porter, for example, can find employment and subsistence in no other place. A village is by much too narrow a sphere for him; even an ordinary market town is scarce large enough to afford him constant occupation. In the lone houses and very small villages which are scattered about in so desert a country as the Highlands of Scotland, every farmer must be butcher, baker and brewer for his own family. In such situations we can scarce expect to find even a smith, a carpenter, or a mason, within less than twenty miles of another of the same trade. The scattered families that live at eight or ten miles distance from the nearest of them, must learn to perform themselves a great number of little pieces of work, for which, in more populous countries, they would call in the assistance of those workmen. Country workmen are almost every where obliged to apply themselves to all the different branches of industry that have so much affinity to one another as to be employed about the same sort of materials. A country carpenter deals in every sort of work that is made of wood: a country smith in every sort of work that is made of iron. The former is not only a carpenter, but a joiner, a cabinet maker, and even a carver in wood, as well as a wheelwright, a ploughwright, a cart and wagon maker. The employments of the latter are still more various. It is impossible there should be such a trade as even that of the nailer in the remote and inland parts of the Highlands of Scotland. Such a workman at the rate of a thousand nails a day, and three hundred working days in the year, will make three hundred thousand nails in the year. But in such a situation it would be impossible to dispose of one thousand, that is, of one day's work in the year.

As by means of water-carriage a more extensive market is opened to every sort of industry than what land-carriage alone can afford it, so it is upon the sea-coast, and along the banks of navigable rivers, that industry of every kind naturally begins to subdivide and improve itself, and it is frequently not till a

long time after that those improvements extend themselves to the inland parts of the country.

. . .

On joint stock companies

(From book V, chapter 1)

Joint stock companies, established either by royal charter or by act of parliament, differ in several respects, not only from regulated companies, but from private copartneries.

First, in a private copartnery, no partner, without the consent of the company, can transfer his share to another person, or introduce a new member into the company. Each member, however, may, upon proper warning, withdraw from the copartnery, and demand payment from them of his share of the common stock. In a joint stock company, on the contrary, no member can demand payment of his share from the company; but each member can, without their consent, transfer his share to another person, and thereby introduce a new member. The value of a share in a joint stock is always the price which it will bring in the market; and this may be either greater or less, in any proportion, than the sum which its owner stands credited for in the stock of the company.

Secondly, in a private copartnery, each partner is bound for the debts contracted by the company to the whole extent of his fortune. In a joint stock company, on the contrary, each partner is bound only to the extend of his share.

The trade of a joint stock company is always managed by a court of directors. This court, indeed, is frequently subject, in many respects, to the controul of a general court of proprietors. But the greater part of those proprietors seldom pretend to understand any thing of the business of the company; and when the spirit of faction happens not to prevail among them, give themselves no trouble about it, but receive contentedly such half yearly or yearly dividend, as the directors think proper to make to them. This total exemption from trouble and from risk, beyond a limited sum, encouraged many people to become adventurers in joint stock companies, who would, upon no account, hazard their fortunes in any private copartnery. Such companies, therefore, commonly draw to themselves much greater stocks than any private copartnery can boast of. The trading stock of the South Sea Company, at one time, amounted to upwards of thirty-three millions eight hundred thousand pounds. The dividend capital of the Bank of England amounts, at present, to ten millions seven hundred and eighty thousand pounds. The directors of such companies, however, being the managers rather of other people's money than of their own, it cannot well be expected, that they should watch over it with

the same anxious vigilance with which the partners in a private copartnery frequently watch over their own. Like the stewards of a rich man, they are apt to consider attention to small matters as not for their master's honour, and very easily give themselves a dispensation from having it. Negligence and profusion, therefore, must always prevail, more or less, in the management of the affairs of such a company. It is upon this account that joint stock companies for foreign trade have seldom been able to maintain the competition against private adventurers. They have, accordingly, very seldom succeeded without an exclusive privilege; and frequently have not succeeded with one. Without an exclusive privilege they have commonly mismanaged the trade. With an exclusive privilege they have both mismanaged and confined it.

. . .

But a joint stock company, consisting of a small number of proprietors, with a moderate capital, approaches very nearly to the nature of the private copartnery, and may be capable of nearly the same degree of vigilance and attention.

. . .

On education

(From book V, chapter 1)

In the progress of the division of labour, the employment of the far greater part of those who live by labour, that is, of the great body of the people, comes to be confined to a few very simple operations, frequently to one or two. But the understandings of the greater part of men are necessarily formed by their ordinary employments. The man whose whole life is spent in performing a few simple operations, of which the effects too are, perhaps, always the same, or very nearly the same, has no occasion to exert his understanding, or to exercise his invention in finding out expedients for removing difficulties which never occur. He naturally loses, therefore, the habit of such exertion, and generally becomes as stupid and ignorant as it is possible for a human creature to become. The torpor of his mind renders him, not only incapable of relishing or bearing a part in any rational conversation, but of conceiving any generous, noble, or tender sentiment, and consequently of forming any just judgment concerning many even of the ordinary duties of private life. Of the great and extensive interests of his country he is altogether incapable of judging; and unless very particular pains have been taken to render him otherwise, he is equally incapable of defending his country in war. The uniformity of his stationary life naturally corrupts the courage of his mind, and makes him regard with abhorrence the irregular, uncertain, and

adventurous life of a soldier. It corrupts even the activity of his body, and renders him incapable of exerting his strength with vigour and perseverance, in any other employment than that to which he has been bred. His dexterity at his own particular trade seems, in this manner, to be acquired at the expense of his intellectual, social, and marital virtues. But in every improved and civilized society this is the state into which the labouring poor, that is, the great body of the people, must necessarily fall, unless government takes some pains to prevent it.

It is otherwise in the barbarous societies, as they are commonly called, of hunters, of shepherds, and even of husbandmen in that rude state of husbandry which precedes the improvement of manufactures, and the extension of foreign commerce. In such societies the varied occupations of every man oblige every man to exert his capacity, and to invent expedients for removing difficulties which are continually occurring. Invention is kept alive, and the mind is not suffered to fall into that drowsy stupidity, which, in a civilized society, seems to benumb the understanding of almost all the inferior ranks of people. In those barbarous societies, as they are called, every man, it has already been observed, is a warrior. Every man too is in some measure a statesman, and can form a tolerable judgment concerning the interest of the society, and the conduct of those who govern it. How far their chiefs are good judges in peace, or good leaders in war, is obvious to the observation of almost every single man among them. In such a society indeed, no man can well acquire that improved and refined understanding, which a few men sometimes possess in a more civilized state. Though in a rude society there is a good deal of variety in the occupations of every individual, there is not a good deal in those of the whole society. Every man does, or is capable of doing, almost every thing which any other man does, or is capable of doing. Every man has a considerable degree of knowledge, ingenuity, and invention; but scarce any man has a great degree. The degree, however, which is commonly possessed, is generally sufficient for conducting the whole simple business of the society. In a civilized state, on the contrary, though there is little variety in the occupations of the greater part of individuals, there is an almost infinite variety in those of the whole society. These varied occupations present an almost infinite variety of objects to the contemplation of those few, who, being attached to no particular occupation themselves, have leisure and inclination to examine the occupations of other people. The contemplation of so great a variety of objects necessarily exercises their minds in endless comparisons and combinations, and renders their understandings, in an extraordinary degree, both acute and comprehensive. Unless those few, however, happen to be placed in some very particular situations, their great abilities, though honourable to themselves, may contribute very little to the good government or happiness of their society. Notwithstanding the great abilities of those few,

all the nobler parts of the human character may be, in a great measure, obliterated and extinguished in the great body of the people.

The education of the common people requires, perhaps, in a civilized and commercial society, the attention of the public more than that of people of some rank and fortune. People of some rank and fortune are generally eighteen or nineteen years of age before they enter upon that particular business, profession, or trade, by which they propose to distinguish themselves in the world. They have before that full time to acquire, or at least to fit themselves for afterwards acquiring, every accomplishment which can recommend them to the public esteem, or render them worthy of it.

. . .

It is otherwise with the common people. They have little time to spare for education. Their parents can scarce afford to maintain them even in infancy. As soon as they are able to work, they must apply to some trade by which they can earn their subsistence. That trade too is generally so simple uniform as to give little exercise to the understanding; while, at the same time, their labour is both so constant and so severe, that it leaves them little leisure and less inclination to apply to, or even to think of any thing else.

But though the common people cannot, in any civilized society, be so well instructed as people of some rank and fortune, the most essential parts of education, however, to read, write, and account, can be acquired at so early a period of life, that the greater part even of those who are to be bred to the lowest occupations, have time to acquire them before they can be employed in those occupations. For a very small expense the public can facilitate, can encourage, and can even impose upon almost the whole body of the people, the necessity of acquiring those most essential parts of education.

. . .

CHAPTER 5

The nature of the firm

RONALD COASE

Ronald Coase was born in Middlesex, England in 1910. He received a Bachelor's (B. Com.) degree from the University of London in 1932, and a Doctor of Science degree in economics from the same school in 1951. When this article was published, he was Assistant Lecturer at the London School of Economics. He is currently Professor Emeritus of Economics and Senior Fellow in Law and Economics at the University of Chicago Law School.

. . .

I

It is convenient if, in searching for a definition of a firm, we first consider the economic system as it is normally treated by the economist. Let us consider the description of the economic system given by Sir Arthur Salter.[1] "The normal economic system works itself. For its current operation it is under no central control, it needs no central survey. Over the whole range of human activity and human need, supply is adjusted to demand, and production to consumption, by a process that is automatic, elastic and responsive." An economist thinks of the economic system as being co-ordinated by the price mechanism and society becomes not an organization but an organism.[2] The economic system "works itself." This does not mean that there is no planning by individuals. These exercise foresight and choose between alternatives. This is necessarily so if there is to be order in the system. But this theory assumes that the direction of resources is dependent directly on the price mechanism. Indeed, it is often considered to be an objection to economic planning that it merely tries to do what is already done by the price mechanism.[3] Sir Arthur Salter's description, however, gives a very incomplete picture of our economic system. Within a firm, the description does not fit at all. For instance, in economic theory we find that the allocation of factors of production between different uses is determined by the price mechanism. The price of factor A becomes higher in X than in Y. As a result, A moves from Y to X until the difference between the prices in X and Y, except in so far as it compensates for other differential advantages, disappears. Yet in the real world, we find that there are many areas where this does not apply. If a workman moves from department Y to department X, he does not go because of a change in relative prices, but because he is ordered to do so. Those who object to economic planning on the grounds that the problem is solved by price movements can be answered by pointing out that there is planning within our economic system which is quite different from the individual planning mentioned above and which is akin to what is normally called economic planning. The example given above is typical of a large sphere in our modern economic system. Of course, this fact has not been ignored by economists. Marshall introduces organization as a fourth factor of production; J. B. Clark gives the co-ordinating function to the entrepreneur; Professor Knight introduces managers who co-ordinate. As D. H. Robertson points out, we find "islands of conscious power in this ocean of unconscious co-operation like lumps of butter coagulating in a pail of butter-milk."[4] But in view of the fact that it is usually argued that co-ordination will be done by the price mechanism, why is such organization necessary? Why are there these "island of conscious power"? Outside the firm, price movements direct production, which is co-ordinated through a series of exchange transactions on the market. Within a firm, these market transactions are eliminated and in place of the complicated market structure with exchange transactions is substituted the entrepreneur-co-ordinator, who directs production.[5] It is clear that these are alternative methods of co-ordinating production. Yet having regard to the fact that if production is regulated by price movements, production could be carried on without any organization at all, well might we ask, why is there any organization?

Of course, the degree to which the price mechanism is superseded varies greatly. In a department store, the allocation of the different sections to the various locations in the building may be done by the controlling authority or it may be the result of competitive price bidding for space. In the Lancashire cotton industry, a weaver can rent power and shop-room and can obtain looms and yarn on credit.[6] This co-ordination of the various factors of production is, however, normally carried out without the intervention of the price mechanism. As is evident, the amount of "vertical" integration, involving as it does the supersession of the price mechanism, varies greatly from industry to industry and from firm to firm.

Reprinted with minor abridgements from Ronald Coase, "The Nature of the Firm," *Economica*, 4 (1937): 386–405, by permission of the publisher.

[1] This description is quoted with approval by D. H. Robertson (1930), p. 85, and by Arnold Plant (1932). It appears in *Allied Shipping Control*, pp. 16–17.

[2] See F. A. Hayek (1933).

[3] See F. A. Hayek (1933).

[4] D. H. Robertson (1930) p. 85.

[5] In the rest of this paper I shall use the term entrepreneur to refer to the person or persons who, in a competitive system, take the place of the price mechanism in the direction of resources.

[6] *Survey of Textile Industries*, p. 26.

It can, I think, be assumed that the distinguishing mark of the firm is the supersession of the price mechanism. It is, of course, as Professor Robbins points out, "related to an outside network of relative prices and costs,"[7] but it is important to discover the exact nature of this relationship. This distinction between the allocation of resources in a firm and the allocation in the economic system has been very vividly described by Mr. Maurice Dobb when discussing Adam Smith's conception of the capitalist:

> . . . It began to be seen that there was something more important than the relations inside each factory or unit captained by an undertaker; there were the relations of the undertaker with the rest of the economic world outside his immediate sphere . . . the undertaker busies himself with the division of labour inside each firm and he plans and organises consciously,

but

> . . . he is related to the much larger economic specialisation of which he himself is merely one specialised unit. Here, he plays his part as a single cell in a larger organism, mainly unconscious of the wider role he fills.[8]

In view of the fact that while economists treat the price mechanism as a co-ordinating instrument, they also admit the co-ordinating function of the "entrepreneur," it is surely important to inquire why co-ordination is the work of the price mechanism in one case and of the entrepreneur in another. The purpose of this paper is to bridge what appears to be a gap in economic theory between the assumption (made for some purposes) that resources are allocated by means of the price mechanism and the assumption (made for other purposes) that this allocation is dependent on the entrepreneur-coodinator. We have to explain the basis on which, in practice, this choice between alternatives is effected.[9]

II

Our task is to attempt to discover why a firm emerges at all in a specialized exchange economy. The price mechanism (considered purely from the side of

[7] L. Robbins (1932) p. 71.

[8] Maurice Dobb (1925), p. 20 Cf., also, Henderson, (1922) pp. 3–5.

[9] It is easy to see when the State takes over the direction of an industry that, in planning it, it is doing something which was previously done by the price mechanism. What is usually not realized is that any business man in organizing the relations between his departments is also doing something which could be organized through the price mechanism. There is therefore point in Mr. Durbin's answer to those who emphasize the problems involved in economic planning that the same problems have to be solved by business men in the competitive system. (See Durbin, 1936). The important difference between these two cases is that economic planning is imposed on industry while firms arise voluntarily because they represent a more efficient method of organizing production. In a competitive system, there is an "optimum" amount of planning!

the direction of resources) might be superseded if the relationship which replaced it was desired for its own sake. This would be the case, for example, if some people preferred to work under the direction of some other person. Such individuals would accept less in order to work under someone, and firms would arise naturally from this. But it would appear that this cannot be a very important reason, for it would rather seem that the opposite tendency is operating if one judges from the stress normally laid on the advantage of "being one's own master."[10] Of course, if the desire was not to be controlled but to control, to exercise power over others, then people might be willing to give up something in order to direct others; that is, they would be willing to pay others more than they could get under the price mechanism in order to be able to direct them. But this implies that those who direct pay in order to be able to do this and are not paid to direct, which is clearly not true in the majority of cases.[11] Firms might also exist if purchasers preferred commodities which are produced by firms to those not so produced; but even in spheres where one would expect such preferences (if they exist) to be of negligible importance, firms are to be found in the real world.[12] Therefore there must be other elements involved.

The main reason why it is profitable to establish a firm would seem to be that there is a cost of using the price mechanism. The most obvious cost of "organizing" production through the price mechanism is that of discovering what the relevant prices are.[13] This cost may be reduced but it will not be eliminated by the emergence of specialists who will sell this information. The costs of negotiating and concluding a separate contract for each exchange transaction which takes place on a market must also be taken into account.[14] Again, in certain markets, e.g., produce exchanges, a technique is devised for minimizing these contract costs; but they are not eliminated. It is true that contracts are not eliminated when there is a firm but they are greatly reduced. A factor of production (or the owner thereof) does not have to make a series of contracts with the factors with whom he is co-operating within the firm, as would be necessary, of course, if this co-operation were as a direct result of

[10] See Harry Dawes (1934), who instances "the trek to retail shopkeeping and insurance work by the better paid of skilled men due to the desire (often the main aim in life of a worker) to be independent" (p. 86).

[11] None the less, this is not altogether fanciful. Some small shopkeepers are said to earn less than their assistants.

[12] G. F. Shove (1933) p. 116, note I, points out that such preferences may exist, although the example he gives is almost the reverse of the instance given in the text.

[13] According to N. Kaldor (1934) it is one of the assumptions of static theory that "All the relevant prices are known to all individuals." But this is clearly not true of the real world.

[14] This influence was noted by Professor Usher when discussing the development of capitalism. He says: "The successive buying and selling of partly finished products were sheer waste of energy" (1921, p. 13). But he does not develop the idea nor consider why it is that buying and selling operations still exist.

the working of the price mechanism. For this series of contracts is substituted one. At this stage, it is important to note the character of the contract into which a factor enters that is employed within a firm. The contract is 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*.[15] The essence of the contract is that it should only state the limits to the powers of the entrepreneur. Within these limits, he can therefore direct the other factors of production.

There are, however, other disadvantages – or costs – of using the price mechanism. It may be desired to make a long-term contract for the supply of some article or service. This may be due to the fact that if one contract is made for a longer period, instead of several shorter ones, then certain costs of making each contract will be avoided. Or, owing to the risk attitude of the people concerned, they may prefer to make a long rather than a short-term contract. Now, owing to the difficulty of forecasting, the longer the period of the contract is for the supply of the commodity or service, the less possible, and indeed, the less desirable it is for the person purchasing to specify what the other contracting party is expected to do. It may well be a matter of indifference to the person supplying the service or commodity which of several courses of action is taken, but not to the purchaser of that service or commodity. But the purchaser will not know which of these several courses he will want the supplier to take. Therefore, the service which is being provided is expressed in general terms, the exact details being left until a later date. All that is stated in the contract is the limits to what the persons supplying the commodity or service are expected to do. The details of what the supplier is expected to do are not stated in the contract but are decided later by the purchaser. When the direction of resources (within the limits of the contract) becomes dependent on the buyer in this way, that relationship which I term a "firm" may be obtained.[16] A firm is likely therefore to emerge in those cases where a very short term contract would be unsatisfactory. It is obviously of more importance in the case of services – labor – than it is in the case of the buying of commodities. In the case of commodities, the main items can be stated in advance and the details which will be decided later will be of minor significance.

We may sum up this section of the argument by saying that the operation of a market costs something and by forming an organization and allowing

[15] It would be possible for no limits to the powers of the entrepreneur to be fixed. This would be voluntary slavery. According to Professor Batt (1929) p. 18, such a contract would be void and unenforceable.

[16] Of course, it is not possible to draw a hard and fast line which determines whether there is a firm or not. There may be more or less direction. It is similar to the legal question of whether there is the relationship of master and servant or principal and agent. See the discussion of this problem presented later.

some authority (an "entrepreneur") to direct the resources, certain marketing costs are saved. The entrepreneur has to carry out his function at less cost, taking into account the fact that he may get factors of production at a lower price than the market transactions which he supersedes, because it is always possible to revert to the open market if he fails to do this.

The question of uncertainty is one which is often considered to be very relevant to the study of the equilibrium of the firm. It seems improbable that a firm would emerge without the existence of uncertainty. But those, for instance, Professor Knight, who make the *mode of payment* the distinguishing mark of the firm – fixed incomes being guaranteed to some of those engaged in production by a person who takes the residual, and fluctuating, income – would appear to be introducing a point which is irrelevant to the problem we are considering. One entrepreneur may sell his services to another for a certain sum of money, while the payment to his employees may be mainly or wholly a share in profits.[17] The significant question would appear to be why the allocation of resources is not done directly by the price mechanism.

Another factor that should be noted is that exchange transactions on a market and the same transactions organized within a firm are often treated differently by Governments or other bodies with regulatory powers. If we consider the operation of a sales tax, it is clear that it is a tax on market transactions and not on the same transactions organized within the firm. Now since these are alternative methods of "organization" – by the price mechanism or by the entrepreneur – such a regulation would bring into existence firms which otherwise would have no *raison d'être*. It would furnish a reason for the emergence of a firm in a specialized exchange economy. Of course, to the extent that firms already exist, such a measure as a sales tax would merely tend to make them larger than they would otherwise be. Similarly, quota schemes, and methods of price control which imply that there is rationing, and which do not apply to firms producing such products for themselves, by allowing advantages to those who organize within the firm and not through the market, necessarily encourage the growth of firms. But it is difficult to believe that it is measures such as have been mentioned in this paragraph which have brought firms into existence. Such measures would, however, tend to have this result if they did not exist for other reasons.

These, then, are the reasons why organizations such as firms exist in a specialized exchange economy in which it is generally assumed that the distribution of resources is "organized" by the price mechanism. A firm, therefore, consists of the system of relationships which comes into existence when the direction of resources is dependent on an entrepreneur.

The approach which has just been sketched would appear to offer an ad-

[17] The views of Professor Knight are examined later in more detail.

vantage in that it is possible to give a scientific meaning to what is meant by saying that a firm gets larger or smaller. A firm becomes larger as additional transactions (which could be exchange transactions co-ordinated through the price mechanism) are organized by the entrepreneur and becomes smaller as he abandons the organization of such transactions. The question which arises is whether it is possible to study the forces which determine the size of the firm. Why does the entrepreneur not organize one less transaction or one more? It is interesting to note that Professor Knight considers that:

> . . . the relation between efficiency and size is one of the most serious problems of theory, being, in contrast with the relation for a plant, largely a matter of personality and historical accident rather than of intelligible general principles. But the question is peculiarly vital because the possibility of monopoly gain offers a powerful incentive to *continuous and unlimited* expansion of the firm, which force must be offset by some equally powerful one making for decreased efficiency (in the production of money income) with growth in size, if even boundary competition is to exist.[18]

Professor Knight would appear to consider that it is impossible to treat scientifically the determinants of the size of the firm. On the basis of the concept of the firm developed above, this task will now be attempted.

It was suggested that the introduction of the firm was due primarily to the existence of marketing costs. A pertinent question to ask would appear to be (quite apart from the monopoly considerations raised by Professor Knight), why, if by organizing one can eliminate certain costs and in fact reduce the cost of production, are there any market transactions at all?[19] Why is not all production carried on by one big firm? There would appear to be certain possible explanations.

First, as a firm gets larger, there may be decreasing returns to the entrepreneur function, that is, the costs of organizing additional transactions within the firm may rise.[20] Naturally a point must be reached where the costs of organizing an extra transaction within the firm are equal to the costs involved in carrying out the transaction in the open market, or, to the costs of organizing by another entrepreneur. Secondly, it may be that as the transactions which are organized increase, the entrepreneur fails to place the factors of production in the uses where their value is greatest, that is, fails to make the best use of the factors of production. Again, a point must be reached where the loss through the waste of resources is equal to the marketing costs of the exchange

[18] Frank Knight (1933).

[19] There are certain marketing costs which could only be eliminated by the abolition of "consumers' choice" and these are the costs of retailing. It is conceivable that these costs might be so high that people would be willing to accept rations because the extra product obtained was worth the loss of their choice.

[20] This argument assumes that exchange transactions on a market can be considered as homogeneous; which is clearly untrue in fact. This complication is taken into account later.

transaction in the open market or to the loss if the transaction was organized by another entrepreneur. Finally, the supply price of one or more of the factors of production may rise, because the "other advantages" of a small firm are greater than those of a large firm.[21] Of course, the actual point where the expansion of the firm ceases might be determined by a combination of the factors mentioned above. The first two reasons given most probably correspond to the economists' phrase of "diminishing returns to management."[22]

The point has been made in the previous paragraph that a firm will tend to expand until the costs of organizing an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organizing in another firm. But if the firm stops its expansion at a point below the costs of marketing in the open market and at a point equal to the costs of organizing in another firm, in most cases (excluding the case of "combination"[23]), this will imply that there is a market transaction between these two producers, each of whom could organize it at less than the actual marketing costs. How is the paradox to be resolved? If we consider an example the reason for this will become clear. Suppose *A* is buying a product from *B* and that both *A* and *B* could organize this marketing transaction at less than its present cost. *B*, we can assume, is not organizing one process or stage of production, but several. If *A* therefore wishes to avoid a market transaction, he will have to take over all the processes of production controlled by *B*. Unless *A* takes over all the processes of production, a market transaction will still remain, although it is a different product that is bought. But we have previously assumed that as each producer expands he becomes less efficient; the additional costs of organizing extra transactions increase. It is probable that *A*'s cost of organizing the transactions previously organized by *B* will be greater than *B*'s cost of doing the same thing. *A* therefore will take over the whole of *B*'s organization only if his cost of organizing *B*'s work is not greater than *B*'s cost by an amount equal to the costs of carrying out an exchange transaction on the open market. But once it becomes economical to have a market transaction, it also pays to divide production in such a way that the cost of organizing an extra transaction in each firm is the same.

[21] For a discussion of the variation of the supply price of factors of production to firms of varying size, see E. A. G. Robinson (1931). It is sometimes said that the supply price of organizing ability increases as the size of the firm increases because men prefer to be the heads of small independent businesses rather than the heads of departments in a large business. See Jones (1927) p. 531, and Macgregor (1906) p. 63. This is a common argument of those who advocate Rationalization. It is said that larger units would be more efficient, but owing to the individualistic spirit of the smaller entrepreneurs, they prefer to remain independent, apparently in spite of the higher income which their increased efficiency under Rationalization makes possible.

[22] This discussion is, of course, brief and incomplete. For a more thorough discussion of this particular problem, see N. Kaldor (1934) and E. A. G. Robinson (1934).

[23] A definition of this term is given later.

Up to now it has been assumed that the exchange transactions which take place through the price mechanism are homogeneous. In fact, nothing could be more diverse than the actual transactions which take place in our modern world. This would seem to imply that the costs of carrying out exchange transactions through the price mechanism will vary considerably as will also the costs of organizing these transactions within the firm. It seems therefore possible that quite apart from the question of diminishing returns the costs of organizing these transactions within the firm may be greater than the costs of carrying out the exchange transactions in the open market. This would necessarily imply that there were exchange transactions carried out through the price mechanism, but would it mean that there would have to be more than one firm? Clearly not, for all those areas in the economic system where the direction of resources was not dependent directly on the price mechanism could be organized within one firm. The factors which were discussed earlier would seem to be the important ones, though it is difficult to say whether "diminishing returns to management" or the rising supply price of factors is likely to be the more important.

Other things being equal, therefore, a firm will tend to be larger:

(a) the less the costs of organizing and the slower these costs rise with an increase in the transactions organized;
(b) the less likely the entrepreneur is to make mistakes and the smaller the increase in mistakes with an increase in the transactions organized;
(c) the greater the lowering (or the less the rise) in the supply price of factors of production to firms of larger size.

Apart from variations in the supply price of factors of production to firms of different sizes, it would appear that the costs of organizing and the losses through mistakes will increase with an increase in the spatial distribution of the transactions organized, in the dissimilarity of the transactions, and in the probability of changes in the relevant prices.[24] As more transactions are organized by an entrepreneur, it would appear that the transactions would tend to be either different in kind or in different places. This furnishes an additional reason why efficiency will tend to decrease as the firm gets larger. Inventions which tend to bring factors of production nearer together, by lessening spatial distribution, tend to increase the size of the firm.[25] Changes like the telephone

[24] This aspect of the problem is emphasized by N. Kaldor (1934). Its importance in this connection had been previously noted by E. A. G. Robinson (1931) pp. 83–106. This assumes that an increase in the probability of price movements increases the costs of organizing within a firm more than it increases the cost of carrying out an exchange transaction on the market – which is probable.

[25] This would appear to be the importance of the treatment of the technical unit by E. A. G. Robinson (1931) pp. 27–33. The larger the technical unit, the greater the concentration of factors and therefore the firm is likely to be larger.

and the telegraph which tend to reduce the cost of organizing spatially will tend to increase the size of the firm. All changes which improve managerial technique will tend to increase the size of the firm.[26,27]

It should be noted that the definition of a firm which was given above can be used to give more precise meanings to the terms "combination" and "integration."[28] There is a combination when transactions which were previously organized by two or more entrepreneurs become organized by one. This becomes integration when it involves the organization of transactions which were previously carried out between the entrepreneurs on a market. A firm can expand in either or both of these two ways. The whole of the "structure of competitive industry" becomes tractable by the ordinary technique of economic analysis.

III

The problem which has been investigated in the previous section has not been entirely neglected by economists and it is now necessary to consider why the reasons given above for the emergence of a firm in a specialized exchange economy are to be preferred to the other explanations which have been offered.

It is sometimes said that the reason for the existence of a firm is to be found in the division of labor. This is the view of Professor Usher, a view which has been adopted and expanded by Mr. Maurice Dobb. The firm becomes

> . . . the result of an increasing complexity of the division of labour. . . . The growth of this economic differentiation creates the need for some integrating force without which differentiation would collapse into chaos; and it is as the integrating force in a differentiated economy that industrial forms are chiefly significant.[29]

[26] It should be noted that most inventions will change both the costs of organizing and the costs of using the price mechanism. In such cases, whether the invention tends to make firms larger or smaller will depend on the relative effect on these two sets of costs. For instance, if the telephone reduces the costs of using the price mechanism more than it reduces the costs of organizing, then it will have the effect of reducing the size of the firm.

[27] An illustration of these dynamic forces is furnished by Maurice Dobb (1928) p. 68. "With the passing of bonded labour the factory, as an establishment where work was organised under the whip of he overseer, lost its *raison d'être* until this was restored to it with the introduction of power machinery after 1846." It seems important to realize that the passage from the domestic system to the factory system is not a mere historical accident, but is conditioned by economic forces. This is shown by the fact that it is possible to move from the factory system to the domestic system, as in the Russian example, as well as vice versa. It is the essence of serfdom that the price mechanism is not allowed to operate. Therefore, there has to be direction from some organizer. When, however, serfdom passed, the price mechanism was allowed to operate. It was not until machinery drew workers into one locality that it paid to supersede the price mechanism and the firm again emerged.

[28] This is often called "vertical integration," combination being termed "lateral integration."

[29] Maurice Dobb (1928) p. 10. Professor Usher's views are to be found in Usher (1921) pp. 1–18.

The answer to this argument is an obvious one. The "integrating force in a differentiated economy" already exists in the form of the price mechanism. It is perhaps the main achievement of economic science that it has shown that there is no reason to suppose that specialization must lead to chaos.[30] The reason given by Mr. Maurice Dobb is therefore inadmissible. What has to be explained is why one integrating force (the entrepreneur) should be substituted for another integrating force (the price mechanism).

The most interesting reasons (and probably the most widely accepted) which have been given to explain this fact are those to be found in Professor Knight's *Risk, Uncertainty and Profit*. His views will be examined in some detail.

Professor Knight starts with a system in which there is no uncertainty:

> . . . acting as individuals under absolute freedom but without collusion men are supposed to have organised economic life with the primary and secondary division of labour, the use of capital, etc., developed to the point familiar in present-day America. The principal fact which calls for the exercise of the imagination is the internal organisation of the productive groups or establishments. With uncertainty entirely absent, every individual being in possession of perfect knowledge of the situation, there would be no occasion for anything of the nature of responsible management or control of productive activity. Even marketing transactions in any realistic sense would not be found. The flow of raw materials and productive services to the consumer would be entirely automatic.[31]

Professor Knight says that we can imagine this adjustment as being "the result of a long process of experimentation worked out by trial-and-error methods alone," while it is not necessary "to imagine every worker doing exactly the right thing at the right time in a sort of 'pre-established harmony' with the work of others. There might be managers, superintendents, etc., for the purpose of co-ordinating the activities of individuals," though these managers would be performing a purely routine function, "without responsibility of any sort."[32]

Professor Knight then continues:

> With the introduction of uncertainty – the fact of ignorance and the necessity of acting upon opinion rather than knowledge – into this Eden-like situation, its character is entirely changed. . . . With uncertainty present doing things, the actual execution of activity, becomes in a real sense a secondary part of life; the primary problem or function is deciding what to do and how to do it.[33]

This fact of uncertainty brings about the two most important characteristics of social organization.

[30] Cf. J. B. Clark (1900) p. 19, who speaks of the theory of exchange as being the "theory of the organisation of industrial society."

[31] Frank Knight (1921) p. 267. [32] Knight (1921) pp. 267–8. [33] Knight (1921) p. 268.

> In the first place, goods are produced for a market, on the basis of entirely impersonal prediction of wants, not for the satisfaction of the wants of the producers themselves. The producer takes the responsibility of forecasting the consumers' wants. In the second place, the work of forecasting and at the same time a large part of the technological direction and control of production are still further concentrated upon a very narrow class of producers, and we meet with a new economic functionary, the entrepreneur. . . . When uncertainty is present and the task of deciding what to do and how to do it takes the ascendancy over that of execution the internal organisation of the productive groups is no longer a matter of indifference or a mechanical detail. Centralisation of this deciding and controlling function is imperative, a process of "cephalisation" is inevitable.[34]

The most fundamental change is:

> . . . the 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. . . . With human nature as we know it it would be impracticable 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. . . . The result of this manifold specialisation of function is the enterprise and wage system of industry. Its existence in the world is the direct result of the fact of uncertainty.[35]

These quotations give the essence of Professor Knight's theory. The fact of uncertainty means that people have to forecast future wants. Therefore, you get a special class springing up who direct the activities of others to whom they give guaranteed wages. It acts because good judgment is generally associated with confidence in one's judgment.[36]

Professor Knight would appear to leave himself open to criticism on several grounds. First of all, as he himself points out, the fact that certain people have better judgment or better knowledge does not mean that they can only get an income from it by themselves actively taking part in production. They can sell advice or knowledge. Every business buys the services of a host of advisers. We can imagine a system where all advice or knowledge was bought as required. Again, it is possible to get a reward from better knowledge or judgment not by actively taking part in production but by making contracts with people who are producing. A merchant buying for future delivery represents an example of this. But this merely illustrates the point that it is quite possible to give a guaranteed reward providing that certain acts are performed without directing the performance of those acts. Professor Knight says that "with human nature as we know it it would be impracticable or very unusual for one man to guarantee to another a definite result of the latter's actions without

[34] Knight (1921) pp. 268–95. [35] Knight (1921) pp. 269–70. [36] Knight (1921) p. 270.

being given power to direct his work." This is surely incorrect. A large proportion of jobs are done to contract, that is, the contractor is guaranteed a certain sum providing he performs certain acts. But this does not involve any direction. It does mean, however, that the system of relative prices has been changed and that there will be a new arrangement of the factors of production.[37] The fact that Professor Knight mentions that the "second party would not place himself under the direction of the first without such a guarantee" is irrelevant to the problem we are considering. Finally, it seems important to notice that even in the case of an economic system where there is no uncertainty Professor Knight considers that there would be co-ordinators, though they would perform only a routine function. He immediately adds that they would be "without responsibility of any sort," which raises the question by whom are they paid and why? It seems that nowhere does Professor Knight give a reason why the price mechanism should be superseded.

. . .

V

Only one task now remains; and that is, to see whether the concept of a firm which has been developed fits in with that existing in the real world. We can best approach the question of what constitutes a firm in practice by considering the legal relationship normally called that of "master and servant" or "employer and employee."[38] The essentials of this relationship have been given as follows:

> (1) The servant must be under the duty of rendering personal services to the master or to others on behalf of the master, otherwise the contract is a contract for sale of goods or the like.
>
> (2) The master must have the right to control the servant's work, either personally or by another servant or agent. It is this right of control or interference, of being entitled to tell the servant when to work (within the hours of service) and when not to work, and what work to do and how to do it (within the terms of such service) which is the dominant characteristic in this relation and marks off the servant from an independent contractor, or from one employed merely to give to his employer the fruits of his labour. In the latter case, the contractor or performer is not under the employer's control in doing the work or effecting the service; he has to shape and manage his work so as to give the result he has contracted to effect.[39]

We thus see that it is the fact of direction which is the essence of the legal concept of "employer and employee," just as it was in the economic concept which was developed above. It is interesting to note that Professor Batt says further:

> That which distinguishes an agent from a servant is not the absence or presence of a fixed wage or the payment only of commission on business done, but rather the freedom with which an agent may carry out his employment.[40]

We can, therefore, conclude that the definition we have given is one which approximates closely to the firm as it is considered in the real world.

Our definition is, therefore, realistic. Is it manageable? This ought to be clear. When we are considering how large a firm will be the principle of marginalism works smoothly. The question always is, will it pay to bring an extra exchange transaction under the organizing authority? At the margin, the costs of organizing within the firm will be equal either to the costs of organizing in another firm or to the costs involved in leaving the transaction to be "organized" by the price mechanism. Business men will be constantly experimenting, controlling more or less, and in this way, equilibrium will be maintained. This gives the position of equilibrium for static analysis. But it is clear that the dynamic factors are also of considerable importance, and an investigation of the effect changes have on the cost of organizing within the firm and on marketing costs generally will enable one to explain why firms get larger and smaller. We thus have a theory of moving equilibrium. The above analysis would also appear to have clarified the relationship between initiative or enterprise and management. Initiative means forecasting and operates through the price mechanism by the making of new contracts. Management proper merely reacts to price changes, rearranging the factors of production under its control. That the business man normally combines both functions is an obvious result of the marketing costs which were discussed above. Finally, this analysis enables us to state more exactly what is meant by the "marginal product" of the entrepreneur. But an elaboration of this point would take us far from our comparatively simple task of definition and clarification.

[37] This shows that it is possible to have a private enterprise system without the existence of firms. Though, in practice, the two functions of enterprise, which actually influences the system of relative prices by forecasting wants and acting in accordance with such forecasts, and management, which accepts the system of relative prices as being given, are normally carried out by the same persons, yet it seems important to keep them separate in theory. This point is further discussed later.

[38] The legal concept of "employer and employee" and the economic concept of a firm are not identical, in that the firm may imply control over another person's property as well as over their labor. But the identity of these two concepts is sufficiently close for an examination of the legal concept to be of value in appraising the worth of the economic concept.

[39] Batt (1929), p. 6. [40] Batt (1929) p. 7.

amount of assets not invested in the first period; in our case, by the range of elements included in the set X (area of acceptance).

9. Conclusion

We have constructed a model that incorporates rational grounds for the choice by two individuals between an employment contract and a contract of the ordinary kind (which we have called a sales contract). By a generalization of this model we are able to account for the fact that in an employment contract certain aspects of the worker's behavior are stipulated in the contract terms, certain other aspects are placed within the authority of the employer, and still other aspects are left to the worker's choice. Since administrative theory has been interested in explaining behavior within the framework of employment relations, and economic theory in explaining behavior within the area of market relations, the model suggests one possible way of relating these two bodies of theory. The most serious limitations of the model lie in the assumptions of rational utility-maximizing behavior incorporated in it.

CHAPTER 9

Production, information costs, and economic organization

ARMEN ALCHIAN and HAROLD DEMSETZ

Armen Alchian was born in 1914 in Fresno, California. He received a Ph.D. in economics at Stanford University in 1943. When this article was published, he was Professor of Economics at the University of California, Los Angeles, where he is currently Professor Emeritus and John M. Olin Distinguished Scholar in the Graduate School of Management.

Harold Demsetz was born in Chicago, Illinois in 1930. He received a Ph.D. in economics at Northwestern University in 1959. This article was written during his last months at the University of Chicago and his first months at the University of California, Los Angeles, where he has been Professor of Economics since 1971.

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

Reprinted with minor abridgements from Armen Alchian and Harold Demsetz, "Production, Information Costs, and Economic Organization," *The American Economic Review*, 62 (1972): 777–95, by permission of the editors.

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 customers? 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

[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.

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 system 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?

[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 (1963).

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.

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 $Z = aX_i^2$ 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]

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.

[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 (1970).

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 exchanges 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 produc-

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

tion) 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 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 mar-

[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.

ket competitive process, revised teams with greater production 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] 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

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

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 a 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 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 nonconcentralized 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

[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 (1969) and by H. B. Malmgren (1961).

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 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 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 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 payments based on gross sales in retail stores, come close to profit sharing . However, it is gross output sharing rather than profit shar-

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]

ing. We are unable to specify the implications of the difference. We refer the reader to S. N. Cheung (1969).

[13] Incidentally, investment activity will be changed. The inability to capitalize the investment value as "take-home" private 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 payoffs if the alternative of investing is paying out the firm's income to employees to take home and use as private property. For a development of this proposition, see the papers by Eirik Furubotn and Svetozar Pejovich (1974) and by Pejovich (1969).

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 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 of 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 stability 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]

[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,

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 capitalized into present wealth of stockholders. (As if to make more difficult that competition by new would-be monitors, multiple 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 en-

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 (1967). 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.

deavor where more shirking is desired and where redirected use 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 non-pecuniary, 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 non-shirking 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]

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.

[15] *Sports Leagues:* Professional sports contests among teams are 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 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 reserves 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 (1970) and formulated by Michael Canes (1970)). 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]

The similarity between the preceding analysis and the question of absentee landlordism and of sharecropping arrangements is no accident. The same fac-

[16] Professional athletes in baseball, football, and basketball, where athletes have 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 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 athletes' future services 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.

tors 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

[17] The analysis used by Cheung (1969) 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.

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 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 inputs 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

[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.

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 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 enhancing 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?

CHAPTER 10

Understanding the employment relation: the analysis of idiosyncratic exchange

OLIVER WILLIAMSON, MICHAEL WACHTER, and JEFFREY HARRIS

Oliver Williamson was born in Superior, Wisconsin in 1932. He received the Ph.D. in economics at Carnegie-Mellon University in 1963. When this paper was published, he was Professor of Economics, Law, and Public Policy at the University of Pennsylvania. He is currently Gordon Tweedy Professor of Economics of Law and Organization at Yale University.

Michael Wachter was born in 1943 in New York City. He received the Ph.D. from Harvard University in 1970. When this paper was published, he was Associate Professor of Economics at the University of Pennsylvania, where he is currently Professor of Economics, Law and Management and Director of the Institute for Law and Economics.

Jeffrey Harris was born in 1948. He earned an M.D. and a Ph.D. in economics from the University of Pennsylvania in 1974 and 1975, respectively. At the time of publication of this paper, he was Resident in Medicine at Massachusetts General Hospital. He is currently Associate Professor of Economics at the Massachusetts Institute of Technology.

. . .

2. Background

A brief review of the literature. The internal labor market literature has its roots in the industrial relations – labor economics literature of the 1950s and early 1960s. The important contributions in this area include the work of Dunlop, Kerr, Livernash, Meij, Raimon, and Ross.[1] This work, which is descriptively oriented, has since been developed and extended by Doeringer and Piore.[2]

Reprinted with abridgements from Oliver Williamson, Michael Wachter, and Jeffrey Harris, "Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange," *The Bell Journal of Economics*, 6 (1975): 250–78. (A variant of this paper also appears as Chapter 4 of Williamson's *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: The Free Press, 1975.)

[1] See Dunlop (1957, 1958), Kerr (1954), Livernash (1957), Meij (1963), Raimon (1953), and Ross (1958).

[2] Although acknowledging some of the efficiency aspects of the internal labor market, Doeringer and Piore also stress nonneoclassical attributes. Subsequent work in this tradition moves even further away from an efficiency orientation. Harrison (1972), Piore (1973), Thurow (1971),