THE CONTRACTUAL NATURE OF THE FIRM*

STEVEN N. S. CHEUNG
University of Hong Kong

ALMOST half a century has elapsed since R. H. Coase wrote "The Nature of the Firm."[1] The impact of this paper is now increasing.[2] At about twenty years of age and before receiving a bachelor's degree from the London School of Economics, Coase conceived the thesis of that work during a traveling scholarship to the United States in 1931–32.[3] Consider-

* For their helpful comments I am grateful to Armen A. Alchian, Yoram Barzel, Keith Leffler, John S. McGee, and Dean Worcester.

[1] Ronald H. Coase, The Nature of the Firm, 4 Economica 386 (1937). The paper, however, was written several years earlier (see note 3 *infra*).

[2] The *Social Sciences Citation Index* provides the following information. Breaking down the period of 1966–80 into three five-year periods, the total citations of Coase's "firm" paper for each subperiod are: 1966–70, 17 citations; 1971–75, 47 citations; and 1976–80, 105 citations.

[3] In personal correspondence Coase wrote me his recollections as follows:

About the firm. I spent the year 1931–1932 in the United States where I studied the problem of vertical and lateral integration in American industry. From pondering on these problems came my views on the nature of the firm. The basic idea in the article on the Nature of the Firm was certainly worked out by October, 1932 as a letter which I then wrote (and which has been preserved) shows. Also I completed a draft of the article by the Spring of 1934 while still in Dundee. As to why I waited until 1937 to publish, the reasons are various. I did not feel under any great pressure to publish it. I did in fact publish some other pieces before the firm article but the ideas in these papers were in fact developed after those on the firm. Remember that I was a young lecturer having to learn a lot in order to teach courses on subjects on which I knew very little and this took up a good deal of time. And I was interested in many other aspects of economics. I have never been one to rush into print—indeed I find it difficult to express exactly what I have in mind and have never thought that the world would suffer much if it didn't have the opportunity to read my views. I did in fact make some changes in the period between 1934 and 1937 but there was no change in my main position. But I did not feel any need to publish before I had expressed my thoughts as well as I could. Some of the projects on which I worked in the 1930s have still not been completed although I hope one day to do so.

About dates. As I completed the first year of University work while still at high school (not unusual at that time), I completed my University work and passed the degree examinations in 1931 (after 2 years at LSE). I was awarded a travelling scholarship by the University of London in 1931 and spent the next year in the United States. At that time, 3 years residence was required for the award of a degree but the regulations were loosely interpreted so as to allow my year in the United States to be counted as a year of residence

[*Journal of Law & Economics*, vol. XXVI (April 1983)]

ing the work started out as the equivalent of an undergraduate term paper one stands in awe of the insights that prompted it.

The issues surrounding what has been called "the firm" and the scope of economic behavior encompassed by Coase's thesis are far from resolved. At present, different versions abound of what Coase meant, along with varied criticisms.[4] I do not propose to evaluate this growing body of literature or to argue that my interpretation is necessarily an accurate presentation of Coase's position. Rather, once inspired by Coase's early work to do research on contracts, I wish to return and expound on that work in light of my own findings.

Anyone who investigates the economics of contractual arrangements will recognize that Coase's "firm" paper relates to the choice of contracts.[5] Yet Coase, aware of the significance and relevance of the measurement problem, suggested in 1969 that I look into the various measurements adopted in the timber industry. However, the difficulty of obtaining data blocked that attempt. By 1974 I had reached the view that piece-rate contracts offer a useful gateway to an understanding of the firm as an organization, because payment by the piece falls squarely in between the

---

at LSE. This explains why I was awarded my degree in 1932 although I had passed the examination in 1931. When I worked out my views on the nature of the firm, I was 20 or 21. A letter written in October 1932 (and the independent recollections of Duncan Black) indicate I had them by October 1932 and presumably I developed them earlier—between October 1931 and the summer of 1932. My date of birth is December 29th, 1910.

[4] For some examples which seem most directly related to Coase's work, see H. B. Malmgren, Information, Expectations and the Theory of the Firm, 75 Q. J. Econ. 399 (1961); G. B. Richardson, The Organization of Industry, 82 Econ. J. 883 (1972); Armen A. Alchian & Harold Demsetz, Production, Information Costs, and Economic Organization, 62 Am. Econ. Rev. 777 (1972); Paul H. Rubin, The Expansion of Firms, 81 J. Pol. Econ. 936 (1973); A. Michael Spence, The Economics of Internal Organization: An Introduction, 6 Bell. J. Econ. 163 (1975); John C. McManus, The Costs of Alternative Economic Organizations, 8 Can. J. Econ. 334 (1975); Oliver E. Williamson, Markets and Hierarchies (1975); Lloyd R. Cohen, The Firm: A Revised Definition, 46 S. Econ. J. 580 (1979); and Yoram Barzel, Measurement Costs and the Organization of Markets, 25 J. Law & Econ. 27 (1982). Each of these works has a different interpretation of Coase, and they all disagree with Coase in different ways. This lack of consensus leaves room for further interpretation, although in the area of transaction costs the lack of a standardized terminology makes it difficult to tell where the disagreements lie.

[5] See Steven N. S. Cheung, Transaction Costs, Risk Aversion, and the Choice of Contractual Arrangements, 12 J. Law & Econ. 23 (1969). With great reluctance I brought in risk sharing to explain the choice of share contracts in agriculture: several attempts made in 1968 to discard "risk" had not succeeded. Certain types of risk, such as the uncertainty of whether a promise will be honored or whether a product will live up to its advertisement, can be viewed as transaction costs. In such a case my preference is always for treating the problem in that light because testable implications are easier to derive. In agriculture, however, risk resulting from nature poses a different problem.

market and what Coase called the firm. My investigation of the piece-rate contract began in 1975 and continues.[6]

What I propose to do here, therefore, is to try to interpret Coase's arguments in light of my findings about contracts in general and about the piece-rate contract in particular. I shall then argue that we do not exactly know what the firm is—nor is it vital to know. The word "firm" is simply a shorthand description of a way to organize activities under contractual arrangements that differ from those of ordinary product markets.

I. The Firm as a Form of Contractual Arrangement

Coase's central thesis is that differences in the costs of operating institutions (transaction costs) lead to the emergence of a firm to supersede a market. On the one hand, market transactions involve products or commodities; on the other, "firm transactions" involve factors of production. The growth of a firm may then be viewed as the replacement of a product market by a factor market, resulting in a saving in transaction costs. This thesis is not easy to understand because Coase does not define "the firm"; nor, as we shall see, is there a clear distinction between a product market and a factor market.

Private ownership of productive inputs is assumed. Each input owner therefore has the option of (1) producing and marketing goods himself, (2) selling his input outright, or (3) entering into a contractual arrangement surrendering the use of his input to an agent in exchange for an income. The firm emerges with the third option: the entrepreneur or the agent who holds a limited set of use rights by contract directs production activities without immediate reference to the price of each activity, and the commodities so produced are then sold in the market. Herein lies the puzzle. If private property rights are absent and hence the above options are not available, it is relatively easy to understand why the activities of a worker or an input are directed by an agent instead of by market prices. But why should a private property owner *voluntarily* surrender his rights and be told what to do by a visible hand?

This choice, according to Coase, is made to reduce transaction costs.

[6] Joseph E. Stiglitz analyzes the piece-rate versus the wage contract in terms of risk and information, while explicitly stating that he ignores the costs of supervision. It appears to me that his "incentive and risk" problems can be treated as problems of transaction costs. In any event, the Stiglitz approach is quite different from that in the present paper. See Joseph E. Stiglitz, Incentives, Risk, and Information: Notes towards a Theory of Hierarchy, 6 Bell J. Econ. 552 (1975). Another interesting paper on the piece-rate contract, again based on a different orientation from mine, is found in Edward P. Lazear & Sherwin Rosen, Rank-Order Tournaments as Optimum Labor Contracts, 89 J. Pol. Econ. 841 (1981).

Some have suggested that such an argument is tautological. But it is not, because other reasons are also advanced for the emergence of the firm, including division of labor, risk, and the coordination of production activities. Coase considered these factors and rejected them all. To him, transaction costs constitute the prime consideration. His argument is subject to refutation because the list of other plausible factors makes it possible to conceive that total transaction costs might rise as the firm emerges.

Even if transaction costs constitute the only relevant factor, or if all other factors are forcibly interpreted as belonging in the domain of transaction costs, Coase's argument is still not tautological if one can identify different types of transaction costs and how they will vary under different circumstances. Generality in the extreme renders an argument tautological, whereas a total lack of general applicability renders an argument ad hoc. Testable implications are to be found somewhere in between, and how the costs are to be specified is a matter of choice, depending on the problem at hand.

An emphasis on transaction costs does not negate the potential gain from specialization through the division of labor or from more efficient coordination of productive efforts. Consider, for example, the classic "pin factory" in which each of the multiple input owners specializes by working on only one part. If all costs of transaction were zero, a customer buying a pin would make a separate payment to each of the many contributing to its production. Comparative advantage guides each to specialize in his own skill, and if it appears desirable to hire a coordinator of activities, the buyer of the pin will simply make an additional payment to him. In such a case, a large number of product prices would direct the production of the single pin.

In such a world it would be redundant to speak of a product market and a factor market. The two would be inseparable: the buyer would be paying simultaneously for the product and for the contribution of the input owner. To separate a product from a factor market requires that an agent pay input owners while receiving from customers payment for the products he hands out. Whereas the standard approach assumes the number of products as given, Coase's view is that this number is determinate only if transaction costs are explicitly incorporated into the analysis.

The problem can be stated more fundamentally. Any productive input is a *private* property if, within well-defined limits, its owner has (1) the right to exclude others so that he alone may decide on its use, (2) the right to extract exclusive income from its use, and (3) the right to transfer the property (including labor) to or to exchange with anyone he sees fit. The right to exchange implies the right to contract, and property rights may be transacted through a wide variety of contractual arrangements. When

these rights are exchanged it is, of course, for the purpose of yielding higher income to the owner, and the choice of contracts will be constrained by the costs of transaction.

Our main concern here is not the transaction itself but the contractual arrangements through which the right to use the input is delegated to another party and commodities produced by this use are sold to consumers. No outright transfer is made, and when the input owner retains some other rights the contract becomes a structured document.[7] A delimited set of use rights is surrendered in exchange for an income, under a form of contract that binds the input owner to follow directions instead of determining his own course by continual reference to the market prices of a variety of activities he may perform.

The surrender of use rights is a matter of degree, and the delimitation of the rights delegated is the subject of contracting, often supplemented by an implicit understanding, by custom, and by common law. It goes without saying that a clerk will not be asked to do work normally reserved for a janitor. The payment tendered is often based on a measured property (for example, hours per day) that bears no resemblance to the measured properties of the final commodities priced for sale. For this reason, observed market prices cannot directly guide the owner of the input to perform in the same manner as if every activity he performs were measured and priced. Hence the surrender of use rights often implies the delegation of the right to decide what to do.

With private property rights, economic analysis asserts as an axiom that when an input owner enters into a contract of the type described (joining the firm) he expects a gain relative to his other options, for he has had the option of not joining. What must be explained is why that increase occurs. Could it be that letting someone else make the decision is often more productive? The answer is no. We have noted that delegating the right to make a decision is the result of the difference in pricing and measuring properties. It cannot in general be the case that management decisions are superior to consumer decisions reached through the price mechanism. Errors are bound to be less frequent when price information guides every activity performed.

Could it be that specialization, coordination, and economy of scale achieved by pooling input resources from many owners will yield higher incomes for all, so that each chooses to join the firm? Again the answer is no. As noted earlier, if every activity is measured and priced, then benefits arising from specialization and coordination can be realized with-

[7] See Steven N. S. Cheung, The Structure of a Contract and the Theory of a Non-Exclusive Resource, 13 J. Law & Econ. 49 (1970).

out the "factor market"—the right to decide and use one's input need not be delegated to some agent or entrepreneur, because in a product market the input owner will receive a payment for every contribution.

Could the firm emerge because people shirk, cheat, or are opportunistic—as some recent theses imply?[8] Maybe. But the problem is that this sort of behavior is ubiquitous and will vary only in degree and in kind depending on the form of contract chosen or on how the property transacted is measured and priced. The behavior of a factory worker whose shirking requires the service of a monitor results from the worker's delegation of the right to use his labor. He would not shirk, or at least he would shirk differently, if for every small contribution he were paid a price.

Coase's answer is bold: "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 'organising' production through the price mechanism is that of *discovering what the relevant prices are*" (emphasis added).[9] He offers several reasons why the discovery and negotiation of prices are costly, although his explanation seems incomplete. I propose to offer here four general reasons, at least two of them derivatives of Coase's reasoning.

Perhaps the most obvious reason why the cost of discovering prices is higher in the absence of a firm is that significantly more transactions are required, each calling for a separate price. If the consumer were to pay for each contribution to or each component of the commodity, instead of for a single finished product, the cost would often be prohibitive. As an alternative, all cooperating input owners might contract with one another, each agreeing on a price for his services, and all the prices then might be aggregated for the final product. As a substitute for these contracts a central agent might contract with each input owner, paying him a price for the use rights rendered and selling the final product at another price. On this form of contract reduction Coase wrote: "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 the working of the price mechanism. For this series of contracts is substituted one."[10]

A second factor apparently not considered by Coase is the information

[8] See, in particular, Williamson, *supra* note 4; and Benjamin Klein, Robert G. Crawford, & Armen A. Alchian, Vertical Integration, Appropriable Rents, and the Competitive Contracting Process, 21 J. Law & Econ. 297 (1978).

[9] Coase, *supra* note 1, at 390.

[10] Coase, *supra* note 1, at 391.

cost of knowing a product. When parts or components of a product considered separately have no easily recognizable usage, agreement on price between producer and consumer of each and every part tends to be more costly than the whole product. Reaching agreement on the price of a spring inside a camera incurs a proportionately higher cost than does the camera. Although the consumer has the final say in assessing the worth of the whole product, he cannot be expected to recognize the value of each component part—he may not even know what some of them are or even that they exist. It simply costs too much to learn about everything in every commodity we buy. As we shall see, in piece-rate contracts the parts or components of a commodity are often directly measured and priced, but these negotiations are between specialist agents and input owners. That is, for a component which by itself has no readily identifiable value, agreement on price is less costly between specialists and input owners than it would be between input owners and consumers or between specialists and consumers. The one who produces component parts tends to know more about them than the one who consumes.

A third cost of discovering price is that of measurement. In every transaction, some characteristics or attributes must be measured, whether the deal is between an agent and a customer, an agent and an input owner, or an input owner and a customer. If the activities performed by an input owner change frequently, if these activities vary greatly, or if some of the activities to be performed cannot be conveniently stipulated in advance, it tends to be more economical to forgo any direct measurement of these activities and substitute another measurement to serve as a proxy. Thus an agent may hire a worker by the hour or lease a building by the square foot without going through the process of metering every contribution the input performs. The varied input activities which, in fact, are the source of product value are not priced at all because measurement would be too costly. The input owner receives payment for his varied contributions by measuring an attribute totally different from those of his actual contributions or of the final product sold to customers. The agent deals with two different sets of measurements, absorbing any gain or loss by directing and monitoring the performance of input owners and providing to consumers total commodities with specified characteristics.

Coase seemed to have in mind the use of a "proxy" measurement to avoid the cost of measuring greatly varied activities when he wrote the following:

> 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 person supplying the commodity or service is expected to do. The details of what the supplier is expected to do is not stated in the contract but is 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. . . . It is obviously of more importance in the case of services—labour—than it is in the case 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.[11]

Finally, the problem of separating contributions generates cost in reaching price agreement. When input owners work in collaboration, in some situations the contribution of each may not be easily delineated and each may claim more than he deserves. It is true that competition among input owners will reduce excessive claims, but the problem will not be eliminated. An agent hiring the collaborating participants may therefore exercise a right similar to that of eminent domain, offering a price for each on a take-it-or-leave-it basis by again measuring a proxy instead of the contribution itself.[12]

As examples of the gain from collaboration and the difficulty of delineating contributions, Alchian and Demsetz[13] cite the examples of loading and of fishing. My own favorite example is riverboat pulling in China before the communist regime, when a large group of workers marched along the shore towing a good-sized wooden boat. The unique interest of this example is that the collaborators actually agreed to the hiring of a monitor to whip them. The point here is that even if every puller were perfectly "honest," it would still be too costly to measure the effort each has contributed to the movement of the boat, but to choose a different measurement agreeable to all would be so difficult that the arbitration of an agent is essential.

With regard to this example, Alchian and Demsetz argue that the moni-

[11] Coase, *supra* note 1, at 391–92. Coase argues that the diverse activities described in this quotation lead to the use of a long-term contract—which I do not consider important to his thesis. The training of workers or the remodeling of a rented building, like improvements made on agricultural land by a tenant, appears to be a more incisive reason for the choice of a long-term contract because of the difficulty of recapturing the investments committed. See Cheung, *supra* note 5; and Klein, Crawford, & Alchian, *supra* note 8. The latter paper employs the concept of quasi rent to interpret the choice of a long-term contract.

[12] In a 1968 discussion, Coase expressed doubt that the power of eminent domain would be necessary to reduce transaction costs. Yet in my later study of the problem of evicting tenants for housing reconstruction under rent control the evidence is incontrovertible that in certain circumstances the power of eminent domain reduces transaction costs. See Steven N. S. Cheung, Rent Control and Housing Reconstruction: The Postwar Experience of Prewar Premises in Hong Kong, 22 J. Law & Econ. 27 (1979), particularly at 50–52.

[13] Alchian & Demsetz, *supra* note 4.

tor (that is, the "firm") enters the picture because of the shirking.[14] My argument goes further back, asserting that the behavior has itself originated because a proxy instead of effort was chosen for measurement; the concept of shirking is, therefore, an indirect way of expressing that there is a cost in discovering prices for relative contributions. Within what Coase calls a firm, then, shirking takes place in a way that differs from dishonesty in the market.

Let me summarize the arguments of this lengthy section. In principle, all contributions of input owners as well as the services of the coordinator can be separately priced and sold to customers by measuring directly various attributes related to each contribution.[15] In this case product and factor markets coincide. But the determination of prices is costly because of the number of transactions, because consumers lack detailed information on the use of each component or contribution to a commodity, because of the difficulty of measuring varied and changing activities, and because of the need to separate contributions.

One effective way to reduce the costs of discovering prices is to substitute some device other than the direct and separate pricing of activities. This substitution may be as simple as the use of the piece-rate contract or as complex as the establishment of a communist regime.[16] Coase's main concern is substitution that entails the delegation of use rights from *private* input owners, to the extent that the contribution or activity of each is not directly priced. This implies measurement by a proxy. An "entrepreneur" therefore emerges.

Specialization and coordination are relevant here to the extent that they typically involve multiple input owners and tend to compound the costs of discovering prices. Had there been no cost of measuring and pricing performance there would be no firm, and the value of the social outputs would be maximal. But these costs do exist, and product market transactions decline as they are being partly superseded by factor-market contracts. The agency costs of a monitor, a director, or a manager—which

[14] Following Knight, Alchian & Demsetz (*supra* note 4) also define a firm to have a residual claimant. Coase notes, however, that under share contracts no residual claimant exists (*supra* note 1, at 392).

[15] While in the total absence of transaction costs the services of a monitor or an agent would not be necessary, those of a coordinator might still be wanted. That is, the services of a conductor in a symphony orchestra or of a coach in a team sport have little to do with transaction costs. In practice, of course, one who coordinates may also perform some functions of an agent.

[16] The communist regime is a "superfirm" within which citizens lack the option of not joining. I have argued elsewhere that this lack necessarily implies higher transaction costs in the operation of the economy. See Steven N. S. Cheung, Will China Go "Capitalist"? (1982).

are also transaction costs—rise as the costs of discovering prices fall.[17] And the supersession will go no further when at the margin the saving in one type of cost equals the rise in the other.

Textbooks typically depict the product market as consisting of outright sales of goods and services at prices determined by supply and demand. No reference is made to any contractual arrangement under which the producer (supplier) may be paid by measuring a proxy rather than the product. The factor market then crops up in another chapter, typically under the theory of marginal productivity. Each input owner receives remuneration equaling the value of his marginal product, and no mention is made of any delegation of rights to an agent. Indeed, how an input owner gets paid and by whom seem irrelevant.

It is *not* quite correct to say that a "firm" supersedes "the market." Rather, one type of contract supersedes another type. Coase's main concern is a type of contract under which an input owner surrenders a delimited set of rights to use his input in exchange for income. He is therefore directed by a visible hand, not by the invisible hand of a price mechanism. It takes remarkable insight to see that as this type of contract increases there will be fewer product markets.

II. The Piece-Rate Contract as an Illustration

One appealing aspect of the piece-rate arrangement is that no other labor contract so clearly reveals productivity differentials among workers. This provides useful observations for testing a number of implications of the marginal productivity theory. Our present interest is in another important aspect: piece-rate workers in a factory (a "firm") are paid on the basis of direct count of their production contributions (a "market"). Hence, to identify a "firm," which we shall later argue is a futile task, one would have difficulty coping with the piece-rate arrangement. But precisely because this arrangement does have a dual nature, tracing the transition from the simplest form of a piece-rate contract to a wage contract will usefully illustrate our earlier interpretation of the organization of production. What is or is not a firm is immaterial: what counts are the various ways of organizing economic activities under different transaction costs.

Since my investigation on the piece-rate contract was conducted in Hong Kong, all factual examples provided here will be taken from that

[17] The term "agency costs" is taken from Michael C. Jensen & William H. Meckling, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, 3 J. Financial Econ. 305 (1976). In Coase's early work, an agent is regarded as the same as an entrepreneur.

city. This choice was made to avoid complications attendant on labor unions and minimum wage laws.[18]

Let us begin with the case of a middleman in Hong Kong who buys shirts from manufacturers by piece count and sells them to an importer in the United States. He shops around, gathers samples, makes offers, and quotes prices to his clients. Seldom will he tell a manufacturer what to produce, and never will he tell the factory workers what to do. He makes his living by specialization: he makes contacts, learns the preferences of a particular market, and has a good knowledge of prices. In short, he does not direct resource use; he merely transmits price signals. Few would claim that the middleman uses a piece-rate contract even though, in fact, he pays for the shirts by the piece.

The transmission of price signals through specialization in contacts and price information may involve several steps. Consider the laying of hardwood floors. A landlord who wants to build a high-rise finds a building contractor. This contractor subcontracts with a hardwood floor contractor on an agreed price per square foot—a piece count. The subcontractor, who imports the wood materials and adds finishing work to the wood on a piece-rate basis, in turn finds a sub-subcontractor, provides him wood, and offers him a price per square foot laid. Finally, the sub-subcontractor hires workers and again pays them per square foot laid.

At each stage of contracting a price signal is transmitted, at a cost, and each successive contractor holds different contacts and information. Of course, a middleman or a contractor may conceal price information to his own advantage, but a price signal is nonetheless transmitted because his decision to buy or to subcontract is based on his own price information. Furthermore, competition in contracting will curb the success of his effort to conceal price.

Other than the number of steps involved, it is difficult to see any difference between the hardwood floor example and the earlier example of the shirt exporter. But while few would say the middleman buys shirts on a piece-rate contract, everyone somehow agrees that the hardwood floor is produced under piece rate. The key distinction seems to be that the middleman does not make payments directly to the workers who make shirts,

[18] For example, whereas under piece rate an unusually slow worker may receive an income below minimum wage, the establishment of a minimum piece rate would be nearly impossible because of the great variety of such work. It is said that when the United Auto Workers was formed some forty-five years ago, the abolition of the piece-rate arrangement was regarded as one of the union's major accomplishments. One popular argument against the piece rate is that the rate for a hard worker will be revised downward by the management. This is flatly contradicted by my Hong Kong findings which, as will be elaborated elsewhere, show that a hard worker often receives bonuses when some specified level of output is attained.

whereas the sub-subcontractor pays his workers directly. Yet a resident homeowner who wants his floor replaced may also pay per square foot to a worker directly, as some do, although most choose to order through a contractor. It is not all that easy to separate a factor market from the product market.

The hardwood floor example clearly illustrates the advantage of the piece-rate contract. Square footage is a straightforward measure which leaves little room for argument. In fact, the hardwood floor in Hong Kong is typically set in squares of one square foot each, and both the workers and the contractors are so expert at estimating partial squares that they often agree on the dimensions without using a tape. The quality of a finished floor is also easily judged by the specialist subcontractor in a few minutes of inspection.

Another piecework advantage in this case is that, because of Hong Kong's humid climate, hardwood floor of a particular type is routinely used in all the highrises. The market price of the flooring per square foot is well known, and the price between contractor and worker is easily negotiated because the workmen know their alternatives. Since the worker is seldom given job direction other than time and place, he is delegating little of his use rights. A price is paid for his service per square foot laid.

For this industry, payment by the hour—a wage contract—would not be practical to negotiate and enforce. Different workers have such varied speeds that a uniform wage will guarantee quarrels. On the other hand, the cost of assessing different abilities is sufficient to assure that varying wage rates would have to be determined somewhat arbitrarily and would require considerably more background information about the workers. Moreover, only a long-term relationship between contractor and workman can assure, under a wage contract, that the work will be done diligently. When many among the labor force work on call from job to job such a contract invites lethargy. To monitor the work is costly; yet a general slowdown can add up to a great deal of unproductive time.

The stringing of beads for belts and headbands is another task routinely done on a piece-rate basis. Little equipment is required, and the workers do the task at home. Each piece is usually a complete commodity, and payments are based on piece count and quality inspection. Outstanding pieces are awarded small premiums; they are used as samples or are placed on top of others in packaging. Typically, about the only direction given to the worker over what would be provided by the consumer in the market is a color pattern for the beads. In transmitting the price, the contractor also transmits an estimated consumer preference as to color and pattern. The worker is told what to do about color because the piece rate he receives contains one less signal than would the price if paid directly by the consumer.

It is always the case that the less explicitly the transmitted price indicates consumer preferences or product specifications, the more direction will the worker receive from a "visible hand." Directions may be given by an agent or by a consumer; in either case, the giving of directions means that the input owner has to some extent relinquished his use rights. Direction under the piece-rate contract tends to be far less than under the wage contract because measurement by the piece is a more direct measure of the actual contribution than is measurement by hours of work.

In the garment and toy industries, as well as in a variety of industries involving light metallic and plastic products, piece-rate contracts are popular. In the vast majority of cases the piece-rate worker specializes at any one time in producing only one component of a product. Examples include the painting of a doll's eyes, the assembling of a toy, the polishing of one part of a flashlight, or the cutting of materials for a coat. Indeed, an entire shirt may be the product of a whole combination of piece-rate works. Among the requisite conditions are that each piece can be easily counted and that a standardized set of measures governs inspection (by random sampling, for low-rate works).

The pricing and measurements adopted for piece components differ from those of an assembled product. Nonetheless, under piece rate it is the contribution of the worker which is directly measured and priced. Pieceworkers may work either in a factory where space and tools are provided or at home using their own equipment—understandably in this case at a higher rate to compensate for the rental value of the equipment and space. In either event, the component worker is given specifications because (as the component is seldom sold separately in a final-product market) the piece rate offers little information about consumer preference. Those working at home are also assigned a deadline for delivery to integrate with the schedule for other components. Coordination of activities is improved when all workers perform under one roof, supplementing with work sent out to home craftsmen only at times of peak loads.

Pieceworkers in factories are often assigned specific machines to avoid careless use under random rotation, and fast workers are often assigned the better equipment. We may put this in the context of Coase's thesis: the home worker is paid a piece rate for the contribution of his labor plus his equipment, whereas the factory worker is paid only for his labor. The contribution of the factory equipment is priced in terms of implicit rental values. Thus the assignment of machines to specific workers and the occasional monitoring of the workers themselves may be viewed as "directions" given to the machines in lieu of a direct pricing of their contributions. Predictably, piece-rate works which require the use of heavy machinery tend to be produced in a factory. Such machinery is awkward in a home, dangerous to children, and requires prompt repair; it may also be

so costly that few home workers would willingly invest without assurance of long-term use.

To equalize the marginal productivities of similar machines requires that their rental values be equal. Yet given the same piece rate the rental yield to a machine will be higher with a faster worker. Thus in addition to assigning a better machine to a faster worker, when costly machines are used a bonus over a standard piece rate is paid for quick work (that is, a lump-sum bonus if a worker exceeds a certain number of pieces a day). Thus one with double the output of another often receives more than double total pay. The home worker receives no bonus even though he is using a costly machine, because his piece rate reflects the total contribution of all inputs except the materials provided him.

The determination of the piece rate—a price—illustrates the cost of "discovering" prices. For a part or a component of a market product, there is often no market price to serve as a frame of reference. For many components, however, there are standard piece rates, long used in the industry, established by reference to the daily wage paid or to piece rates offered by competing factories for similar work. Minor modifications in the piecework may promptly lead, through rough estimation, to minor changes in the piece rate. But a greater problem concerns major changes, particularly in the plastic and light metallic industries, where models often change significantly between orders and where each factory tends to originate unique models. The management must decide, in such cases, whether the new component promises sufficient volume to justify the costs of negotiating a new piece rate.

Test runs are an unsuitable basis for determining such a rate, because workers beginning an unfamiliar task will take several days to pick up speed, and they tend to overstate the difficulty of the new job. The standard procedure therefore is to employ a motion specialist, sometimes recruited from outside, to negotiate with a representative of the workers. The basic reference is that an average worker moves his hands or feet approximately 8,500 times on a normal working day, the number varying with the extent of the required motion and with its duration. The number and types of motion required to complete the new component are first assessed, then a daily wage, or the average daily earning of a piece-rate worker in a comparable trade, is used as a frame of reference in setting the new rate of pay.

This negotiation process may be cumbersome because the worker representative must gain consent from the workers. Moreover, a rate so negotiated is seldom binding, and either side has the right to demand renegotiation after a few days of work. These negotiation costs—that is, the cost of discovering price—will of course be lower if the component to

be produced, like a string of beads or a square foot of hardwood floor, commands a well-established price in the final product market which may be used as a reference. But, as noted in the preceding section, few customers would even recognize the function of some components.

That is why piece rates are not applicable when components or contributions either change often or entail a variety of activities by the same worker. The point is illustrated by the polishing of flashlights or gas lamps. Workers doing routine finishing with a motor and wax are paid a piece rate; but when defects are found, such as some minor scratches or burns from electroplating, workers correcting the flaws are paid a wage. Similarly, other employees receive wages: the janitor, the office clerk, the accountant, the model designer, and the chemist in the electroplating shop.

When components are assembled in a way that makes separation of workers' contributions costly, the piece-rate contract is also difficult to establish. In one sort of collaboration, a group of workers polishes flashlights and places them in separate baskets which, after counting and inspection, are moved to another group which inserts switches. Both activities are paid by the piece. But in another type of collaboration, when the flashlights are being electroplated or anodized, one worker monitors the tank of chemicals, another rotates the articles in the solution, and a third rinses them as they are handed to him on a hanger. The relative contribution of each worker is thus difficult to separate. Similar difficulty attends the planning of a new product: one member searches the market for ideas, a committee evaluates alternatives, and the export manager solicits feedback from potential buyers. Where each of these workers has incentive to claim more credit because his share is difficult to delineate, it becomes less costly to price by measuring a proxy under the arbitration of an agent.

A proxy such as an hour or a day is easily measured; it does not change as components or contributions do; it can be separately measured for each worker; and it can embrace a host of activities to be performed by each worker, some of them too trivial to be subject to contracting. These advantages result in lower costs of determining wage rates than would be possible if contributions were measured in the complicated situations just described. But a unit of time in itself contributes no productive value; it merely represents what a worker presumably can perform. More stringent monitoring and direction are thus required under a wage contract because the activity itself is not directly priced. From this we infer that in some cases the piece rate may be chosen even if the costs of determining that rate are higher than those for a wage rate.

A piece rate transmits a price signal, although we have shown that the information it provides may vary greatly and, as a result, specifications

and directions vary in degree. While the inspection of output is more demanding for piece-rate work because the worker has an incentive to rush, the total costs of monitoring are necessarily lower than under a wage contract. This is because the piece-rate worker's pay is proportional to his contribution, which reduces the need to contract for the use of his input. On the other hand, pricing by measurement of a proxy necessitates greater control of the use rights before profitable contributions can be expected. This delegation of rights by the input owners is a delimited package within which the entrepreneur can pick and direct production activities without pricing contributions. Management is a cost; decision error is also a cost. These costs are higher in pricing a proxy; the benefit is a reduction in the costs of discovering prices.

In this section I have traced the transitions from the simple service of a middleman, through various piece-rate arrangements, to wage contracts. The measurements adopted range respectively from a whole product to a component to a proxy measure. The information of the price signal moves from a full valuation of the contribution to no direct signal at all.[19] Direction and monitoring by the agent correspondingly rise in complexity. Delegation of the right to use the input also increases until full control is granted in terms of some contractual limits. The supersession of market transactions ranges from a possible doubling of transactions under the middleman arrangement to the making of piece-rate payments on a collective scale to the clear emergence of factor-market transactions superseding product-market transactions.

Throughout these transitions I have, following Coase, emphasized the costs of discovering prices, including the costs of information, of measurement, and of negotiation. When these costs change, different contracts appear. I do *not* claim that price determination is the only transaction cost which matters in all choices of the forms of contracts or organizations. But with regard to the central point made by Coase, my own investigation supports his view: under private property rights, any movement toward the contractual delegation of use rights results mainly from the constraints of pricing costs.

III. The Ambiguity of Firm Size

Some ten years ago I posed to Coase the following question: If an apple orchard owner contracts with a beekeeper to pollinate his fruits, is the

[19] Strictly speaking, not all valuable characteristics in a product are directly measured and priced. Properties such as vitamin C in an orange or the use of salt in a restaurant are not measured in transaction, although their presence is known and implicit in the product prices. This and related implications should not detain us here.

result one firm or two firms? This question has no clear answer. The contract involved may be a hive-rental contract, a wage contract, a contract sharing the apple yield, or, in principle, some combination of these and still other arrangements. In each case the beekeeper receives a remuneration for his service, and the orders he expects from the orchard owner vary with the form of contract.

The taxing agent of a government may treat the beekeeper and the orchard owner as one or two firms, depending on tax laws and business registrations adopted in the relevant area. But what is the viewpoint of economics? Most economists would probably opt for only one firm if the beekeeper is hired on a wage contract but for two if the hives are rented. Does it make sense to say that the number of firms, hence firm size, depends on the chosen form of contract? If so, in what sense? Is it not true that different directions will be given by the orchard owner under different contracts? What about the case of the sharing arrangement? And what if, instead of renting beehives, the orchard operator rents out the orchard itself? The basic problem here is that examples can easily be found to counter almost any definition of a firm that one can offer.

Consider our earlier example of the hardwood floor. There are three stages of subcontracting between the building contractor and the workers. How many firms are there? One? Or four? Hong Kong laws hold that all parties are separately responsible for tax, and all but the worker hold business licenses. Yet economists may well argue that because they are all vertically integrated by contracts, with transfer pricing, only one firm exists.

Consider, further, the case of a big department store bearing only one name but consisting in fact of separate sellers, each leasing a space under one roof, paying a rent to one central agent, and governed by a set of rules on the line of products each can sell and his hours of operation. That would seem to represent a single firm. However, *exactly* the same arrangements are found in most shopping centers, except that the shops bear various names. Why should that matter in determining firm size? Of course, some names are worth more than others, as is evident in the case of franchises, but it is also true that a single corporation may establish a number of subsidiaries, each bearing a different name for a different business.

The truth is that according to one's view a "firm" may be as small as a contractual relationship between two input owners or, if the chain of contracts is allowed to spread, as big as the whole economy. We have noted that the delegation of use rights and the transmission of price information are matters of degree. Given, then, that every individual is an input owner, it is highly likely that each has some contractual relationship

with others in addition to the goods he sells or buys in the product market. Seldom in a modern society would an individual operate single-handedly in production and in exchange of his output in the marketplace.

Thus it is futile to press the issue of what is or is not a firm. If each individual is a private input owner—of his own labor, if nothing else—then almost all individuals in society are bound by contracts when they compete and interact.[20] The important questions are why contracts take the forms observed and what are the economic implications of different contractual and pricing arrangements.

In the domain of price theory, the traditional concerns of economists have been income distribution and resource allocation. Contractual arrangements, and their effects on income distribution and resource allocation, have been slighted. This oversight would be irrelevant if all transactions were costless: in that case, various contractual arrangements would never have been devised. But transaction costs *are* positive and significant. Hence, real-world observations are often not satisfactorily explained by the traditional approach. In an indirect way, Coase's early paper on the firm made this same complaint.

In spite of his attempt to seek a definition of the firm, Coase was well aware of the difficulty of drawing a boundary. He stated in a footnote that "it is not possible to draw a hard fast line which determines whether there is a firm or not."[21] His concern was not so much with the firm as with production and exchange activities observed to have been organized in different ways. In particular, he concentrated on the observations of direction by price signal versus direction by an entrepreneur. We need not reiterate here the different types of contracts and their implications.

We have noted that equilibrium is reached when the saving of transaction (pricing) costs incurred in the product market is equaled at the margin by the rising agency costs (also transaction costs) within an emerging "factor market." This is identical to Coase's condition that "a firm will tend to expand until the costs of organising 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."[22] But the condition does not really determine firm size. Rather, it determines the extent of contract substitution.

We have stated that the costs of discovering and agreeing on prices in

[20] In the Chinese communist regime, not even labor is privately owned: workers have no rights either to choose their profession or to contract and negotiate their wages. See Cheung, *supra* note 16, at 33.

[21] Coase, *supra* note 1, at 392.

[22] Coase, *supra* note 1, at 395.

the product markets include those of information, measurement, and negotiation. In different dimensions and magnitudes these costs also exist in the factor markets. For factor markets Coase cites various agency costs, including those of organizing activities and of making errors. He also adds changes in the supply price of factors, the relevance of which I do not see. Coase equates "rising supply price" with "diminishing returns to management,"[23] which appear to have been popular subjects among British economists of that period. But this point seems redundant once we recognize that (1) pricing by measurement of a proxy incurs extra costs of management and increased chance of decision errors and (2) these costs rise as the price signals transmitted contain less information, thus requiring greater control of use rights delegated by input owners.

Perhaps the most interesting confirmation of the fruitfulness of the approach presented here is that we cannot say much about firm size because we do not know precisely what a firm is. Stated differently, under the Coase approach we *can* identify clearly what a firm is and hence its size if the following conditions are met. First, producers or agents sell directly to consumers in outright transactions. Second, agents or entrepreneurs hold only wage or rental contracts with input owners. And third, there is no contractual relationship between agents. Ignoring the producer who owns all his input resources outright, these conditions yield the polar cases of (1) consumers buying commodities from agents (the product market) and (2) each separate agent holding wage and rental contracts with input owners (the firm). With any contractual relationship between agents assumed away, firms become separately identifiable and are literally "islands of conscious power."[24] To say that one type of contract supersedes another will then be the same as saying that the firm supersedes the product market. Firm size becomes determinate at the moment contractual supersession is determinate.

But the world is more complex. The polar cases are complicated by middlemen and subcontractors; agents contract among themselves; and any type of input may support a variety of contractual arrangements. We surmise that these very complications, which render the "firm" ambiguous, have arisen from attempts to save transaction costs that were not avoidable in the polar cases.[25] As one example, we have shown that the

[23] Coase, *supra* note 1, at 396–97.

[24] Coase cites D. H. Robertson at 333: "[Firms are like] islands of conscious power in [the] ocean of unconscious co-operation like lumps of butter coagulating in a pail of buttermilk."

[25] Governmental regulations may complicate the choice of contracts, but this we ignore here.

piece-rate contract lies somewhere in between, with a lower pricing cost than if the components were sold directly in the market and a lower agency cost than wage or rental contracts. The total of transaction costs is reduced by the piece-rate contract. The Coase thesis is therefore strengthened at the trivial sacrifice of acknowledging that we now know less about the definition and size of the firm.

IV. Concluding Remarks

In Hong Kong in 1969, I sat on an empty wooden box by the roadside and let a boy shine my shoes at an agreed price of twenty cents. As soon as his work began another boy approached and, without saying a word, started to polish my other shoe. "How do I pay?" I asked. "Ten cents for each of us," one boy replied. Not knowing the pricing practice, I was surprised to find out that the two boys did not know each other at all. But it dawned on me that this must illustrate what Coase meant by the market: the splitting of one transaction into two could not have gone so smoothly had I worn a shoe on one foot and an uncommon boot on the other. This little episode led me to reread "The Nature of the Firm" with a different orientation, and the subsequent research on contracts, which forms the basis of my present interpretation, followed to satisfy my own curiosity.

Few would deny that Coase's work on social cost is his most important and his most famous.[26] But I hold the "firm" paper in special regard because its insights are the spring at the headwaters of his most important subsequent writings. The surging interest in the "firm" paper during the past decade is one measure of the change in view among economists about the importance of the real-world applicability of theory which Coase and others have helped to promote.

It is not long since external economies or diseconomies constituted a prime subject in the field of economic development, presented along with such dubious terms as "bottleneck," "take off," "vicious circle," and "balanced versus unbalanced growth." These seem archaic concepts now. Only twenty years ago the word "contract" was confined to the "contract curve" in the Edgeworth-Bowley box; now it has become a household word in the discipline. Property rights were an untouchable area for doctoral theses; now centers have emerged and the subject is taught even at the undergraduate level. The transaction-cost paradigm is here to stay.

[26] Coase, The Problem of Social Cost, 3 J. Law & Econ. 1 (1960). This may well be the most cited economic paper of our time: 1966–70, 88 citations; 1971–75, 286 citations; and 1976–80, 331 citations.

Coase was not alone in promoting this paradigm. Knight and Hayek, among others, shared an early interest in Coase's subject matter. But came the Keynesian revolution, and whatever observations economists did not understand found refuge in various "imperfections." It was a different era when, in 1960, Coase published his paper on social cost. In the same period Stigler wrote on information and Arrow on the appropriability of returns.[27] Students once frustrated by the inept explanatory power of price theory were now told of something other than imperfections. At roughly the same time, Aaron Director's interest in tie-in sales was spreading, and Armen Alchian was persistently interpreting price and competition in terms of property rights. Students at UCLA were buying duplicate copies of Friedman's price theory lectures in the gray market, sweating over the questions provided at the end in fear that some of them might be used in an upcoming preliminary examination. That Friedman's lectures had little to do with his questions on mergers and various forms of pricing behavior foretold what was to come.

Coase was fortunate in that his important work on social cost (and transaction costs) took shape just as a tide was turning. The call for policy recommendation, which had dominated economic thinking, was subsiding as a desire for economic explanation began to swell. The turning tide surged far more strongly because of what Coase had to say. My view is that transaction costs and contracting will someday be regarded as a basis for analysis rivaling marginalism in neoclassical economics.

[27] George J. Stigler, The Economics of Information, 69 J. Pol. Econ. 213 (June 1961); and Kenneth J. Arrow, Economic Welfare and the Allocation of Resources for Invention, in The Rate and Direction of Inventive Activity 609–25 (1962).