RESOURCE COORDINATION AND TRANSACTION COSTS
A Framework for Analyzing the Firm/Market Boundary

Lee J. ALSTON*
University of Illinois at Urbana-Champaign, Urbana, IL 61820, USA

William GILLESPIE*
Yale University, New Haven, CT 06520, USA

Received April 1987, final version received June 1988

We develop a framework to classify transaction costs. Our framework is based on the recognition that the types of transaction costs encountered in organizing production vary across the factors of production and over the stages of the production. Our framework allows us to distinguish between the transaction costs of using the market and those of using the firm. Most importantly, it enables us to see that transaction costs are a function of what factors an individual supplies to the production process. This insight, by illuminating the interdependence of transaction costs, contributes towards understanding the firm/market boundary.

1. Introduction

With the introduction of R.H. Coase’s (1937) paper The Nature of the Firm, the study of transaction cost economics began. Coase articulated a powerful challenge to conventional price theory with a simple question: If the price mechanism can coordinate production so well, why does the firm exist [Coase (1937, p. 333)]? Coase examined several of the received explanations of the firm, including the division of labor and risk based explanations, and rejected them as inadequate. Instead, he proposed that ‘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’ [Coase (1937, p. 336)].

Coase’s insight that transaction costs are the economic basis of organizations like the firm has been developed in a large and expanding body of literature. Despite the intellectual effort devoted to the theory of transaction costs, a definitive position on the nature of the firm has yet to emerge. One problem with the transaction cost literature is that specific costs are often

*For comments we thank Ralph Bradburd, David Fairris, Peter Gorringe, Robert Higgs, Timur Kuran, Alvin Klevorick, Gary Libecap, Nancy Lutz, Douglass North, Louis Putterman, David Ross, Padmanabhan Srinagesh, Gordon Winston and participants at seminars at the Australian National University, LaTrobe University and Washington University. Lee Alston is grateful for support as a visiting fellow at the Australian National University during 1986–1987.

invoked in an ad hoc fashion to explain economic organizations. We provide
a framework for examining transaction costs, so that future empirical testing
can identify which costs are critical to different organizational forms.

2. A structure of transaction costs

To an extent, previous explanations for the existence of firms are
complementary rather than competing approaches. Each theory offers a
valuable insight into the reasons why firms exist, and each makes a
contribution by identifying specific types of transaction costs. Nevertheless,
the common problem of these theories is that they fail to take each other
into consideration. The metering problem identified with Alchian and
Demsetz (1972), for example, is certainly an element in the larger explanation
of the theory of the firm, but it is not the only reason why firms exist.
Similarly, though Williamson’s (1979, 1981a, 1981b and 1985) concept of
asset specificity can explain some aspects of vertical integration, other costs
identified by Barzel (1982), Cheung (1983) and others also play a role.

In this section we outline a framework which organizes the diverse group
of costs relevant to the firm/market decision which appear under the banner
of transaction cost economics. In addition to suggesting the complementary
nature of many of these costs, this framework provides a structure which we
will use to derive further implications for the theory of the firm. Nevertheless,
we recognize that our framework may be incomplete. Other transaction costs
that have not been identified may also be significant to the theory of the
firm. We hope that our framework will be able to encompass any additions
that subsequent writers wish to make. Such additions, if they can be made
without significantly changing the basic structure of the framework, are really
the ultimate ‘test’ of this type of conceptual apparatus.

Our framework for categorizing transaction costs rests on Coase’s obser-
vation that both the firm and the market are methods of coordinating
production [Coase (1937, p. 334)]. In both the firm and the market, the
factors of production are combined to create goods and services. This
combination does not occur instantaneously as the neoclassical production
function might suggest, but instead, production takes place through a process
of explicit and implicit contractual arrangements between individuals. The
key to understanding the structure of transaction costs is recognizing that the
types of transaction costs encountered in organizing production vary across the
factors of production and over the stages of the production process.

To categorize transaction costs, several analytical distinctions need to be
made. First, in our framework, the factors of production are divided into
three categories: physical and financial capital, human capital, and work
intensity. Physical and financial capital consists of the machines and tools of
the production process plus all ‘liquid’ capital. Human capital consists of the
knowledge, skills and labor used in the production process which are possessed by each worker. Knowledge and skill may be acquired through learning by doing or through more formal schooling. In addition, human capital includes skills like the entrepreneur's talent for recognizing productive activities that cannot be 'learned' in any conventional sense. Work intensity refers to the withholding or application of work effort. The incentive to withhold effort decreases the more closely effort is linked to reward and the less costly it is to monitor effort. For example, workers have less incentive to withhold effort under piece-rate contracts than under wage contracts. Furthermore, when it is not costly to monitor work effort, workers are not likely to withhold effort because they fear that if detected they will lose their job.

The second distinction we make for analytical purposes is to divide the production process into three periods of time: pre-production period, negotiations between the buyer and the seller take place. After the contract is agreed upon, production takes place. The post-production period is marked by the delivery of the good or service agreed to earlier. Although the actual boundaries between these periods are not always as sharp in reality, production does proceed through this sequence of events.¹

Our framework concentrates on the transaction costs of contracting within and between firms, yet transaction costs alone do not determine the profit-maximizing governance structure but rather the sum of the production and transaction costs [Wallis and North (1988)].² By production costs we mean the costs in an engineering or technological sense of combining inputs to produce output (i.e., the usual neoclassical production function). Transaction costs are the negotiation, monitoring and enforcement costs necessary to assure that contracted goods and services between and within firms are forthcoming. Neoclassical economists have tended to focus on production costs such as scale economies to the neglect of internal and external transaction costs. Just as transportation costs must be weighed with production costs so too do transaction costs. For example, in the days before enforcement of contracts by government, the high transaction costs of knowing the reliability of one's trading partner limited the ability to reap fully the advantages of scale economies. Our goal is not to determine efficiency in any global sense but rather to posit how transaction costs affect the likelihood of contracting within or between firms.³

¹Williamson (1985) continually stresses that contracts must be viewed in their entirety. This is what we mean by stages of the production process.
²This abstracts from the potent criticisms raised by Putterman (1984) and Dow (1987) about why labor managed firms may be more efficient than capitalist firms. Putterman and Dow argue that appropriability problems may prevent labor managed firms from emerging.
³Framing the problem in this fashion enables us to examine the comparative static properties of transaction costs without worrying about the agency problems of who monitors the monitors and whether an alternate nonexistent governance structure is Pareto optimal.
By crossing the three factors of production with the three stages of the production process, we obtain a framework which we use to classify transaction costs (see table 1). This framework is useful for several reasons:

1. It enables us to identify those transaction costs which have some bearing on the firm/market boundary;
2. It allows us to distinguish between the transaction costs of using the market and transaction costs of the firm;
3. Most importantly, it enables us to see that transaction costs are a function of what factors an individual supplies to the production process.

The resulting interdependence among the transaction costs influences the decision 'to make or buy'.

Other authors have identified the various transaction costs listed in table 1 but no one has put them into a coherent framework. The items in table 1 referred to as 'pre-production' and 'post-production' are factors that encourage production within firms. Alternatively, their absence encourages market transactions. The items under 'production' are costs of using the firm. Market transactions do not entail the costs under column 2 because the party to receive the good has no incentive per se to oversee the use of the resources in the production process. All the receiving party cares about is the availability and quality of the goods received.

2.1. Pre-production costs of using the market\(^4\)

The role played by asset specificity in promoting integration was first elaborated by Klein, Crawford and Alchian (1978) and Williamson (1979, 1981a, 1981b and 1985). Asset specificity refers to the problem that is created when one party in a transaction makes an investment in human or physical

\(^4\)In this and the following two sub-sections we will elaborate on the transaction costs in table 1.
capital which cannot be fully recovered if the transaction is terminated [Williamson (1981a, p. 555)]. Transaction specific investments create a problem for market procurement. An investment in assets specific to a transaction locks both the buyer and the seller into a bilateral monopoly because the cost of the asset cannot be fully recovered if the transaction is terminated [Williamson (1979, p. 240)]. In the terminology of Klein, Crawford and Alchian (1978, p. 298), a stream of quasi-rents is created which can be appropriated through opportunistic behavior by either party. In order to make such investments without incurring the hazards of post-contractual opportunistic behavior, firms will organize such transactions internally through vertical integration.

Another set of transaction costs may be associated with constraints on the availability of labor. An inelastic labor supply creates the potential for workers to 'hold up' their employer at critical times in the production process if there are only a limited number of workers available. One example is migrant farm workers who may be able to extort almost the full wholesale value of the crop because they are the only labor force available to the farmer at harvest time. Klein, Crawford and Alchian (1978, p. 313) use a similar example in which a farm workers union is able to extort quasi-rents because it can exclude other workers. They conclude that the union was able to extract rents because the crop was 'specialized' to the union. While we agree, we believe that in the short-run, an exogenously determined labor market may also be an important determinant of contract form [Alston (1981, p. 217)].

The other type of pre-production transaction costs noted in table 1 are the information constraints associated with the acquisition and use of human capital. These costs are classified as pre-production because information constraints influence the way in which contracts are negotiated between and within firms. For analytical purposes, we separate these costs into two broad categories: knowledge paradox and information transmission.

Knowledge paradox refers to the potential for the appropriation of rents created by the inherent asymmetries involved in selling information. In order to transfer information, the seller must reveal the content of the information in the pre-production stage, but if the seller reveals too much, the buyer no longer needs the seller. If the buyer cannot be fully informed, however, he takes the risk that the information will not be worth what he paid [Teece (1980, p. 228), Winter (1964)].

Information transmission is an information constraint which affects both firms and markets. When information is difficult to communicate across markets, internal organization may be the best way to realize the productive potential of the knowledge. Entrepreneurs who realize that there is an opportunity for profit from a new product must first gain control of the resources needed for production. The transaction costs of transmitting
information in the market about specific services that they need may be more costly than forming firms and directing production themselves. By directing production, entrepreneurs reduce transmission costs because information about the steps of a process may be easier to convey than information about the entire process [Silver (1984, pp. 13–15)]. Furthermore, the other participants in the production process need not bear any risk – apart from employment risk – when entrepreneurs own the physical capital.

This argument, however, assumes that the entrepreneur either owns or has access to sufficient capital to integrate the entire production process. With outside investors, the entrepreneur must assume the transaction cost of convincing lenders that he has a good idea. To the extent that the entrepreneur can convince the lenders of capital, firms provide a vehicle for entrepreneurs to use other people's capital to realize their ideas. Alternatively, ventures that would be profitable if only production costs are considered may never materialize because the information transmission costs required to convince the lenders of capital to commit their resources are too high.

Information transmission costs are also generated by the complex nature of many types of information. Teece (1980, pp. 226–228) argues that team 'know-how' may be difficult or impossible to communicate across markets because it cannot be expressed simply or concisely in a blueprint or an instruction manual. Instead, team know-how involves a large learning by doing component and requires not one but many people working together as a team to realize its productive potential. Since this type of team know-how may be prohibitively costly to transmit, firms may be the least costly method of transforming team know-how [Teece (1980, pp. 230–232)]. In other instances, possession of information or 'know-how' about the production of a single product may be applied to the production of other products, thereby enabling firms to capture economies of scope [Teece (1980, pp. 228–229)].

Firms, however, do not always have advantages over markets in transmitting information. Dispersed information, like the specific knowledge of time or place possessed by each individual, may be difficult to internalize into firms. Individual discretion based on specific knowledge and motivated by self-interest is necessary for timely action [Hayek (1945, pp. 521–522, 526)].

2.2 Post-production costs of the market

A necessary condition for a market transaction is that participants be able to measure the quality and quantity of output. If it is difficult to judge the quality of goods directly there is an incentive for integration because
otherwise both the producer and the purchaser need to expend resources in measuring the quality of goods by monitoring the inputs. To avoid duplication in measuring all steps of production, integration arises over the steps where direct measurement is difficult [Barzel (1982)]. Barzel argues that brand names and guarantees arose to reduce the otherwise high cost of measurement when products are transferred between owners in a market. In addition, when output is produced by a team of workers the problems of measuring relative contributions provide another incentive to internalize production [Alchian and Demsetz (1972)]. More market transactions relative to internal production will transpire if individuals can be paid directly for their contribution to output. When measurement costs preclude paying directly for output, firms internalize production and pay workers based on a proxy for output like hours worked [Cheung (1983, pp. 7-9)].

Low measurement costs alone do not determine whether market exchange will prevail over internal organization. Contractors must be able to enforce contract stipulations [North (1984, p. 7)]. Contract enforcement includes those costs incurred to prevent or rectify a breach of contractual stipulations. Contract enforcement is categorized in table 1 as Post-Production because contractual disagreements usually arise only after production has begun. For example, legal redress is applied only after a contract has been broken. A carrot as well as a stick can induce contract compliance. The expectation of future gains from trade serves as a powerful incentive to live up to the spirit and letter of a contract [Klein and Leffler (1981), Telser (1980) and Winston (1988)].

2.3 Production costs of using the firm

Transaction costs arise within firms for three reasons:

(1) resources in the production process are owned by different parties and owners have an incentive to monitor the use of their assets;
(2) people have limited ability to process and act on information received;
(3) when effort is not directly tied to reward, workers have an incentive to shirk and consequently employers have an incentive to monitor work intensity.

These costs are associated with the factors of production listed in table 1—physical and financial capital, human capital, and work intensity.

Agency costs occur when liquid capital is at the disposal of someone other than the owner. Separation of ownership and control creates the possibility that the manager of the firm may not run the firm in accordance with the owner's interests. Consequently, profitability and capital value of the firm may be adversely affected [Jensen and Meckling (1976), Fama and Jensen (1983)]. When physical capital is used by someone other than the owner,
abusive treatment may arise out of carelessness, haste, or malice. To prevent excessive depreciation of the value of their physical capital, owners will monitor the use of their capital [Alston and Higgs (1982a, pp. 339–340)]. When either abuse of physical capital or agency costs become large relative to economies of scale in production, there is an incentive to engage in more market transactions or to form smaller firms.

Another factor limiting the growth of firms is coordination costs. Coase argued that as the firm gets larger, 'there may be decreasing returns to the entrepreneur function'. In particular, as the number of transactions that are organized increase, 'the entrepreneur fails to place the factors of production where their value is the greatest' [Coase (1937, p. 340)]. Coordination costs can be reduced by contracting out specific portions of the production process [Masten (1984, p. 405)] – thereby increasing market transactions – or by creating hierarchical structures designed to economize on the amount of information required for making decisions [Arrow (1974, pp. 68–70)].

As indicated earlier, when the services of labor are difficult to measure by inspecting output – either because quality is difficult to assess or because of team production – workers are paid by hours worked and firms become residual claimants. Consequently workers have an incentive to shirk and firms have an incentive to monitor to increase work intensity. Shirking or withholding effort applies to unskilled and especially skilled labor. Shirking on the part of skilled labor is more difficult to detect than stinting on tasks requiring physical exertion because problem solving and creativity are not readily observable. Shirking can be thought of as a deviation from expected contractual behavior. Without the ability to enforce contractual behavior, being able to detect shirking would not ensure the viability of internal organizations. In the context of an ongoing employer–employee relationship where the contract is specified incompletely, contract enforcement can include the costs of trilateral and bilateral governance structures designed to work out disagreements. Examples of these arrangements include third party arbitration and union grievance procedures [Williamson (1979), pp. 249–252]).

3. A resource coordination model of the firm

The framework developed in the previous section encompasses the hypotheses suggested by others. Moreover it enables us to derive implications which have been ignored by other theorists. Before advancing our hypotheses about resource coordination we will demonstrate how earlier theories fit into our framework.

6We thank Louis Puttermann for bringing to our attention the different costs of monitoring across skill categories.
3.1 **Transaction costs in the market and the emergence of firms**

The basic postulate of transaction cost economics is that unless there are costs associated with using the market, transactions will not be organized through firms. Organization through a firm creates depreciation, agency, coordination, and shirking costs which will not be incurred unless there are larger costs associated with market transactions.

One source of market transaction costs is asset specificity. If the degree of asset specificity foreseen by the entrepreneur prior to the investment in physical capital is large, organization of transactions within a firm will be chosen to avoid opportunistic behavior. We call this the holdup problem. Support for this hypothesis is provided by Masten's (1984) study of the aerospace industry. Masten found that internal procurement by the prime contractor was preferred to contracting with other firms when the part was highly specialized to the firm's requirements.

The knowledge paradox is another source of transaction costs in the market. If the transfer of information creates a high potential for the appropriation of rents in the process, internal organization will be chosen as the mechanism for transferring knowledge. One of the most conspicuous examples of this phenomenon is the multinational corporation. If knowledge could be transferred without the risks inherent in the disclosure of information, part of the reason for multinational corporations would disappear [Teece (1976)]. Tsokhas (1986) provides a compelling illustration of how the difficulties in transferring technology shaped multinational contracts in the Australian mining industry.

The problems of information transmission lead to another information constraint hypothesis: if information can be more readily transmitted through a firm than through a market, internal organization will be favored. Silver (1984) argues that within firms entrepreneurs need only transmit information about the steps of a process rather than the entire process. This encourages internal production rather than market contracting. The possible economies of scope in know-how suggested by Teece (1980) also foster integration. In addition, high information costs in markets may explain whether the labor market will be 'internalized' within a firm. If information about the productivity of employees is imperfectly transmitted across markets, then internal promotions will be preferred to hiring in the market because information on someone outside the firm may be misleading or inaccurate [Williamson, Wachter and Harris (1975, pp. 273–274)].

The transaction costs associated with the measurement of output may also provide a reason for the creation of firms. If the quality or quantity of output between successive stages of production is difficult to measure, vertical integration will be favored for two reasons. First, the nature of the product

---

7 The following discussion is meant to be illustrative, not exhaustive.
may be such that the most efficient way to measure the output is to monitor inputs and direct the production process [Alchian and Demsetz (1972, p. 785)]. The second reason why firms may be favored is that each time a product changes hands, it must be measured. The costs of repetitious measurement can be avoided if one firm handles all the stages of production, thereby measuring the quality of the intermediate product only once at each stage [Barzel (1982, p. 41)].

Although each of these hypotheses offers a plausible explanation of why firms exist, each one gives us only one piece of the larger picture. In order to complete the transaction cost theory of the firm, the transaction costs incurred within the firm must also be taken into account. The monocausal hypotheses offered above consider only the costs of using the market, and as a result, none can adequately explain either the extent of vertical integration or the final shape of contracts within the firm.

To illustrate this point, consider the following solution to an asset specificity problem.* An entrepreneur who wants to produce a specialized part for bicycles requires a die to be manufactured so that the part can be molded. If the die is owned by the tool and die company which is going to produce the part, an asset specificity problem will exist. The tool and die company could be held up by the entrepreneur if he threatens not to purchase the part, but the entrepreneur could be held up by the tool and die firm if they stop production. The problem is resolved by using a contractual arrangement common in the tool and die industry; after the firm manufactures the die, the die is sold to the entrepreneur. Since the firm no longer owns the die, and the die can be transferred to another firm, the holdup problem disappears. In this case, full scale integration is not required to resolve the asset specificity problem. By looking only at the costs of using the market, however, the asset specificity hypothesis cannot explain why similar problems arising from transaction specific investments are not routinely solved simply by having the entrepreneur buy the capital.

3.2 Resource ownership and the joint determination of transaction costs

To explain specific questions regarding contractual form like the contract between the bicycle manufacturer and the tool and die company, a transaction cost theory of the firm must recognize that production involves coordinating resources owned by different parties. Before an entrepreneur makes the decision to engage in production himself, he must consider the

---

*Information on the following example was acquired by one of the authors through personal experience in negotiating a contract for a bicycle components manufacturer.

The theory of resource ownership and the joint determination of transaction costs was initially developed by Alston and Higgs (1982) to explain contractual mix in agriculture.
The entrepreneur who wants to produce a product has two choices: he can alter his own resource bundle, or he can acquire control over resources which belong to other individuals. Both of these options are costly. Acquiring human capital takes time, and by the time the entrepreneur learns what he needs to know, the opportunity may be lost. Similarly, because no one has unlimited amounts of time or money, the amount of work effort and capital which the entrepreneur can supply is severely restricted. These constraints force the entrepreneur to employ resources belonging to other people.

In acquiring control over the factors of production the entrepreneur also acquires the transaction costs of the firm. Unlike the transaction costs of the market, the transaction costs of the firm are not independent of one another; instead, the transaction costs of the firm are *jointly determined* by who owns the factors of production. Since there are costs associated with altering resource endowments, the ownership of the factors of production can be treated as fixed at any moment in time. Thus, once the decision to engage in production has been made, the existing factor endowments simultaneously determine the structure of transaction costs within the firm.

To generate specific, verifiable implications about contractual form from this observation, a simplifying assumption will be made: Although each of the transaction costs of the firm can influence the others, we will treat the cost of monitoring workers to prevent shirking as the 'dependent variable' while taking the other transaction costs as given. This procedure can be justified by noting that because the work effort of other people can never be owned in the same way that physical capital can be owned, the shirking problem always exists.

The first hypothesis that this approach suggests is that if the entrepreneur supplies physical capital to the production process, the need for the entrepreneur to monitor the use of his capital will substantially lower the marginal cost of monitoring his employee's work effort. Therefore, internal organization will be favored whenever capital is subject to excessive depreciation through abuse or mismanagement because there exist economics of scope in monitoring across assets. With respect to this hypothesis, the earlier example of the entrepreneur and the tool and die firm can be considered to be the exception that proves the rule. What makes the example unusual is that after the entrepreneur buys the die, he allows the firm to use it without his supervision. This arrangement is feasible because, unlike many forms of capital, solid steel dies are difficult to damage. Since the entrepreneur does not need to monitor the use of his capital to prevent abuse, the cost of monitoring employees (which would be necessary if he bought the company) is very high. An example of the diametrically opposite case is provided by the organization of plantations in the Southern U.S. after the
Civil War. When the plantation owner supplied capital in the form of land and mules to his workers, he had to monitor workers to prevent his capital from being damaged. In such cases, the economies of scope arising from the low marginal cost of monitoring work effort resulted in a shift toward wage contracts and away from tenant contracts. [Alston and Higgs (1982a, p. 340)]. This is analogous to a shift towards firm and away from market production.

The second joint factor hypothesis that our framework suggests is that if the entrepreneur provides human capital to the production process, the marginal cost of supervising work effort falls. This effect assumes a central role when we realize that the entrepreneur is usually coordinating resources under conditions of uncertainty. If the entrepreneur is the only one who knows how to respond to contingencies and must be present at the production site to do so, then the marginal cost of supervising workers is very low [Alston and Higgs (1982b, p. 4)]. A more familiar, though slightly different approach to the joint aspects of coordination and monitoring costs is provided by Alchian and Demsetz's concept of team production. When the entrepreneur must monitor his employees to determine the most productive team combination of inputs and meter rewards, then the marginal cost of supervising those workers to prevent shirking is reduced [Alchian and Demsetz (1972, pp. 779–780, 793)].

To illustrate our jointness hypothesis consider an entrepreneur who faces the make or buy decision (see fig. 1). On the vertical axis is $P_b$, the relative price of a market transaction. On the horizontal axis is the number of market transactions. When more market transactions take place, less is produced internally by the firm.\(^{10}\) The demand for market transactions depends on both their price $P_b$ and the internal transaction and production costs incurred by the firm. With a lower price, more market transactions will take place. Similarly, as the costs of internal production rise due to either rising production or transaction costs, the demand for market transactions will rise. The supply of market transactions is perfectly elastic at $P_b$. $P_b$ includes the transaction costs of using the market which form a wedge between what the seller charges and the total costs to the buyer.\(^{11}\) The relevant costs affecting the demand and supply of market transactions are listed in table 1. Obviously the greater the transaction costs of using the market – holdups from asset specificity, information constraints, measurement costs and contract enforcement costs – the bigger the wedge between the price offered by the supplier and the price faced by the buyer. As a consequence, a 'make' decision is more likely. Conversely, when the transac-

---

\(^{10}\)For simplicity we view production as fixed so any increase in market transactions will decrease internal production.

\(^{11}\)In a way, the transaction costs of using the market are analogous to transportation costs.
tion costs of internal production – abuse, agency, coordination and monitoring costs – are greater, a ‘buy’ decision is more likely. Changes in transaction costs shift either the demand or supply of market transactions.

This framework is still incomplete for it fails to consider the jointness across transaction costs. In particular who owns what factors of production affects the costs of monitoring shirking. For example if the entrepreneur supplies (owns) the physical capital this will increase the likelihood of internal production because it reduces the costs of monitoring. Why? Because there are economies of scope in monitoring. An entrepreneur can simultaneously monitor for the abuse of her capital and labor shirking. In the above diagram this will shift the demand for market transactions to $D'$ with the result that market transactions will decrease to $q_m$ and firm transactions will increase. The difference between $D$ and $D'$ is a recognition of the economies of scope in monitoring. Viewing transaction costs singularly can overstate the total transaction costs. The distribution of human capital within the firm, for instance, may affect monitoring costs. Most employees hired in managerial capacities contribute effort to the production process in addition to their role in monitoring other employees. A recent study indicates that ‘supervisors spend barely one and a half hours actually supervising subordinates’ [The Wall Street Journal (1987, p. 1)].

If effort and supervision do not interact, the joining of work effort and monitoring would be difficult to explain in large organizations. Why not simply hire people who specialize in each function? If economies of scope exist between applying work effort and supervising employees, however, then only by utilizing working managers can the firm internalize its jointness effect on supervision and reduce the transaction costs of the firm. Models which treat managers only as monitors neglect this interaction between working and monitoring. If this effect is important, a major factor in determining the
structure of the firm's internal hierarchy has been omitted from the models which attempt to explain such hierarchies.\footnote{See for example Alchian and Demsetz (1972), Calvo and Wellisz (1978), Holmström (1982) and Bohn (1987).}

Application of this jointness hypothesis is not limited exclusively to the transaction costs of the firm; the transaction costs of the market can also act in conjunction with the transaction costs of the firm. Problems associated with asset specificity, for example, may result in the purchase of a piece of physical capital, but having purchased the capital, the entrepreneur may need to monitor its use to prevent abuse. When such monitoring takes place, the marginal cost of supervising workers falls and the demand for market transactions falls. Similarly, when material inputs and the production process must be monitored because the quality of output is difficult to measure, the marginal cost of monitoring shirking is lower than it otherwise would be, thereby decreasing the demand for market exchange.

3.3 Changes in transaction costs

The model presented in the previous section demonstrates how an owner-managed firm balances the transaction costs of internal production and obtaining inputs from the market. Implicit was the assumption that the factor endowments and the technology of production were fixed. How firms adapt over time to changes in transaction costs has been the subject of some debate. Some writers, including Williamson (1981b, 1985) have argued that governance structures evolve in order to minimize the transaction costs. One example of such evolution may be the change in corporate structure from a unified arrangement (the U-form) to a multidivisional form (the M-form) [Williamson (1985, p. 295)]. This type of dynamic argument usually rests on the assumption that optimal organizational design, organizational learning, or competitive market pressure will force firms to minimize the sum of transaction and production costs. This approach has been criticized by Dow (1987, pp. 25-33) who argues that governance structures may not always evolve toward efficiency if considerations like bounded rationality, small numbers bargaining, or sunk costs are present in organizations.\footnote{We find Dow's criticisms provocative, but beyond the scope of our paper. We do not argue that the organizational form that exists is the most efficient, but rather, given the organizational forms in existence (e.g. capitalist firms versus market), actors will choose the more efficient of the two. Functional explanations of the sort we posit have merit when certain conditions are met [see Elster (1983), especially pp. 55-58]. In our case, the threat of takeovers ensures that actors choose the most efficient governance structure of those in existence.}

Our interpretation of the impact of transaction costs starts from the observation that organizational forms (broadly conceived as firms or markets) have not changed much over the past 100 years. From this observation we focus on the impact that transaction costs have on the
L.J. Alston and W. Gillespie, Resource coordination and transaction costs

preponderance of contracts observed within and between economic units. One example of this type of change might be the shift from tenant to wage contracts noted previously. When changes in transaction costs are not large enough to warrant a new type of governance structure, the types of governance structure may be held fixed and an exercise in comparative statics can reveal how the firm/market boundary shifts in response to a change in transaction costs (fig. 1). When such changes are sufficiently small, the considerations impeding change noted by Dow (1987) may be less significant so that movements over time in the firm/market boundary may reflect changing relative efficiencies.

From the resource coordination perspective, changes in the technology of production and factor endowments are particularly interesting because such changes may alter the transaction costs and/or production costs associated with the firm or the market. Changes in transaction costs resulting from changes in the ownership of the factors of production. Changes in transaction costs could strain existing contractual agreements. When a large amount of human capital is acquired during production, a worker may want to alter the terms of his employment to reflect this change. Renegotiation of contracts may be difficult and expensive as Dow (1987) argues, especially if both sides have an incentive to behave in an opportunistic fashion. Such costs, however, should be considered as part of the transaction costs of the firm. If such costs rise substantially, there will be an incentive for one or both parties to alter the firm's status quo. If one of the major advantages of team production within the firm is the joint production of coordination and monitoring by the owner, then the acqui-

14Changes in transaction costs can also affect the adoption of technology. For a discussion of the interplay between technological choice and transaction costs, see Alston and Robertson (1988).
sition of human capital by the worker may decrease his need to have someone else supply direction and coordination. Self-employment becomes more attractive since the employee would no longer have to incur the cost of someone else supervising his work. Thus, in a business where there are not other major obstacles to entry, establishing another firm may be a viable alternative. In such a case, the resource coordination model would predict that existing firms may disintegrate if workers acquire significant amounts of human capital and the transaction costs of the firm (in terms of the renegotiation of contracts, and so forth) increase substantially. One example of this phenomenon is computer software companies. After an employee (who may have been hired for his knowledge of computers) learns about the business of selling software, he may strike out on his own and set up his own company. Similar examples can also be cited from the fields of management consulting and accounting.

As a final observation, we note that in some cases, transaction costs can be reduced without any necessary implications for whether the transaction will occur within a market or within a firm. The prime example of this result is provided by the literature on relational exchange. As Telser (1980) and Klein and Leffler (1981) observe, repetitious exchange between parties which possess specific identities can mitigate opportunistic behavior and make market exchange feasible. And as Winston (1988) notes, in a repetitive series of exchanges, the assumptions of neoclassical economics with respect to perfectly informed parties making rational, profit maximizing decisions are fulfilled. In such cases, the transaction costs of using the market may virtually disappear. Relational exchange, however, can reduce the transaction costs of the firm as well by reducing abuse, agency and shirking costs. When employees have some bonds — whether of affection or fear — with their employers, they may be less likely to impose costs on the firm. In part this may explain the success of family businesses and the ability of family farms to out-compete corporate farms until recently.

4. Competing frameworks

Our framework is superior to other theories for two reasons. First, most other explanations for the firm/market boundary rest on a single transaction cost. Our framework is not monocausal and enables us to encompass existing explanations. Second, our framework illuminates the jointness of transaction costs that arise from who owns what in the production process.

\[15\] For the initial treatment of relational exchange, see Goldberg (1980).

\[16\] A comprehensive discussion of the issues related to specific identity can be found in Ben-Porath (1980).
Because Williamson (1979, 1981a, 1981b and 1985) provides the most comprehensive alternative framework, we offer some criticisms of the weaknesses in his framework. Williamson views the market and the firm as alternative governance structures. Which governance structure — firm or market — is preferable depends on the frequency, uncertainty, and degree of asset specificity involved in the transaction [Williamson (1979, pp. 247-254)]. Williamson argues that in the absence of transaction-specific investments, firms deciding whether to make or buy a component will usually choose to buy the component in the market because markets have several advantages over internal organization. By buying rather than making, the firm can exhaust static scale economies and avoid the bureaucratic costs associated with internal organization. In addition, markets can pool risks by aggregating uncorrelated demands and they can realize economies of scope by providing a related set of activities [Williamson (1981b, p. 1547)].

An investment in assets specific to a transaction locks both the buyer and the seller into a bilateral monopoly because the cost of the asset cannot be fully recovered if the transaction is terminated [Williamson (1979, p. 240)]. According to Williamson, this creates an incentive to organize transactions within firms. Williamson identifies frequency and uncertainty as the two other 'dimensions' of transactions, but in his model, the frequency and uncertainty of transactions affect the choice of governance structure only insofar as they appear in conjunction with asset specificity. When transactions requiring specific assets occur infrequently, Williamson argues, the cost of elaborate governance structures cannot be justified. Arbitration through a third party (trilateral governance) will be used to resolve disputes. On the other hand, with frequent transactions, Williamson suggests that complex bilateral agreements or vertical integration will be preferred depending on the degree of asset specificity [Williamson (1979, pp. 247-253)]. In his model uncertainty about future conditions primarily affects transactions with an intermediate degree of asset specificity; changing conditions impose strains on market agreements regarding specific assets. Since firms enjoy advantages over markets in writing and executing contracts under conditions of uncertainty, vertical integration may result. Within a firm, common ownership of assets reduces the incentive to behave opportunistically, and costly negotiations over contractual changes can be avoided because the firm can resolve disputes through fiat [Williamson (1981b, pp. 1548-1549)].

Basing a model of the firm on the transaction costs of asset specificity, however, poses several problems. In particular, while the hazards connected with specific assets undoubtedly explain why firms choose to expand by investing in specialized assets prior to the actual production process, the asset specificity argument alone cannot explain post-production vertical integration. To illustrate this point, suppose that firm A invents or produces a machine to produce a specialized part for firm B. If that part is critical to
firm B’s production, firm A can hold up firm B and extract all the quasi-rents associated with that machine. If firm B attempts to vertically integrate its production process by acquiring firm A, the owner of firm A will attempt to extract the present value of all the quasi-rents arising from his monopoly power [Alston and Higgs (1982b, p. 3)]. Despite the extraction of quasi-rents by firm A, it may be profitable for firm B to purchase firm A. It depends on whether firm B can run the operations of firm A more profitably as part of its organization than firm A can run itself. Firm B may possess a cost advantage arising from the joint determination of transaction costs. For example if measuring the quality of firm A’s output is difficult without direct supervision, then redundant measurement costs can be reduced through integration. Once direct measurement is necessary to assure quality the marginal cost of monitoring shirking falls. An additional rationale for vertical integration arises from the deadweight loss which firm A may impose on firm B. If firm B’s production technology is variable proportions, firm A will induce a distortion in factor utilization if it monopoly prices [Vernon and Graham (1971), p. 924]. Integration of the two firms would permit optimal factor utilization and thereby increase profits.

With respect to specialized human capital, the holdup problem never disappears unless slavery is permitted. If both the worker and his employer are aware of the value of the worker’s specialized human capital, the fact that the transaction is organized within the firm will not prevent each from trying to extract the full value of those quasi-rents [Alston and Higgs (1982b, p. 3)]. Williamson, Wachter and Harris (1975, p. 270) argue that this type of behavior is reduced by ‘shifting to a system where wage rates attach mainly to jobs rather than to workers’, but this response does not address the question of why workers don’t negotiate their wage rates more frequently. A more plausible answer offered by Williamson, Wachter and Harris (1975, p. 273) is that workers do not haggle because the firm creates powerful incentives to be a ‘team player’ through an internal promotion ladder. The benefits of long term contractual relationships, however, need not be realized exclusively within firms; relational exchange and personal trust relations evolve in market relationships as well. Another equally plausible reason why workers do not attempt to extract the rents commensurate with the full value of their capital might be that information asymmetries exist between employers and employees. If an employee does not recognize the value of his specialized capital, the employer may be able to ‘exploit’ that ignorance.

Another problem with Williamson’s framework is that the effects of frequency and uncertainty are not as clear as he assumes. When the frequency of transactions increases, market governance may be employed even when specific assets are involved. In long term relationships, opportunism is mitigated because both parties stand to gain from continuing the relationship. Opportunism, though potentially lucrative in the short run, may
result in a net loss if the victimized party and other potential customers refuse to do business with an opportunist [Telser (1980)].

Similarly, uncertainty about future events does not preclude contractual arrangements designed to handle unspecified contingencies. Some contracts contain clauses which, in effect, stipulate that 'the parties will behave reasonably in the event of unforeseen circumstance'. Agricultural tenancy contracts in the Southern U.S. for example, commonly required the tenant to act in a ‘farmer like manner’ or ‘to work faithfully and diligently...to cultivate the land properly’ [Reid (1973, pp. 117–118)]. Contracting parties knew what these general provisions required of them under specific contingencies, as did the local courts. To the extent that the spirit as well as the letter of the law is enforceable under common law, opportunism may be precluded [Alston and Higgs (1982b, p. 4)].

5. Toward defining the firm

Although we have discussed ‘the firm’ and ‘the market’ at great length, we have not offered a definition of either. The question of what constitutes a firm has troubled many economists because the wide range of contracts observed makes it difficult to define, precisely, what is and what is not a firm. This lack of consensus reflects, in large measure, the diverse and sometimes contradictory views of the firm which all share a transaction cost orientation. From the perspective of Alchian and Demsetz, for example, defining the firm and the market is a rather futile exercise because fundamentally, there is no difference between the two. Alchian and Demsetz (1972, p. 777) argue that ‘Telling an employee to type this letter rather than that document is like my telling a grocer to sell me this brand of tuna rather than that brand of bread’. In contrast to this view, the resource coordination model suggests that there is a difference between buying a can of tuna and directing an employee to type a letter. When I buy a can of tuna from my grocer, I am not contributing any of the factors which make production possible. That can of tuna will be on the shelf whether or not I decide to buy tuna on that particular day. When an employer tells his secretary to type a letter, however, both parties are contributing resources to the production process. The employer supplies directions about what to type, and, typically, the typewriter as well. The secretary supplies human capital in the form of typing skills and work effort. Unless all of these factors come together in the proper fashion, the letter will not be produced.

This distinction suggests that one way to define a firm is to begin by defining a market transaction. A transaction can be classified as a pure market transaction when a good or service is produced exclusively by one party and sold to another party which did not contribute any of the factors of production. A market, therefore, is simply a collection of such transac-
A firm, on the other hand, is an organization which provides conscious direction in coordinating physical and financial capital, human capital, and work intensity in order to produce goods and services for sale. In the context of the resource coordination model of the firm, a pure firm can be distinguished by two characteristics. First, a pure firm acquires all of its physical inputs through pure market transactions. Second, a pure firm assumes all of the transaction costs related to the production of goods and services (see column 2, table 1). A pure firm does not exist if either the agency costs from supplying capital, the coordination costs from supplying human capital, or the monitoring costs associated with work intensity are borne by an independent organization which subcontracts for part of the production process.

This resource coordination definition of the firm has much in common with Cheung's (1983) approach, but it differs from Cheung's view in several important respects. Like Cheung, the resource coordination model pictures the firm and the market as two ends of a contractual continuum. But in between these extremes Cheung's approach does not allow one to make distinctions. For example, Cheung argues that a construction company which subcontracts the roofing, the electrical wiring, and the plumbing out to three independent contractors is just as much a firm as another company which does all of that work with its own employees. In our resource coordination view, however, the construction company that does not subcontract differs from the one that does in that it has internalized all of the transaction costs associated with the organization of a firm. By providing an objective criterion to differentiate contractual variations (i.e. the degree to which an organization assumes the transaction costs of the firm), the resource coordination definition enables us to distinguish among different types of firms along the contractual continuum.

The second major difference between Cheung's approach and ours is that the resource coordination view of the firm is much more narrowly focused than Cheung's contractual view. Cheung (1983, p. 17) argues 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'. From a resource coordination perspective, this elastic view of contracts obscures many of the important differences in transaction costs that arise from the ownership of the factors of production.

In offering this definition of the firm, our discussion has, in one sense, returned to the point where it began. In 1937, Coase offered a definition of what characterizes a firm and what characterizes a market, and in many respects, our view of the firm is very similar to Coase's definition. Like Coase, the resource coordination model treats the firm and the market as two distinctly recognizable types of organization between which it is difficult to draw any 'hard and fast line'. Nevertheless, we also believe that the
resource coordination model is more than just a synthesis of work that has gone on in transaction cost economics since Coase's 1937 paper. Besides incorporating the elements of other transaction cost theories, the resource coordination model distinguishes between the transaction costs of the firm and the transaction costs of the market. That distinction, in turn, highlights the way in which the distribution of the factors of production can simultaneously determine the organizational structure of the firm. Only by looking at the way in which the transaction costs of the firm and the market are jointly determined can a transaction cost theory of the firm begin to explain the firm/market boundary.

References

North, D.C., 1984, Transaction costs, institutions, and economic history, Journal of Institutional and Theoretical Economics 1, 7–17.
Silver, M., 1984, Enterprise and the scope of the firm (Martin Robertson Press, Oxford).
Teece, D.J., 1976, The multinational corporation and the resource cost of international technology transfer (Ballinger, Cambridge, MA).