

The Theory of the Firm¹

by

Nicolai J. Foss

Department of Industrial Economics and Strategy
Copenhagen Business School

Henrik Lando

Department of Finance
Copenhagen Business School

Steen Thomsen

Institute of International Business
Aarhus Business School

August 24, 1998

¹ The authors have written this article in close cooperation and share joint responsibility for all parts of it.

I. Introduction: the Emergence of the Theory of the Firm

Along with households, firms have for a long time been a crucial part of the explanatory set-up of economics. For example, in basic price theory, firms are part of the apparatus that helps us trace out the effect on endogenous variables of changes in exogenous variables. But the explanatory atoms, firms and households, are not themselves treated in any detail. Thus, we have for a long time had an economics *with* firms, as it were; what was missing until the nineteen-seventies was an economics *of* firms. It is only relatively recently, in other words, that economists have felt the need for an economic theory addressing the reasons for the *existence* of the institution known as the (multi-person) business firm, its *boundaries* relative to the market, and its *internal organization* – to mention the issues that are generally seen as the main ones in the modern economics of organization (e.g., Milgrom and Roberts 1988; Hart 1989; Holmström and Tirole 1989).

Although pioneering early work was done already by Frank Knight (1921) and Ronald Coase (1937), it was not until the mid-1970s that work really blossomed within the field, stimulated by advances in the economics of market failures, property rights, information and uncertainty which made possible a more rigorous understanding of the sources and nature of transaction costs and of the incentive properties of alternative types of economic organization. The work of Coase did not belong to this formal stream of work, and as late as in 1972, Coase lamented that his 1937 paper had been “much cited and little used”.

However, at the time of Coase’s lamentation, serious work on the theory of firm had begun to take off, notably with Williamson (1971) and Alchian and Demsetz (1972), two seminal contributions that helped found distinct perspectives within the modern economics of organization. It is fair to say that today organizational economics is one of the richest and most rapidly expanding fields in modern economics. It may be seen as part of broader attempt (sometimes called “new institutional economics”) to move beyond the confines of the market institution and also inquire into the rationales and functioning of alternative institutions for resource allocation, generalizing standard neoclassical economics in the process (Arrow 1987; Eggertson 1990).

The pure analysis of the market institution leaves almost no room for the firm (Debreu 1959). Under the assumption of a perfect set of contingent markets, as well as certain other restrictive assumptions², the model describes how markets

² The most important of the technology constraints is that production possibilities should be convex (i.e. that linear combinations of feasible production plans should also be feasible) which implies non-increasing returns to scale. Small firms are assumed to produce as least as efficiently as large ones. This is consistent with the competition constraint that firms act as

may produce efficient outcomes. The question how organizations should be structured does not arise, because market-contracting perfectly solves all incentive and coordination issues. By assumption, firm behaviour (profit maximization) is invariant to institutional form (e.g. ownership structure). The whole economy can operate efficiently as one great system of markets, in which autonomous agents enter into very elaborate contracts with each other. However, by treating the firm itself as a black box, where internal structure, contracts, etc. disappear from the picture, there are many other issues that the theory cannot address. For example, the theory does not tell us why firms exist.

As already indicated this is not a new recognition, but rather a basic point in Coase (1937). What Coase observed was indeed that, in the world of neoclassical price theory, firms have no reason to exist. According to the textbook, the decentralized price system is the ideal structure for carrying out economic coordination. Why then do we observe some transactions to be removed from the price system to the interior of organizations called firms? The answer, Coase reasoned, must be that there is a “cost to using the price mechanism” (Coase 1937: 390). Thus was born the idea of transaction costs: costs that stand separate from and in addition to ordinary production costs. Firm organization may avoid these costs, and exists for this reason. However, as is well-known, Coase’s insights were neglected for more than three decades.

During the 1960s, standard neoclassical theory was criticized for ignoring conflicts of interest between owners and managers, a point reaching back to Berle and Means (1932) (Marris 1963, Williamson 1964). Thus, the standard theory assumed that profit maximization was the main objective of managers. At the same time, critics from behavioral studies attacked the assumption of perfect rationality. In their view, the existence of organizations such as firms was primarily a matter of economizing with bounded rationality (Simon and March 1959; Cyert and March 1963). A little later, Williamson (1971, 1975) following Coase (1937) questioned the view (which he called “technological determinism”) that technologies are determinative of economic organization. This did not, according to Williamson explain, why the services of production factors are coordinated in a firm rather than offered to the market by self-sufficient factor-owners. More or less simultaneously, Kenneth Arrow’s work

price takers, that is, that no firm is large enough to influence the market price. In contrast, economies of scale imply that large firms can produce at lower units costs which is likely to lead to market concentration and alternative modes of competition (oligopoly or monopoly). This “technological theory” was refined by Baumol, Panzar and Willig (1982) and Panzar (1989) who reexamine the relationship between technology (production possibilities) and cost structures. They specify exact technological conditions under which large firms may enjoy cost advantages from economies of scope (multiple products) as well as scale (volume of the same product). Panzar (1989) appears to follow Koopmans (1951) in attributing economies of scale or scope to indivisibilities.

on welfare economics and information (Arrow 1969, 1974) emphasized various limitations of the market mechanism and argued that firms can be understood in terms of market failures which arise under conditions of externality, economies of scale and information asymmetries.

All these influences and insights set up a rich agenda for research. Work on the theory of firm (as it has been defined here) began to take off during the 1970s, notably with seminal contributions by Williamson (1971, 1975), Alchian and Demsetz (1972), Ross (1973), Arrow (1974), Jensen and Meckling (1976), and, from a rather different perspective, Nelson and Winter (1982; summarizing their work in the 1970s). It is fair to say that the emerging economics of organization is now one of the richest and most rapidly expanding fields in modern economics. It may be seen as part of broader attempt (sometimes called “new institutional economics”) to move beyond the confines of the market institution and also inquire into the rationales and functioning of alternative institutions for resource allocation, generalizing standard neoclassical economics in the process (Arrow 1987; Eggertson 1990).³

While there exist various streams of research, they share the basic principle that economic organization, including paradigmatically the choice between firm or market, will turn on the maximization of joint surplus from cooperation in production (Coase 1960). Although modern theories of organization are partial equilibrium theories, and although they emphasize bilateral aspects of transactions, many of them have a foundation in general equilibrium theory, both historically and conceptually (Hart and Holmström 1987; Guesnerie 1994). While this may not be historically true of the contributions of, for example Coase and Alchian and Demsetz, nevertheless the various approaches to the theory of the firm can be *rationalized* as different departures from the Arrow-Debreu model. This is the interpretation we adopt in the following.

Since firms would not exist in a full and perfect information setting (implying perfect and costless contracting), one can understand the existence of firms as a consequence of one or more of the Arrow-Debreu assumptions not holding true. We focus in on two basic Arrow-Debreu assumptions which we use to structure our account of the different theories of the firm.

1. The assumption of complete contracting: Agents can foresee all future contingencies and can costlessly write contracts which cover all contingencies. Thus, there are no incomplete contracts.
2. The assumption of symmetry information concerning “states of nature”. Thus, there are no principal-agent incentive problems.

Theories of the firm can accordingly be classified in a rough way into two

³ In this connection, it should be mentioned that what we here call “the theory of the firm” really is a general theory of economic organization that also include, for example, “intermediate forms”, such as franchising arrangements or joint-ventures.

categories:

1. *Incomplete contracting models* which are founded on the assumption that it is costly to write elaborate contracts, and that there is therefore a need for *ex post* governance. The work of Williamson, Grossman-Hart-Moore, Coase, and Simon, as well as implicit contract-theories, and the theory of communication in hierarchies, belong to this category.
2. *Principal-agent models* which allow agents to write elaborate contracts characterized by *ex ante* incentive alignment under the constraints imposed by the presence of asymmetric information. The work of, for example, Alchian and Demsetz, Holmström, and Milgrom belong in this category.

One way of interpreting this division of the literature is to say that they have concentrated on different kinds of the transaction costs that Coase (1937) identified but didn't elaborate on. Thus, where incomplete contracting theories typically emphasize the costs of writing costs and the costs of *ex post* adaptations, the principal-agent traditions neglect these costs and emphasize instead costs of monitoring and costs of setting up incentive arrangements (costs that are largely neglected in the incomplete contracting tradition). Clearly, these perspectives are complementary, and should be integrated. There are signs that this integration process is slowly beginning (e.g., Holmström and Milgrom 1994). However, we stick in this paper to the above dichotomization,⁴ and after surveying theories in each category, we end up discussing prospects and criticisms.

II. Complete Contracting and Economic Organization: Principal/Agent Theories

A. General Aspects of Principal/Agent Theories

Historically, principal/agent theories (or simply, agency theory) reach back to early debates on the shareholders/managers relation. Following the observation by Berle and Means (1932) that ownership of US firms had become separated from management and control, managerialist theories have modeled firm behavior as the maximization of managerial objectives (firm size, growth) under a profit constraint (Marris and Mueller 1980; Williamson 1964). Managerial objectives are believed to be variables like firm size or growth, partly because they are positively correlated with managerial compensation

⁴ For more on the disagreement between the two modeling strategies, see Tirole (1994).

and power. The conflict of interest between owners and managers is but one example of an incentive or principal-agent conflict which arises in the context of firms and which may explain important aspects of their organization.

During the nineteen-seventies formal agency theory blossomed (Ross 1973 is an early contribution). In its paradigmatic version, the theory deals with a relationship between a principal (e.g. the owner) and an agent (e.g. the manager) who works on a well-defined task, although the model can be generalized to encompass, for example, many agents, hierarchical layers (for example, a middle-manager who is both agent and principal), and multiple tasks. Indeed, much of the formal agency work in the nineteen-seventies consisted in such extensions of the basic model (see Hart and Holmström 1987).

A basic assumption is that some information asymmetry exists between the two, so that the principal cannot directly observe the activities of the agent or that the agent knows some other aspect of the situation which is unknown to the principal. The literature makes a distinction between “hidden action” and “hidden information” models (Arrow 1985). The principal may, of course, infer something by observing output (e.g. profits), but under uncertainty output is an imperfect signal of the agent’s actions. Under these conditions the first best (output maximizing) contract would be to let the agent compensate the principal with a fixed lump sum payment and to be awarded the residual; however, risk aversion on behalf of the agent may rule out this solution. Other solutions are possible but all entail cost. In practice it has been suggested that managers are disciplined (i.e. induced to behave efficiently) by takeover threats (Manne 1965; Jensen 1986), product market competition (Hart 1980), debt pressure (Jensen 1986) and competitive managerial labor markets (Fama 1980).

B. The Nexus of Contracts View

Alchian and Demsetz (1972), Jensen and Meckling (1976), Barzel (1997), and, most forcefully, Fama (1980) and Cheung (1983)⁵ argue that from the perspective of economics, it is essentially misguided to draw a hard line between firms and markets. Although firms are surely legal entities, and although this of course has important economic consequences (e.g., limited liability, the right to deduct input purchases from tax statements, infinite lifetime, etc.), that firms are nevertheless merely special kinds of market contracting. What may distinguish them relative to other market contracts lies primarily in the continuity of association among input owners.

In Alchian and Demsetz’s formulation, the authority relation associated with the employment relationship is e.g. in no way the defining characteristic of the

⁵ All these contributions explicitly draw on property rights theory (e.g., Coase 1960; Alchian 1965).

firm. An employer has no more authority over an employee than a customer over a grocer. Both the employer and the customer rely on their ability to punish “disobedience” by firing, that is, by no longer dealing with their counter-part. A customer “firing” a grocer is not in economic terms any different from an employer firing an employee, it is asserted. In short, the firm is nothing but a nexus of contracts, backed up by special legal status and characterized by continuity of association among input owners.⁶ We may talk about a nexus of contracts being more “firm-like” when, for example, residual claimancy becomes more concentrated, but it is not in general productive to talk about “firms” as distinctive entities.

While there is much truth in the view that defining the exact firm/environment boundary is theoretically problematic, it would seem that this view “under-determines” real world firms: authority does seem to have a real existence and firms seem to be characterized by more than their legal status, for example, some argue, by specific technologies. While we treat the first of these two issues later, we briefly consider the second one in the following section.

C. The Firm as a Solution to Moral Hazard in Teams (Alchian-Demsetz and Holmström)

Alchian and Demsetz (1972) emphasize that the firm is indeed characterized by something more than legal status, namely the technology of team-production, by which they mean production with inseparable individual production functions. This implies, they argue, that marginal products are costly to measure. This may create a free rider problem since team-production can be a cover for shirking. The solution to this problem is to appoint a monitor who is given the right to fire and hire members of the team, based on his observation of employees’ marginal products. Giving him rights to the residual income of the team furthermore means that he is given incentives to perform the efficient amount of monitoring. This arrangement results in the distribution of rights known as “the classical capitalist firm” (Alchian and Demsetz 1972).

In a formal contribution, Holmström (1982) discusses the incentive problems of monitoring and possible solutions (abstracting from team synergies by assuming an additive production function). Under the assumption that the monitor is uninformed about individual effort levels under team-production, Holmström demonstrates that only under restrictive assumptions will the monitor be able to induce efficient effort levels. He can do this by devising clever incentive mechanisms. Specifically, Holmström proves that in a team-

⁶ In a recent variation over this theme, Yoram Barzel (1997: 81) argues that one should “... define the firm by its guarantee capital and by the scope of its guarantees. The scope of the firm comprises the set of contracts whose variability is contractually guaranteed by common equity capital. The firm, then, is a nexus of outcome guarantees”.

production situation with unobservable effort levels, the triple requirement of the incentive system of Nash equilibrium, budget balancing (that is, the revenues should be distributed among the team-members by the incentive system) and Pareto optimality cannot be met. A budget-balancing incentive system cannot reconcile Nash equilibrium and Pareto optimality. This is simply because every team-member equalizes marginal costs and benefits of additional effort: if one team member's effort generates some extra revenue for the team, he should be given that revenue in order to be properly motivated – but this cannot be done for all team-members under budget-balancing. In this perspective, the central advantage of the firm is that third parties can inject capital whereby the employees as a group do not have to balance their budget.

However, neither the theory of Alchian-Demsetz nor that of Holmström provides us with a theory of the boundary of the firm. The basic question why the incentive problems cannot be just as well solved in the market through contracting as within firms cannot be answered by their theory. Thus, in Alchian-Demsetz' theory it is not clear why the monitor must be the employer of the firm where he performs his monitoring services. He could be the employee of a firm, specialized in monitoring services. Moreover, why aren't the employees able to monitor each other? Is it really meaningless to speak of authority if the employer/monitor has the right to deprive the employee of the right to work with his tools and equipment to which the employee may be strongly specialized? Still, as will become clear, unmeasurability of individual effort and contribution can be made part of a theory which explains the boundary of the firm (this is done in the Holmström-Milgrom model).

D. The Firm as an Incentive System (Holmström and Milgrom)

The connection between ownership and employment is examined by Holmström and Milgrom (1994) who stress the importance of viewing the firm as “a system”, specifically as a coherent set of complementary contractual arrangements which mitigate incentive conflicts. In their opinion, it is misleading to focus on any one single aspect of the coherent whole: the firm is characterized by the employee not owning the assets, by the employee being subject to a low-powered incentive scheme,⁷ and by the employee being subject to the authority of the employer.

These “incentive instruments” are *complementary*. For example, in the presence of measurement costs (Barzel 1997), it is important that a person who does not own the assets which he uses is not subject to high-powered incentives, since he then is likely to care too little for the assets. Likewise, low-powered incentives

⁷ The subject of low-powered incentives was analyzed in Holmström and Milgrom (1991), in which such incentives are shown to make sense in the presence of costs of measuring tasks that are vital to the firm.

make it important for the employer to be able to exercise authority over the use +of the employee's time, since the employee will lack the proper incentive to be productive. Due to this complementarity it is logical that independent contracting has the exact opposite constellation of instruments from the employment relationship.

The choice between the two different incentive systems depends importantly on the extent to which every dimension of a person's contribution can be measured. When an important dimension is unmeasurable, it might be counterproductive to remunerate the person through a high-powered incentive scheme since the person is likely to allocate too little attention on the unmeasurable activity. Thus, according to Holmström and Milgrom lack of measurability is an important variable determining the size of the firm (see also Barzel 1997). They cite empirical evidence that sales-agents (the sum of whose productive contributions can be measured with relatively high accuracy) are independent and *vice versa*.

It should be mentioned that the Holmström and Milgrom model also incorporates the importance of the allocation of property rights to physical assets in determining bargaining powers and hence incentives, as in the class of models we consider later, those associated with the work of Williamson and Grossman-Hart-Moore model. Hence, it is not only a principal-agent but also an incomplete-contracting theory, and perhaps a sign of an increasing awareness of the need to join ideas from the complete contracting tradition with ideas from the incomplete contracting tradition. We consider the latter next.

III. Incomplete Contracting Theories

A. General Aspects

This section reviews theories that may be reconstructed as departing from the assumption of complete contracting. Thus, for different reasons, not all contingencies can be contractually covered. This makes it possible to theorize about authority and asset-ownership. The view of Coase (1937) and Simon (1951) that the essence of the firm is the employment contract and the associated authority relation also belongs to the incomplete contracting category. This is because it emphasizes the costs of entering into a comprehensive contract and the consequent need of adaptation to changing circumstances that are not covered by the contract. The work of Williamson and Grossman and Hart also belong here, although their focus is somewhat different: they focus in on the bargaining power that asset-ownership may provide in those situations in which contingencies arise that are not covered by

the contract.

B. The Authority View: The Firm as an Employment Relation

The view that the employment contract is the characteristic which defines the firm is generally associated with Coase (1937) and Simon (1951). From this point of view, the size of the firm (the boundaries of the firm) is determined by the number of employees of the firm. An employee is distinguished from an independent contractor by the nature of his contract: While the employee is subject to the authority of the manager of the firm, an independent contractor acts autonomously. Coase sees the advantage of establishing an employment contract, that is, an semi-authoritarian (hierarchical) relationship, in the saving of transaction costs, not least the costs of haggling over the terms of the contract. The disadvantage that is also part of the hierarchical relationship is seen as “information overload”: as a company expands, the manager will find it difficult to direct the actions of all employees subject to his authority, because he will be incapable of gathering all relevant information. This helps establishing the location of the boundaries of the firm.

Simon defines the employment relationship more closely and compares its efficiency with the efficiency of a contract written between two autonomous actors. The latter contract specifies the action to be performed in the future and its price while the employment contract specifies a range of acceptable orders and establishes the right of the employer and the duty of the employee to accept orders within this range. The advantage of the employment relationship lies in its flexibility. The action of the employee can be adapted to whatever state of nature will occur. Intuitively, the benefit of flexibility is greater the greater the uncertainty. On the other hand, Simon stresses that the employment relationship is to some extent reliant upon the employer’s reputation for not abusing his authority. The need for trusting the employer is less if the employee is nearly indifferent between different tasks.

Simon’s theory was criticized by Williamson (1975) for favorably comparing the flexibility of a short term (spot market) employment contract to the rigidity of a (long term) non-contingent output contract. But conceivably the flexibility of an output contract contracts could also be achieved by a series of spot market contracts.

A modern formulation by Wernerfelt (1997) overcomes some of the objections that can be raised against a theory of the firm that builds on Simon’s theory. Wernerfelt thinks of governance mechanisms as institutional mechanisms (gameforms) according to which players adapt to changes in the environment and communicate about these changes. His conjecture here is that different gameforms will be systematically characterized by different levels of costs of making adaptations. For example, in the case of the hierarchy, the employer and the employee avoid the costs of negotiating either a very complex

agreement or a series of short term contracts. Instead, the parties negotiate a once-and-for-all wage contract. In this context, authority is simply an implicit contract which states that one of the parties should have the authority to tell the other what to do (as in Coase 1937). This gameform requires less bargaining over prices than the market gameform. The employment relationship is hence a game-form which parties decide to adhere to in order to save on communication (adaptation) costs. The agreement to play by the least costly adaptation-mechanism is upheld by the parties' concern for reputation in a repeated game. As in Coase (1937), the size of the firm is limited by administrative information overload.

C. The firm as a Governance Mechanism (Williamson)

In spite of the contributions mentioned above, Coase's idea that transaction costs can sometimes be diminished by interacting within firms rather than across markets, has been difficult to develop theoretically. In the terms of Oliver Williamson, Coase's basic story for long awaited its "operationalization". In particular, the nature and determinants of transaction costs were not, according to Williamson, adequately identified in Coase's work.

In a string of contributions, Williamson (e.g., 1971, 1975, 1985, 1996) has erected an impressive edifice on Coasian foundations.⁸ The behavioral starting points in Williamson's theorizing are, first, Herbert Simon's concept of bounded rationality, which produces contractual incompleteness and a need for adaptive, sequential decision-making, and, secondly, opportunism, defined as "self-interest seeking with guile", which has the implication that contractual agreements need various types of safeguards, such as "hostages" (e.g., the posting of a bond with the other party). Williamson call contractual arrangements and the safeguards they embody "governance structures", and the basic idea is to assign transactions to alternative governance structures on the basis of their transaction properties.

Given bounded rationality and uncertainty, these are determined by what has increasingly become the central character in Williamson's analysis, namely asset-specificity. Assets are highly specific when they have value within the context of a particular transaction but have relatively little value outside the transaction. This opens the door to opportunism. If contracts are incomplete due to bounded rationality, they must be renegotiated as uncertainty unfolds, and if a party to the contract (say, a supplier firm) has incurred sunk costs in developing specific assets (including human capital), that other party can

⁸ Thus, Williamson's work incorporates ideas from the so-called behavioural theory of the firm (Cyert and March 1963), the emphasis on operating with a broad notion of self interested behaviour (a broad utility function) in the managerialist school (Williamson 1963), ideas on contract law, and elements of information economics (Arrow 1969).

opportunistically appropriate an undue part of the investment's pay-off ("quasi-rents") by threatening to withdraw from the relationship.

According to Williamson, vertical integration, that is concentration of ownership of the (alienable) assets involved in the relation rather than having them separately owned, does away with these problems, because it removes the incentive for opportunism. Williamson's arguments for the relative benefits of vertical integration for some classes of transactions - those that are characterized by a high degree of asset-specificity - rely on organizational theory and law. Thus, he argues that internal organization is characterized by its own implicit contract law, what he calls "forbearance". Whereas divisions will not normally be granted standing for a court, corporate headquarters and headquarters function as the firm's "ultimate court of appeal". For example, Williamson points out that disputes which arise within the firm, for example, between different divisions, may be easier to resolve than disputes arising between firms which sometimes require the use of the court-system (Williamson 1996).

D. The firm as an Ownership Unit (Grossman, Hart and Moore)

Over the last decade Oliver Hart, John Moore and others have developed an "incomplete contracts" or "property rights" theory of the firm (Grossman and Hart 1986; Hart and Moore 1990; Hart 1995) which draws on key elements in Williamson's work.⁹ As in Williamson's work, a central assumption is that because of transaction costs/bounded rationality, contracts must necessarily be incomplete in the sense that the allocation of control rights cannot specified for all future states of the world. The theory defines ownership as the possession of *residual* rights of control, that is, rights to control the uses of assets under contingencies that are not specified in the contract. As a result, the allocation of formal ownership rights will affect behavior and resource allocation. For example, agents will be less inclined to invest in specific assets if they do not own these and relevant complementary assets.

In the Grossman-Hart-Moore theory a firm is defined as a collection of jointly owned assets. The basic distinction between an independent contractor and an employee, that is to say, between an inter-firm and an intra-firm transaction, turns on who owns the physical assets which the person utilizes in his work. An independent contractor owns his tools etc., while an employee doesn't. Owning an asset is important if one undertakes a non-contractible investment which is specific to the asset; if one does not own the asset, one is subject to the hold-up threat by the owner. Hence, according to this theory, the person who

⁹ Kreps (1996) argues that the Grossman-Hart-Moore theory misses some of the richness of Williamson's work. The ex-ante perspective on the hold-up problem was first developed by Klein, Crawford and Alchian (1978?).

is to make the most important (non-contractible) asset-specific investment should own the asset. Hiring an employee (making the product within the firm) means hiring someone who stands the risk of being held-up by the firm since the manager can threaten to “fire him” (i.e. exclude him from the use of the assets of the firm). Hiring an independent contractor (buying in the market) means granting him some power to hold up the firm by with his assets quitting. The optimal size of the firm balances these two opposing forces.

The Grossman-Hart-Moore theory was the first to provide a formal model that could explain both the advantages and the disadvantages of firm-organization, that is, a theory that could convincingly establish the boundaries of the firm.

E. Implicit Contracts

When it is difficult to write complete state-contingent contracts, for example, when certain variables are either ex-ante unspecifiable or ex-post unverifiable, people often rely on “unwritten codes of conduct”, that is, on implicit contracts. These may be self-enforcing, in the sense that each party lives up to the other party’s (reasonable) expectations from a fear of retaliation and breakdown of cooperation. The basic idea in the implicit-contract theory of the firm is that implicit contracts may function differently within firms than between firms; a person is hired as an employee rather than as an independent contractor when coordinating with him requires an implicit contract that is easier to implement within the firm than in the market-place.

In a recent paper, Baker, Gibbons and Murphy (1997) emphasize that implicit contracts occur both within firms (in the employment relationship) and between firms (“relational contracting”). The difference lies in the options which are open to the parties if the implicit contract breaks down. Contrary to an independent contractor, an employee cannot leave the relationship with the assets belonging to it. Specifically, in their model, independent contracting is defined by the feature that the supplier has the possibility to sell the finished product elsewhere, while firm-contracting is defined by the supplier (the employee) not owning the finished product and hence not having the option of leaving with the asset or product. The strength of the threat to leave the relationship determines the implementability of implicit contracts. For example, if the market for the good is volatile, relational contracts are vulnerable since the supplier is tempted to break out from the implicit contract when the market price is high. If the supplier is a division within the firm, this option does not exist and the implicit contract which holds the (internal transfer) price constant may therefore be self-enforcing. The implicit-contract theory is linked with Williamson’s idea that dispute resolution is easier to carry out within a firm than between firms: Mechanisms for dispute resolution can be seen as part of the system of self-enforcing implicit contracting within a firm.

G. The Firm as a Communication-hierarchy

Recent work on the “theory of communication in hierarchies” is based on the idea, well-known from organization theory (e.g., Simon and March 1958), that one important function of the firm is to adapt to and process new information. In most versions, it builds rather directly on the early work by Marschak and Radner (1972), a classic contribution on team-theory that was published at about the same time that work on incentive alignment began to take off. This book pointed to a completely different approach to economic organization, one in which incentive conflicts were disregarded, or at least assumed to have been solved, and where coordination and communication were highlighted instead. For some time enthusiasm over the new theories of the firm that all centered on incentive-conflicts swamped interest in team-theoretic issues, but recently interesting work that focuses on coordination in a team-theoretic framework has emerged (Arrow 1985; Aoki 1986; Crémer 1990; Radner 1992, 1996; Bolton and Dewatripont 1994, Casson 1997).

These writers view the firm as a communication network that is designed to minimize both the cost of processing new information and the costs of communicating this information among agents. Communication is costly because it takes time for agents to absorb new information sent by others, but this time may be reduced by specializing in the processing of particular types of information. In Bolton and Dewatripont’s (1994) model, for example, each agent handles a particular type of information, and the different types of information are aggregated through the communication network. When the benefits to specializing outweigh the costs of communication, teams (firm like organizations) arise.

It was mentioned earlier that team-theory has difficulty explaining the boundary of the firm. This applies also to the theory of the firm as a communication-hierarchy. It does not explain why communication hierarchies cannot exist between firms. However, if this can be given an explanation, that explanation together with the theory of the firm as a communication hierarchy may constitute a theory of the firm.

IV. Some Criticisms of the Modern Theory of the Firm

It is hard to deny that the modern theory of the firm represents one of the important instances of scientific advance in economics during the last two decades. However, criticisms of the modern theory of the firm have been raised. For example, Milgrom and Roberts (1988: 450), two of the most important representatives of modern formal work on the theory of the firm, made the following prediction ten years ago:

The incentive based transaction costs theory has been made to carry too much of the weight of explanation in the theory of organizations. We expect competing and complementary theories to emerge - theories that are founded on economizing on bounded rationality and that pay more attention to changing technology and to evolutionary considerations.

In the following, we organize our brief discussion around Milgrom and Roberts' points of critique.

To put in admittedly simplistic terms, almost all of the focus in the modern economics of organization has been on interaction characterized by incentive conflicts. Very much attention has been given to the hold-up problem and its ex-ante consequences. Coordination problems not involving incentive conflicts have been much less treated, although they are surely important in most firms. However, as already mentioned, one attempt to study the non-incentive aspects of economic organization is represented by team theory. This sort of work adds a sophisticated, formal treatment of some hitherto ill-treated phenomena to the modern economics of organization. The theory's focus on communication channels, on delays in information transmission, on costs of communicating, etc. makes some sense out of for example the economic function of corporate culture (Cremer 1990).¹⁰ With respect to the challenge from such allegedly "softer" issues, it is also worth noting that work on implicit contracts meets this challenge to some extent (see Baker, Murphy and Gibbons 1997; Segal 1996).

Bounded rationality still awaits its formalization, which may prove important in the development of the theory of the firm. The relevance of ideas on bounded rationality is illustrated by a fundamental problem in much of the modern economics of organization: How are efficient types of organization selected in reality? It is a basic postulate of the modern economics of organization that real-world agents will indeed structure their dealings in accordance with the theory's predictions (e.g., Williamson 1985). But how can this be justified? One idea has been to rely on (untheorized) selection forces that weed out ill-adapted types of economic organization. We now know that social selection forces cannot in general be relied upon to produce efficient results. The other main approach is to simply assert that agents can rationally calculate pay-offs associated with alternative types of economic organization (e.g., firms vs markets) and choose the efficient one (technically, agents can perform dynamic programming) (Kreps 1992). Now, this assertion may perhaps be hard to square with the basic assumption of incomplete contracts, which typically (but not necessarily) involves some notion of unforeseen contingencies. Such problems awaits the development of precise models of

¹⁰ Kreps (1990) is an attempt to conceptualize corporate culture in an incomplete contract setting, where corporate culture has the function of signaling trust-worthy behavior under unforeseen (and therefore non-contractible) contingencies.

bounded rationality.

A primary source of unforeseen contingencies in real world economies is changing technology. It has been a persistent critique that the modern economics of organization “neglects technology”. Relatedly, a number of critics (e.g., Demsetz 1991; Winter 1991) have argued that the differences in different firms’ productive capabilities have been too much suppressed in the modern economics of organization in favor of an overriding concern with incentives. While the firm cannot generally be described by a production function, neither should technology be ignored altogether. The critics emphasize that knowledge may be imperfect in the realm of production – that firms have “differential productive capabilities” – and that this may influence economic organization.

V. Concluding Comments

In this paper, we have surveyed a number of streams of research that together constitute the modern theory of the firm, or, the economics of organization. The tenor of the discussion has been optimistic, and although we have acknowledge the existence of rather deep problems (section IV), we believe that we are on the way to a coherent theory of the firm.

Thus, an argument may be made that at least some of the streams of research that we have discussed are strongly complementary. Together they provide a coherent picture of the advantages and disadvantages of firm (or firm-like) organization.

The perhaps dominant basic perspective on the firm in today’s economics is that of an ownership unit under incomplete contracts, as it has emerged from the work of Williamson and Hart. The complementary nature of different attributes of firm-organization of the firm has been pointed out by Holmström and Milgrom (1994). And given incomplete contracts, Baker and Murphy (1997) have explained how implicit contracting under certain conditions is more likely to emerge within than between firms. Moreover, these contributions blend ideas from both the principal-agent and the incomplete contracting modeling approaches.

Therefore, one may argue that a synthesis is gradually emerging. However, as we indicated, even this synthesis is not complete: Critics argue that the field still needs to provide more room for issues relating to bounded rationality, coordination that is not related to incentive conflicts, cognition, embeddedness and differential capabilities. As it stands, the formal theory of the firm still differs from casual observation of real business firm as well as from the richness of Williamson’s comparative institutional analysis (Williamson 1975,

1985, 1996). This present lack of realism should be seen in a positive way, namely as an invitation to further research.

Bibliography

- Alchian, Armen A. 1965. "Some Economics of Property Rights," in idem. 1977. *Economic Forces at Work*, Indianapolis: Liberty Press.
- Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization," *American Economic Review* **62**: 772-795.
- Aoki, Masahiko. 1986. "Horizontal vs Vertical Information Structure of the Firm," *American Economic Review* **76**: 971-983.
- Arrow, Kenneth J. 1969. "The Organization of Economic Activity," in idem. 1984, *The Economics of Information*, Oxford: Basil Blackwell.
- Arrow, Kenneth J. 1974. *The Limits of Organization*, New York: W.W. Norton & Co.
- Arrow, Kenneth J. 1985a. "The Economics of Agency," in John W. Pratt and Richard J. Zeckhauser, *Principals and Agents: The Structure of Business*, Boston: Harvard Business School Press.
- Arrow, Kenneth J. 1985b. "The Informational Structure of the Firm," *American Economic Review* **75**: 303-307.
- Arrow, Kenneth J. 1987. "Reflections on the Essays," in George Feiwel, ed. *Arrow and the Foundations of Economic Policy*, New York: New York University Press.
- Baker, George, Robert Gibbons and Kevin J. Murphy. 1997. "Implicit Contracts and the Theory of the Firm," *Working Paper*.
- Baumol, William J. 1982. "Contestable markets: An Uprising in the Theory of Industry Structure," *American Economic Review* **72**: 1-15.
- Baumol, William J., John C. Panzar, and Robert D. Willig. 1982. *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich.
- Barzel, Yoram. 1997. *Economic Analysis of Property Rights*. 2nd ed. Cambridge: Cambridge University Press.
- Berle, Adolph A. and Gardiner C. Means. 1932. *The Modern Corporation and Private Property*. New York: Macmillan.
- Bolton, Patrick, and Mathias Dewatripont. 1994. "The Firm as a Communication Network," *Quarterly Journal of Economics* **115**: 809-839.
- Casson, Mark. 1997. *Information and Organization*, Oxford: Oxford University Press.
- Cheung, Stephen S.N. 1983. "The Contractual Nature of the Firm," *Journal of Law and Economics* **26**: 1-22.
- Coase, Ronald H. 1937. "The Nature of the Firm," *Economica* **4**: 386-405.
- Coase, Ronald H. 1960. "The Problem of Social Cost," *Journal of Law and Economics* **3**: 1-44.

- Coase, Ronald H. 1972. "Industrial Organization: A Proposal for Research," in Victor R. Fuchs, ed., *Policy Issues and Research Opportunities in Industrial Organization*. New York: NBER.
- Cyert, Richard and James G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs: Prentice-Hall.
- Debreu, Gerard. 1959. *Theory of Value*, New York: Wiley.
- Demsetz, Harold. 1983. "The Structure of Ownership and The Theory of the Firm", *Journal of Law and Economics*, 26, p. 375-390.
- Demsetz, Harold. 1991. "The Theory of the Firm Revisited," in Oliver E. Williamson and Sidney G. Winter, eds. *The Nature of the Firm*. Oxford: Blackwell.
- Dixit, Avinash K. and Robert S. Pindyck. 1994. *Investment Under Uncertainty*. Princeton: Princeton University Press.
- Dosi, Giovanni and Luigi Marengo. 1994. "Some Elements of an Evolutionary Theory of Organizational Competences," in Richard W. Englander. *Evolutionary Concepts in Contemporary Economics*. Ann Arbor: University of Michigan Press.
- Fama, Eugene. 1980. "Agency Problems and the Theory of the Firm," *Journal of Political Economy* **88**: 288-307.
- Fama, Eugene and Michael Jensen. 1983. "Agency Problems and Residual Claims," *Journal of Law and Economics* **26**: 327-49.
- Fama, Eugene and Michael Jensen. 1985. "Organizational Forms and Investment Decisions," *Journal of Financial Economics* **14**: 101-119
- Guesnerie, Roger. 1994. "The Arrow-Debreu Paradigm Faced With Modern Theories of Contracting," in Lars Werin and Hans Wijkander. 1992. *Contract Economics*, Oxford: Blackwell.
- Grossman, Sanford and Hart, Oliver. The Costs and Benefits of Ownership: A Theory of vertical and Lateral Integration. *Journal of Political Economy* **94**:691-719.
- Hansmann, Henry. 1988. "Ownership of the Firm," *Journal of Law, Economics, and Organization* **4**: 267-305.
- Hansmann, Henry. 1996. *The Ownership of Enterprise*, Cambridge, MA: The Belknap Press of Harvard University Press.
- Hart, Oliver. 1983. "The Market Mechanism as an Incentive Scheme," *Bell Journal of Economics* **74**.
- Hart, Oliver. 1990. "Is "Bounded Rationality" an Important Element of a Theory of Institutions?," *Journal of Institutional and Theoretical Economics* **16**: 696-702.
- Hart, Oliver and Bengt Holmström. 1987. "The Theory of Contracts," in Truman F. Bewley. 1987. *Advances in Economic Theory. Fifth World Congress*. Cambridge: Cambridge University Press.

- Hart, Oliver and John Moore. 1990. "Property Rights and the Nature of the Firm", *Journal of Political Economy* **98**: 1119-1158.
- Hart, Oliver. 1995. *Firms, Contracts and Financial Structure*, Oxford: Clarendon Press.
- Hayek, Friedrich A. von. 1948. *Individualism and Economic Order*, Chicago: University of Chicago Press.
- Holmström, Bengt. 1979. "Moral hazard and Observability," *Bell Journal of Economics* **10**:
- Holmström, Bengt. 1982. "Moral Hazard in Teams," *The Bell Journal of Economics* **13**: 324-340.
- Holmström, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization* **7**: 24-52.
- Holmström, Bengt and Paul Milgrom. 1994. "The Firm as an Incentive System," *American Economic Review* **84**: 972-991.
- Hurwicz, Leonid. 1972. "On Informationally Decentralized Systems," in Charles B. McGuire and Roy Radner. *Decision and Organization*. Amsterdam: North-Holland.
- Jensen, Michael. 1986. "Agency Costs of Free Cash Flow, Corporate Finance and Takeovers," *American Economic Review, Papers and Proceedings* **76**
- Jensen, Michael and William Meckling. 1976. Theory of the Firm: Managerial Behaviour, Agency Costs, and Ownership Structure. *Journal of Financial Economics* **3**: 305-360.
- Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics* **21**: 297-326.
- Knight, Frank H. 1921 [1965]. *Risk, Uncertainty, and Profit*. New York: Augustus M. Kelley.
- Kreps, David M. 1992. "Static Choice in the Presence of Unforeseen Contingencies," in Partha Dasgupta, David Gale, Oliver Hart, and Eric Maskin, eds. *Economic Analysis of Markets and Games: Essays in Honour of Frank Hahn*. Cambridge: MIT Press.
- Kreps, David M. 1990. "Corporate Culture and Economic Theory", in James E. Alt and Kenneth Shepsle, eds. 1990. *Perspectives on Positive Political Economy*, Cambridge: Cambridge University Press.
- Kreps, David M. 1996. "Markets and Hierarchies and (Mathematical) Economic Theory," *Industrial and Corporate Change* **5**: 561-595.
- Langlois, Richard N. 1998. "Capabilities and the Theory of the Firm," in Nicolai J. Foss and Brian J. Loasby, eds. 1998. *Capabilities, Coordination, and Economic Organisation: Essays in Honour of George B Richardson*. London: Routledge.

- Lewis, Tracy, and David Sappington. 1991. "Technological Change and the Boundaries of the Firm," *American Economic Review* **81**: 887-900.
- Loasby, Brian J. 1996. "The Organisation of Industry", in Nicolai J Foss and Christian Knudsen, eds. *Towards a Competence Theory of the Firm*. London: Routledge.
- Manne, Henry. 1965. "Mergers and the Markets for Corporate Control," *Journal of Political Economy* **73**: 110-120.
- Marris, Robin. 1964. *The Economic Theory of Managerial Capitalism*. Glencoe, IL: Free Press of Glencoe.
- Marris, Robin and Dennis C. Mueller. 1980. "The Corporation, Competition and the Invisible Hand," *Journal of Economic Literature* **18**: 32-63.
- Marschak, Jacob and Roy Radner. 1972. *The Theory of Teams*, New Haven: Yale University Press.
- Maskin, Eric and Jean Tirole. 1996. "Unforeseen Contingencies, Property Rights, and Incomplete Contracts," *unpublished manuscript*.
- Milgrom, Paul J. and John D. Roberts. 1988. "Economic Theories of the Firm: Past, Present, and Future," *Canadian Journal of Economics* **21**: 444-458.
- Milgrom, Paul, and John D. Roberts. 1992. *Economics, Organization, and Management*. New York: Prentice Hall.
- Miller, Gary. 1992. *Managerial Dilemmas*. Cambridge: Cambridge University Press.
- Moore, John. 1992. "The Firm as a Collection of Assets," *European Economic Review* **36**: 493-507.
- Panzar, John C. 1989. "Technological Determinants of Firm and Industry Structure," in Richard Schmalensee and Robert D. Willig eds. 1989. *Handbook of Industrial Organization, Volume I*. North-Holland: Elsevier Science Publishers.
- Radner, Roy. 1996. "Bounded Rationality, Indeterminacy, and the Theory of the Firm," *Economic Journal* **106**: 1360-1373.
- Richardson, G. B. 1972. "The Organisation of Industry," *Economic Journal* **82**: 883-96.
- Ross, S. A. The Economic Theory of Agency: The Principal's Problem. *American Economic Review* **63**:134-139.
- Segal, Ilya R. 1996. "Modeling the Managerial Task," *mimeo*, Department of Economics, University of California at Berkeley.
- Schmalensee, Richard and Robert D. Willig, eds. 1989. *Handbook of Industrial Organization, Volume I*. North-Holland: Elsevier Science Publishers, B. V.
- Simon, Herbert A. 1951. "A Formal Theory of the Employment Relationship," *Econometrica* **19**: 293-305.
- Simon, Herbert A. and James G. March. 1958. *Organizations*. New York: Wiley.

- Teece, David J., Gary Pisano, and Amy Shuen. 1997. "Strategic Management and Dynamic Capabilities," in Nicolai J Foss. *Resources, Firms, and Strategies: A Reader in the Resource-Based Perspective*. Oxford: Oxford University Press.
- Tirole, Jean. 1994. "Incomplete Contracts: Where Do We Stand?", *unpublished manuscript*.
- Wernerfelt, Birger. 1997. "On the Nature and Scope of the Firm: An Adjustment-Cost Story," *Journal of Business* **70**: 489-514.
- Williamson, Oliver E. 1964. *The Economics of Discretionary Behavior*. Englewood Cliffs: Prentice-Hall.
- Williamson, Oliver E. 1971. "The Vertical Integration of Production: Market Failure Considerations'," *American Economic Review* **61**: 112-123
- Williamson, Oliver E. 1975. *Markets and Hierarchies*. New York: Free Press.
- Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.
- Williamson, Oliver E. 1996. *The Mechanisms of Governance*. Oxford: Oxford University Press.
- Winter, Sidney G. 1988. "On Coase, Competence, and the Corporation," in Oliver E. Williamson and Sidney G. Winter, eds. *The Nature of the Firm*. Oxford: Blackwell.