Abstract

This chapter is a survey of modern theories of the firm. We categorize these as belonging either to the principal-agent or the incomplete contracting approach. In the former category fall, for example, the Alchian and Demsetz moral hazard in teams theory as well as Holmstrom and Milgrom’s theory of the firm as an incentive system. Belonging to the incomplete contracting branch are theories that stress the importance of the employment relationship (for example, Coase and Simon) as an adaptation mechanism, theories that stress the importance of ownership of assets for affecting incentives when contracts must be renegotiated (Williamson, Grossman and Hart, Hart and Moore), and some recent work on implicit contracts (Baker, Gibbons and Murphy). We argue that these different perspectives on the firm should be viewed as complementary rather than as mutually exclusive and that a synthesis seems to be emerging.

JEL classification: K22, L22

Keywords: Principal-Agent Problems, Incomplete Contracts, Moral Hazard, Employment Relationship, Firm Ownership, Renegotiation

1. 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 1970s 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 (for example, 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 the 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 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).

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 assumptions, the model describes how markets 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 (for example, 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, and so on 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
The Theory of the Firm

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, p. 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, 1964; 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, 1958; 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 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 the 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). (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.)
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 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.

Accordingly, theories of the firm may roughly be classified in a rough way into two categories:

(a) *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 and Moore; Coase; and Simon, as well as implicit contract-theories, and the theory of communication in hierarchies, belong to this category.

(b) *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 did not elaborate on. Clearly, these perspectives are complementary, and should be integrated. There are signs that this integration process is slowly beginning (for example, Holmström and Milgrom, 1994). However, in this paper we stick to the above dichotomization, and after
The Theory of the Firm

surveying theories in each category we end up discussing prospects and criticisms. (For more on the disagreement between the two modeling strategies, see Tirole, 1994.)

2. Complete Contracting and Economic Organization: Principal-Agent Theories

2.1 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, manageralist 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 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 1970s formal agency theory blossomed (Ross, 1973 is an early contribution). In its paradigmatic version, the theory deals with a relationship between a principal (for example, the owner) and an agent (for example, 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 1970s 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, 1985a). The principal may, of course, infer something by observing output (for example, 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 (that is, induced to behave efficiently) by takeover threats (Manne, 1965; Jensen, 1986), product market competition
(Hart, 1983), debt pressure (Jensen, 1986) and competitive managerial labor markets (Fama, 1980).

2.2 The Nexus of Contracts View
Alchian and Demsetz (1972), Jensen and Meckling (1976), Barzel (1997) and, most forcefully, Fama (1980) and Cheung (1983) 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 (for example, limited liability, the right to deduct input purchases from tax statements, infinite lifetime, and so on), 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' formulation, the authority relation associated with the employment relationship is, for example, in no way the defining characteristic of the 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 counterpart. 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. 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 ‘underdetermines’ 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.

2.3 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 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 and 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 and Demsetz’s 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 don’t the employees 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).

2.4 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 a 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’ (Holmström and Milgrom, 1991) 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 is then likely to care too little for the assets. Likewise, low-powered incentives 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 the 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.

3. Incomplete Contracting Theories

3.1 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 belongs here, although their focus is somewhat different: they focus on the bargaining power that asset-ownership may provide in those situations in which contingencies arise that are not covered by the contract.

3.2 The Authority View: The Firm as an Employment Relation

The view that the employment contract is the characteristic feature 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 a 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 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 (game forms) according to which players adapt to changes in the environment and communicate about these changes. His conjecture here is that
different game forms 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 game form requires less bargaining over prices than the market game form. 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.

3.3 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 (for example, 1971, 1975, 1985, 1996) has erected an impressive edifice on Coasian foundations. Thus, Williamson’s work incorporates ideas from the so-called behavioral theory of the firm (Cyert and March, 1963), the emphasis on operating with a broad notion of self-interested behavior (a broad utility function) in the managerialist school (Williamson, 1964), ideas on contract law, and elements of information economics (Arrow, 1969). 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’ (for example, the posting of a bond with the other party). Williamson calls 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 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 to 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).

3.4 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 elements in Williamson’s work. 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 be 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, and so on, while an employee does not. 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 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’ (that is, 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 quitting with his assets. 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.

3.5 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 the 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.
3.6 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 (for example, 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 theory issues, but recently interesting work that focuses on coordination in a team theory framework has emerged (Arrow, 1985b; Aoki, 1986; Crémér, 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 time consumption 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.

4. 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, p. 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 it 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 \textit{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 (see also Crémer, 1990) (Kreps, 1990, makes an attempt to conceptualize corporate culture in an incomplete contract setting, where corporate culture has the function of signaling trustworthy behavior under unforeseen - and therefore non-contractible - contingencies). 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 (for example, 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 (for example, firms v. 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 await the development of precise models of 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’. Moreover, a number of critics (for example, Demsetz, 1991; Winter, 1988) have argued that the differences in different firms’ productive capabilities have been 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.

5. Concluding Comments

In this chapter, 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 acknowledged the existence of rather deep problems (Section 4), 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, Gibbons 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 the casual observation of real business firm as well as from the richness of Williamson’s comparative institutional analysis (Williamson, 1975,
This present lack of realism should be seen in a positive way, namely as an invitation to further research.

Acknowledgements

We are grateful for helpful comments by two anonymous reviewers.

Bibliography on The Theory of the Firm (5610)


Jovanovic, Aleksandra (1993), Participativna ekonomija: Uticaj tipa svojine na ponasanje (efikasnost) preduzeca (Participatory Economics: Influence of a Property Form on Behaviour (Efficacy) of a Firm), Belgrade, Pravni Fakultet.


Knight, Frank H. (1921), Risk, Uncertainty, and Profit, New York, Kelley.


Marchak, Jacob and Radner, Roy (1972), The Theory of Teams, New Haven, CT, Yale University Press.

Maskin, Eric and Tirole, Jean (1996), Unforeseen Contingencies, Property Rights, and Incomplete Contracts, mimeo, Boston, MA, Harvard University.


Miller, Gary J. (1992), Managerial Dilemmas, Cambridge, Cambridge University Press.


The Theory of the Firm


Segal, Ilya R. (1996), Modeling the Managerial Task, California, Department of Economics, Berkeley.


Tirole, Jean (1994), Incomplete Contracts: Where Do We Stand?, unpublished manuscript.

