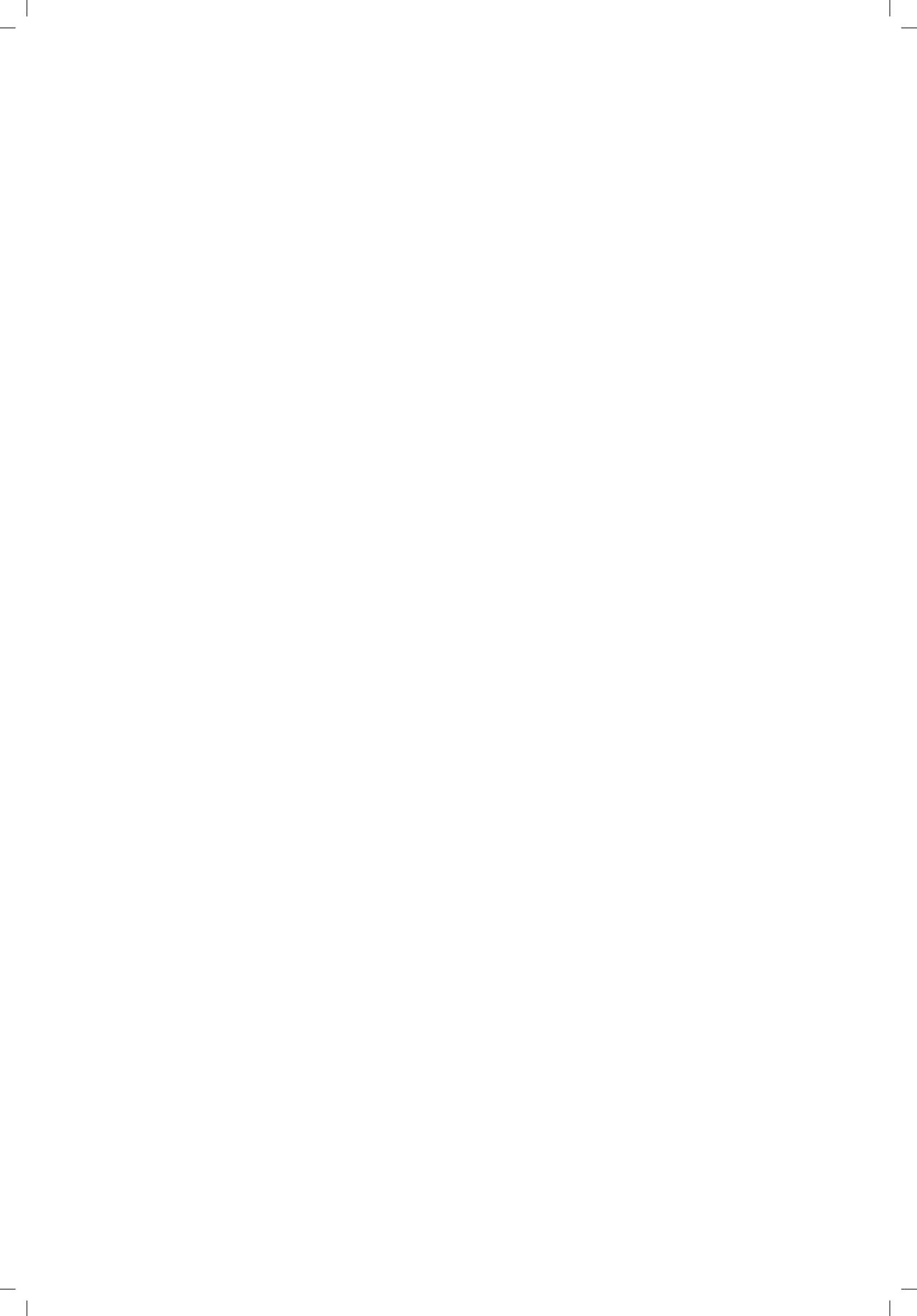


**CONTEMPORARY ECONOMIC
THEORIES OF THE FIRM**



Jacek Prokop

CONTEMPORARY ECONOMIC THEORIES OF THE FIRM

Peer review: Michał Ramsza

This textbook was prepared for the purposes of International Doctoral Programme in Management and Economics organized within the Collegium of World Economy at Warsaw School of Economics.

The textbook is co-financed by the European Union from the European Social Fund.

Contents

1. The Background of the Economic Theory of the Firm	7
1.1. The importance of the theory of the firm	7
1.2. Methodological innovations	8
1.3. Definitions of a firm	11
1.4. The firm as a static synergy	11
1.5. Applicability of the theory of the firm	12
2. The Firm as a Long-Run Relationship	15
2.1. Idiosyncratic investment and asset specificity	15
2.2. Bilateral monopoly pricing and the ex post volume of trade	16
2.3. Specific Investment and the hold-up problem	18
2.4. Limitations of long-run relationships	21
2.5. Applications of contractual view of the firm	21
3. Transaction Cost Economics and the Firm	25
3.1. The sources of incomplete contracts	25
3.2. Arbitration	26
3.3. Authority and its scope	27
3.4. Substitutes for contracts	29
3.5. The property rights approach	30
3.6. Other formal approaches to the theory of the firm	30
4. Principle-Agent Analysis	35
4.1. The basic incentive problem	35
4.2. An example of an optimal incentive scheme	38
4.3. Observability, verifiability, authority	40
4.4. Limits to discretion – yardstick competition	41
4.5. The dynamics of managerial incentives	42
5. Empirical Research in the Theory of the Firm	45
5.1. Evidence on boundaries of the firm	45
5.2. Studies on make-or-buy decision	47
5.3. Empirical analysis of contracting	49
5.4. Problems with empirical testing of the theory of the firm	50
5.5. Opportunities for the future	53
Bibliography	55



The Background of the Economic Theory of the Firm

1.1. The importance of the theory of the firm

Firms are the key elements of any economy, thus a better understanding of what they do and why they exist is crucial. The theory of the firm addresses the existence, the boundaries, and the internal organization of the firm. It has become a favourite preoccupation of modern economists. R. Coase's "The Nature of the Firm" (1937) is conventionally regarded as the founding paper in the theory of the firm. However, Frank Knight in "Risk, Uncertainty, and Profit" (1921) was probably the first economist to explicitly argue that the economic principles can explain the different forms of business organizations found in the real world. He pointed at explanations involving morally hazardous behaviour (Barzel, 1987), non-contractibility of entrepreneurial judgment (Langlois and Cosgel, 1993), and the optimal allocation of risk (Kihlström and Laffont, 1979). The latter explanation was a critical point of departure in the Coase's paper.

The seminal analysis by the 1991 Nobel Prize winner R. Coase was initially neglected for more than three decades. For example, T. Marschak (1965) writing in "The Handbook of Organization" about "Economic Theories of Organization" does not even mention Coase. Serious work on the theory of the firm had begun to take off with Oliver Williamson's 1971 paper on "The Vertical Integration of Production: Market Failure Considerations," and A. Alchian and H. Demsetz' "Production, Information Costs, and Economic Organization" (1972). These two contributions already marked the beginning of a branching in the Coasian theory of the firm. More explicitly, O. Hart in "An Economist's Perspective on the

Theory of the Firm" (1989) introduced a distinction between "transaction cost economics" attributing it to Williamson's paper, and "the firm as a nexus of contracts view" attributing it to the Alchian and Demsetz paper.

Other approaches also took off in the beginning of the 1970s, primarily the team-theoretic approach of J. Marschak and R. Radner (1972) and the evolutionary theory of the firm by R. Nelson and S. Winter (1973). Contract theory also began around that time with the first contributions to the formal principal-agent theory; among the early papers we could name S. Ross' "The Economic Theory of Agency: The Principal's Problem" (1973). Formal advancements in the agency problem were surveyed in B. Holmström and J. Tirole's "The Theory of the Firm" (1989), and an overview of incomplete contracts was provided in O. Hart and B. Holmström's "The Theory of Contracts" (1987).

Perhaps the best known author on the modern theory of the firm is the 2009 Nobel Prize winner O. Williamson. He has not only popularized the thoughts of Coase, but also made his own contributions to the development of the theory of the firm.

The importance of a theory of the firm stems from its key role in providing the most relevant, even foundational inputs, into many business disciplines: strategic management, organizational theory, parts of marketing, international business, etc. Moreover, the firm (organization) is a fundamental unit of analysis in most of management, and many dependent and independent variables in management research are firm-level. Hence, the general importance of a theory of the firm is unquestionable.

1.2. Methodological innovations

The theory of the firm is not homogenous, although recent formal work may be characterized by a convergence towards insights and analytical approaches developed within the framework of contract economics. One common theme that has often been highlighted is that all work in this area begins from a view of human nature that goes beyond the conventional maximizing assumption. Usually it is introduced through assumptions about the potentially opportunistic or morally hazardous behaviour of contracting partners. Since it may not always be perfectly clear when a particular behaviour is "opportunistic" and when it is "non-opportunistic," it is better to say that all problems of economic organization are resulting from incentive conflicts.

The early approach to the concept of a firm in economics (by the end of the 1950s) could be best summarized as the concept of a producer, i.e., an economic agent whose role is to choose (and carry out) a production plan. The producer (with a plan for the whole future) was equivalent to a specification of the quantities of all his inputs and all his outputs; outputs

are represented by positive number, inputs by negative numbers. With this convention a production plan of firm j could be represented by a point, y_j , in the commodity space.

The main assumptions of the early economic cannon could be highlighted as follows:

- purposeful behaviour (often “greed”),
- forward-looking agents,
- full information,
- equilibrium,
- competitive markets.

The modern economic approach to the theory of the firm is still based on:

- purposeful behaviour,
- forward-looking agents,

but it is modified by:

- asymmetric and imperfect information,
- great proliferation of equilibrium concepts,
- small-numbers interaction.

This change in economic thinking has facilitated understanding economic organization. In the last three decades there was an explosion of work on the organization economics induced by the analysis of imperfect contracts, repeated games, incentives and asymmetric information and the small-numbers interaction.

The theory of the firm spans at least three disciplines: sociology, economics, and (social-) psychology (i.e., the dominant base disciplines in management). The work in this field seems to be quite fragmented, although there is certain “integrative” ambition among researchers – particularly in Williamson’s type of transaction cost economics.

Mathematical modelling and formal deductive logic have been a part of economic theory for a long time, but their application by the main stream of the profession accelerated around a half-century ago. The allure of mathematical modelling and reasoning comes from the power it offers to define all terms precisely and unambiguously. That in turn allows formulating precise conclusions and to extend our analyses and unify them. Another point is that economic research (relatively) lacks controlled data. The logical coherence is the second-best substitute for data when one is trying to understand the real world, but it gives some discipline and sharpness to theorizing.

Allied with the rise of formal mathematical models was the rise of statistical methodology in economics. For many decades, most of empirical work was substantially more difficult than theory-building. The reason was a great difficulty in gathering data, but also the relatively high costs of processing them.

The rise of mathematical modelling and statistical methods in economics meant a change in the way economists reasoned and communicated. Traditionally, many subjects within economics, such as international trade,

development, economic history, labour markets, public finance, maintained significant intellectual autonomy. The material a student would learn in, for example, development economics would not resemble that closely what was taught in public finance. Earlier emphasis was placed on typologies and on the institutions that pertained to the subject in question. There were a number of regional dialects of economics, dialects that were close to being distinct languages. As economics was mathematicised the dialect of international trade today is very close to the dialects of industrial organization, development, and labour economics. Nowadays the languages used by different branches of economics sound a lot more like branches of a single methodological, rather than topical, tongue.

This unification of dialects had its advantages, because insights obtained in one context can be transferred to another; that promotes the educational process. Nonetheless, there was a real loss of certain topical concerns. Fortunately, after a narrowing of topical concerns as the language was unified, we observe a widening of concerns as the language develops. That is particularly visible in the subject of institutional economics. It is also worth underscoring that, recently, interest in comparing broadly different systems such as socialism and capitalism, have become much more microeconomic. There is a significant increase of interest in nonmarket economic exchange, i.e., exchange within firms and other nonmarket institutions.

Together with these environmental influences we can name some important methodological innovations. The ability to process data has been increased enormously. Statistical tests are no longer limited to a small number of theoretically manageable parametric specifications. Monte Carlo methods have dramatically increased what can be tested and how.

The amount of data available to researchers is huge due to the revolution in the information technology. Even the data at a microscopic level of aggregation could be obtained. The general availability of transaction-by-transaction trading data and details on purchasing activities of individual households have virtually revolutionised financial market theory and accounting.

Where data do not exist, economists are increasingly creating them using experiments. The growth of computing power has also made it possible to solve large and complex problems and to investigate the properties of decision rules through simulation.

The innovation that has had the greatest impact on economic thought has been the application of tools developed by information economics and noncooperative game theory. It has allowed tackling three important groups of issues: small numbers interactions, privately held information, and dynamics. While the mathematization of economics was certainly a revolution, the subsequent expansion until today has been evolutionary. Now, the economists speak a more or less common dialect and topical fads are common.

1.3. Definitions of a firm

There could be various ways to classify the definitions of a firm, and it is hard to find a fully coherent approach. One of the early classifications was offered by J. Tirole (1997). He has named four not necessarily disjoint views of the firm.

The first approach is to perceive a firm as a loophole for the exercise of monopoly power. The legal environment prompts internalization of certain market practices to circumvent the law. A firm may want to avoid some regulations, e.g., intermediate price controls, or sales tax. Another reason could be the implementation of price discrimination.¹

The remaining three definitions are based on:

- synergy (technological view),
- contractual view, and
- incomplete-contracting (transaction-cost) view.

1.4. The firm as a static synergy

An old wisdom in industrial organization is that the size and the number of firms in an industry are related to the degree of returns to scale or of scope. It has been well documented that higher level of production permit the use of more efficient techniques.² Economies of scale related to the volume of a single product are called product-specific economies. It could include the so-called economies of massed reserves. A plant with a large number of machines can sustain a flow of output proportionally higher than the one with a small number. Similarly, a firm serving several markets with variable demands faces less uncertainty than separate firms serving these markets independently. Alternatively, the gathering of activities allows avoiding duplication of fixed costs, or at least it reduces these costs on the average.

However, returns to scale have their limits. Machines or functional divisions related to different production units can be advantageously pulled together only if they are not employed to their capacities. Furthermore, it is sometimes argued that there can be unique factors, such as managerial talent, that cannot be duplicated as the firm expands.

The formalization of returns to scale and scope is presented below. Initially, consider the single product firm. Let $C(q)$ denote a firm's total cost of producing q units of output; assume that the cost function is twice differentiable. Marginal costs are strictly decreasing if $C''(q) < 0$ for all q . Average costs are strictly decreasing if for all q_1 and q_2 such that $0 < q_1 < q_2$, it holds that:

¹ See Tirole (1997, pp. 17–18).

² See Scherer (1980, pp. 81–84).

$$\frac{C(q_2)}{q_2} < \frac{C(q_1)}{q_1}$$

The cost function $C(q)$ is called strictly subadditive if, for any n -tuple of outputs q_1, \dots, q_n it holds that:

$$\sum_{i=1}^n C(q_i) > C\left(\sum_{i=1}^n q_i\right)$$

Subadditivity thus means that it costs less to produce the various outputs together than to produce them separately.

For a multiproduct firm, subadditivity generalises in a natural way. Suppose that \mathbf{q} is now a production vector, i.e. $\mathbf{q} = (q_1, \dots, q_m)$ for the m outputs. Let $\mathbf{q}^1, \dots, \mathbf{q}^n$ denote n such vectors; superscripts denote production plans, and subscripts index goods. The cost function $C(q)$ is strictly subadditive if, for all \mathbf{q} such that $\sum_i q^i \geq 0$ it holds that:

$$\sum_{i=1}^n C(q^i) > C\left(\sum_{i=1}^n q^i\right)$$

This definition gives one formulation of economies of scope. In the case of two goods, for a strictly subadditive cost function, it holds:

$$C(q_1, 0) + C(0, q_2) > C(q_1, q_2)$$

It costs less to produce the various outputs together than to produce them separately.

The technological view of the firm aims at determining the size of a firm. Economies of scale encourage the gathering of activities. And the limit to the expansion stems from the fact that average-cost curves may start rising at high enough output. There are at least two points of criticism to this approach. First, it is not clear why economies of scale should be exploited within the firm, and not through contracting between legally separate entities. Second, we should not take it for granted that average-cost curves rise at high output, since firms also have an option to split production across various plants.

1.5. Applicability of the theory of the firm

The theory of the firm has received a fair dose of criticism. For example, R. Coase (1991) strongly criticised the reliance in the dominant Williamson-Hart theory of the firm on asset specificity and opportunism. Coase argued that modern theories of the firm do not make an adequate provision for the activity of managing in terms of strategy building and leadership. A related critique has been put forward by Demsetz (1988). P. Milgrom and J. Roberts (1988) – two

of the most important formal contributors to the modern theory of the firm – made the following observation and prediction:

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 economising on bounded rationality and that pay more attention to changing technology and to evolutionary considerations.

The over-concentration on transaction costs to the near exclusion of production costs, which are taken as identical for the same task across firms when in reality firms are likely to have different production costs has been stressed by Demsetz (1988) and Langlois and Foss (1998).

Winter (1988) points out that the literature seems to pay not enough attention to path-dependency, which may lead the theorist to think of organizations as much more flexible as they really are.

There is lack of attention to determinants of economic organization that are not related to incentive conflicts, such as information processing and organizational “codes,” “languages,” etc.

There seems to be a tendency to take a too piecemeal approach, whereby transactions are examined one at a time so that the interaction effects among transactions are missed (Winter, 1988).

Another weakness of the literature on the theory of the firm is the neglect of organizational learning (Williamson, 1998) and other dynamic phenomena and issues, such as innovation (Williamson, 1985).

It is striking that most of the soft spots of the theory of the firm can be reduced or related to the bounded rationality issue. This is a powerful argument why economists must ultimately come to grips with bounded rationality, such as the collection of information, the violation of the axioms of basic decision theory, etc. (Conlisk, 1996). It would be nice to have something more definite than merely defining bounded rationality as virtually any deviation from maximizing behaviour. A first try in this direction may be found in some contributions to the information-processing view of the firm. Progress may arise from attempts to treat both information processing problems and incentive conflicts in the same formal model, even if these are likely to be quite complex.



The Firm as a Long-Run Relationship

In the previous chapter we have seen some reasons why initially independent units might want to merge or coordinate their activities through a static contract. It is however important to consider the long-run relationships between different units. In particular, it is important to justify why the rules that govern trade tomorrow ought to be determined today whenever this is feasible.

2.1. Idiosyncratic investment and asset specificity

Long-run relationships between business entities are often associated with either switching costs or specific investments. Switching costs have been extensively discussed by Williamson (1985). He described the hazards associated with a government's use of repeated bidding to allocate the rights to natural monopoly. The key issue among switching costs is the need of a newcomer to learn and the reluctance of the incumbent to share the information. The regulator does better by avoiding the use of repeated bidding to allocate the monopoly rights, and by the same argument it might not be cost-efficient for a buyer to repeatedly use the spot market to purchase certain goods or services from a supplier. Hence, the long-run agreements are preferred.

Switching costs constitute a case of idiosyncratic investment. Previous engagement in trade between two parties makes it more cost-efficient relative to trading with other business entities. In general, idiosyncratic investments are much more related to the potential for future trading rather than to current exchange. For example, when a supplier must install equipment that is specific to a buyer's particular order and would be practically useless for other purposes. In addition to this kind of specificity, it is possible to distinguish two further types: site specificity and specific investment in human capital.³

³ See Williamson (1975).

All types of specificity lead to the same conclusion: the parties that contract now are aware that there will be potential gains from trade between them in the future. It is crucial to exploit and divide these gains properly in order to induce the efficient amount of specific investments now.

The most important aspect of a specific investment is that there exists a pool of competitive buyers and suppliers *ex ante*, but the selected buyer and supplier end up in a bilateral monopoly *ex post*. Thus, a problem of possible inefficiencies in the trade *ex post*, as well as in the specific investment *ex ante* can arise.

2.2. Bilateral monopoly pricing and the *ex post* volume of trade

Consider the following model of interactions between a supplier and a buyer over two periods: $t = 1, 2$.



A supplier and a buyer may or may not contract at $t = 1$. At the beginning of period $t = 2$, the two parties learn how much they could gain by trading in period $t = 2$. Denoting the value of the good to the buyer by v , and the cost of the good to the supplier by c , the gains from trade to be split between the two parties will be $v - c$. If p is the trading price, the buyer obtains the surplus equal to $v - p$, and the seller receives the surplus of $p - c$. In case of no trade the surplus for each of the parties equals zero. Two main cases are possible.

Case 1: no contract at $t = 1$

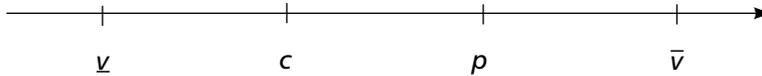
Should the parties decide not to sign any agreement at $t = 1$, there could be some bargaining taking place at $t = 2$ in order to determine whether to trade and at what terms.

If v and c are common knowledge before bargaining, we should expect the efficient amount of trade (i.e. trade occurs if $v \geq c$). Generally, bargaining under symmetric information is efficient – Coase theorem (Coase, 1960). However, asymmetric information may yield bargaining inefficiencies.

When the buyer's value and the supplier's cost are private information, the efficient volume of trade may not be reached because of the bilateral monopoly pricing problem. The inefficiency results from the fact that both parties would like to appropriate the gains from trade but due to asymmetric information face the risk of losing trade in the hunt for a bigger share of the pie in case of trade. They are too greedy.

Suppose that the cost c is common knowledge, the value v is known to the buyer only, and the supplier's beliefs about v are distributed on an interval $[\underline{v}, \bar{v}]$

($\underline{v} < c < \bar{v}$) with a cumulative probability distribution function $F(v)$, and density $f(v) > 0$ ($F(\underline{v}) = 0$ and $F(\bar{v}) = 1$).



Suppose that in period $t = 2$ the supplier makes a „take-it-or-leave-it“ price offer: p , i.e. the supplier has all the bargaining power. The buyer accepts the price offer only if $v \geq p$. Thus the probability of trading is $Pr(p \leq v) = 1 - F(p)$ and the supplier’s expected profit amounts to:

$$(p - c)[1 - F(p)]$$

Maximizing this profit with respect to (w.r.t.) p yields the first order condition (f.o.c.):

$$1 - F(p) - (p - c) f(p) = 0$$

Notice that for the above equation to hold, it must be that $p > c$, but the efficient volume of trade would arise if (if and only if) $p = c$. Thus, the volume of trade is suboptimal. This inefficiency results, because charging a price equal to cost yields no profit to the supplier, however, raising the price above cost yields a profit with some probability; the forgone benefits are costless because the initial price-cost margin is zero.

Case 2: contracting at $t = 1$

The resulting trade inefficiency induces the parties to contract *ex ante* to avoid or limit this inefficiency. It suffices to give the “informed party” – the buyer in this case – the right to choose the price (i.e. to reverse the bargaining power). Since c is common knowledge, the outcome will be (Pareto) efficient. The buyer’s monopoly price is equal to the price that leaves the seller indifferent between accepting and rejecting the offer, i.e. $p = c$. Thus, the buyer *ex post* expropriates the entire gains from trade. The buyer would pay *ex ante* a negotiated amount for the right to choose the price. More generally, the contract is supposed to help create the largest possible “pie”, but the division of this pie depends on *ex ante* relative bargaining powers of the parties.

Analogously, if the buyer’s value is common knowledge and the supplier’s cost is private information, it is efficient to assign the right to fix the price to the supplier. There is some similarity of such rights to a sequential authority relationship in which one party has the authority to choose the price and the other one can only decide whether to trade or not.

Summarizing, *ex post* bargaining may not lead to the efficient volume of trade. In order to improve the efficiency, some constraints on the second-period decision process must be contracted for. The key rule is that the power should go to the informed party.

2.3. Specific Investment and the hold-up problem

Suppose that at $t = 1$, a supplier invests in cost reduction and a buyer invests in value enhancement. These investments are assumed to be specific in the sense that they would affect neither c nor v if the parties were to engage in trade with any outside parties. Consider again two possible cases.

Case 1: no contract at $t = 1$

The parties decided not to sign any contract in the period $t = 1$, therefore they engage in bargaining at $t = 2$ over whether to trade and at what price. Clearly, the *ex post* volume of trade affects the size of investments.

Assume that at $t = 1$, the value of the good to the buyer is known to be $v = 3$ and the cost of the good to the supplier is high, i.e., $c > 3$. At $t = 1$ the supplier could invest by spending $I = 2$ to reduce the cost c to 0 at $t = 2$. If he does not invest, his cost remains high.

In the case of no contract between the parties at $t = 1$, the (*ex post*) bargaining at $t = 2$ results in the Nash bargaining solution (NBS), where any gains from trade are split evenly.

When $c = 0$, the gains from trade ($v - c$) amount to 3, and the NBS-price is 1.5 (thus each party's surplus from trade at $t = 2$ equals 1.5). Thus the net buyer's surplus equals to $0.5(v - c) = 1.5$. Since the supplier had to invest ($I = 2$) to reduce the cost c to zero, his net profit amounts to $0.5(v - c) - I = 1.5 - 2 = -0.5$. Hence he should not invest at $t = 1$. When c stays high ($c > 3$), there are no gains from trade; hence there will be no trade, and each parties' surplus will be zero.

However, it should be noticed that investment is socially desirable, because the total net gains $(v - c) - I = 3 - 2 = 1$ are positive.

A generalized model of specific investment:

Suppose that the cost is a deterministic function satisfying the following conditions: $c(I) \geq 0$, $c'(I) < 0$, $c''(I) > 0$, and assume that $v \geq c(0)$. The price *ex post* is determined by the Nash bargaining solution, i.e., $p(I) = [c(I) + v]/2$. The supplier's profit maximization problem is given as:

$$\max_I [p(I) - c(I) - I] = \max_I [v/2 - c(I)/2 - I]$$

The f.o.c. is given by: $-c'(I)/2 - 1 = 0$, i.e. the privately optimal investment I solves $c'(I) = -2$.

But for socially optimal investment (SO), we have:

$$\max_I [v - c(I) - I]$$

The f.o.c. for the above program is given as: $-c'(I) - 1 = 0$, i.e. the socially optimal investment I_{SO} solves $c'(I_{SO}) = -1$. Since $c'(I) > 0$, hence $I < I_{SO}$; thus I is suboptimal.

The problem identified here is that the investing party does not capture all the cost savings (increments in value) created by his investment. The other party can use the threat of not trading to appropriate some of these savings. It is the so called opportunistic behaviour or opportunism.⁴

Case 2: contracting at $t = 1$

Now assume that the seller and the buyer can write an *ex ante* contract specifying the process through which the amount of trade and the price are determined *ex post*. Investments are viewed as observable, but not verifiable, i.e. they cannot be specified *ex ante* in the contract. The same is assumed about the valuation and cost. Thus, everything is known to the involved parties, but they cannot be measured by a court.

The threat of termination of trade is often a way for some parties to appropriate a share of common gains, therefore it may be important to agree *ex ante* to introduce high penalties for breach of contract, e.g. trade at price p "no matter what" (equivalent to infinite penalties), or lose some collateral. A drawback is that high penalties force parties to trade even when there are no gains from trade. In the case of a non-negligible probability of no gains from trade, more flexible and sensitive mechanisms must be contracted for.

Consider a simple case. Suppose that v and c are random at $t = 1$, and become common knowledge between the two parties (non-verifiable) at the beginning of $t = 2$ (there is some chance that $v < c$). Supplier may invest I in order to reduce c , i.e. c depends stochastically on I . When supplier invests, the sequential authority mechanism by which he chooses the price and the buyer accepts or refuses to trade at this price yields the optimal levels of investment and trade, i.e., is efficient. The supplier offer price $p = v$ if $v \geq c$, and any price $p > v$ if $v < c$. In this case trade takes place if and only if $v \geq c$, and the entire *ex post* gains from trade go to the supplier. Clearly for the relationship between both parties to make sense the gains from trade could be divided *ex ante* through a lump-sum payment (e.g., supplier pays to the buyer an amount equal to 0.5, and by setting the price $p = v = 3$, he is left with the profit of $(p - c) - I - 0.5 = 3 - 2 - 0.5 = 0.5$. Thus the rule of thumb is that the party investing should have the authority over the price.

Consider now an example when the *ex ante* investment by the supplier affects the quality (the value) of the product to the buyer according to the following function:

⁴ The use of these terms in this context is attributed to Williamson (1975).

$$v(I) = 3I - \frac{1}{2}I^2$$

Again we assume that I is observable by the buyer but not verifiable by a court, so that it cannot be specified by a contract. The buyer's surplus in case of trade is $v(I) - p$, and the supplier's surplus is $p - c - I$ (c is a constant production cost $< 1/2$).

In order to determine the efficient amount of investment we maximize net total surplus:

$$\max_I [v(I) - I] = \max_I [3I - \frac{1}{2}I^2 - I] = \max_I [2I - \frac{1}{2}I^2]$$

The f.o.c. is $2 - I = 0$, thus the optimal investment amounts to $I^* = 2$.

When there is no contract and the two parties' *ex post* bargaining results in a NBS, i.e. $v(I) - I = p - c$, i.e.

$$p = \frac{[v(I) + c]}{2}$$

Ex ante, the supplier maximizes:

$$\frac{[v(I) + c]}{2} - c - I = \frac{3I - \frac{1}{2}I^2}{2} - c - I = \frac{I}{2} - \frac{c}{2} - \frac{I^2}{2}$$

The f.o.c. is $\frac{1}{2} - \frac{I}{2} = 0$, thus the optimal investment amounts to $I = 1$, which means that this investment will be suboptimal.

When the parties sign a contract specifying that the buyer has the right to buy the good at a given price p , the supplier either will not invest ($I = 0$), or will invest the minimum amount so that the buyer decides to buy: $v(I) = p$. The latter strategy yields profit equal to $p - c - v^{-1}(p)$. If $p = v(I) = 4$, the investment will be $I = 2$, and the profit amounts to $2 - c$. If $p < 4$, investment will be suboptimal or nonexistent.

Clearly, if the supplier has the right to sell at a given price p , the investment will be zero.

In the case when the supplier is given the right to choose the price *ex post*, the investment will be efficient.

Given the above discussion, it is worth noting that perfect auditing could contribute to expanding common knowledge about the value and the cost of traded goods, and this way prevent inefficiencies associated with incomplete information. Williamson (1975) has argued that integrated firms may be more susceptible to auditing than non-integrated ones, because it seems to be easier to audit an internal division than an outside contractor offering the same services; external auditors are regarded as suspicious and face collusive behaviour on the part of employees, who may hide the information.

2.4. Limitations of long-run relationships

There are at least three main limitations of long-run relationships.

The first limitation is clearly the existence of outside opportunities. Forcing the parties to stick to each other by the use of high penalties for breach may hurt them not only when there are no gains from trade, but also when outside opportunities are available to one of the parties or to both. Thus, the contract must be based on the optimal trade-off between flexibility and the prevention of opportunism.⁵

The second limitation of long-run relationship is that they promote collusion between the units' personnel. A long-run horizon gives them enough occasions to reciprocate favours and to become confident that the collusion is stable.⁶ For example, managers of a firm may in the long term identify with particular suppliers that could potentially create inefficiencies. That calls for the rotation of personnel within each unit, or the occasional switching of suppliers or buyers.

The third limitation may result from the fact that a short-term relationship is generally more advantageous to a party who expects to have good outside opportunities in the future.

2.5. Applications of contractual view of the firm

M. Jensen and W. Meckling (1976, p. 311) popularized "the nexus of contracts view" of the firm by:

The private enterprise or firm is simply one form of legal fiction which serves as a nexus for contracting relationships and which is also characterized by divisible residual claims on the assets and cash flows of the organization which can be sold without permission of the other contracting individuals.

Several other authors supported the above claim that the firm is nothing but a legal person or fiction, e.g. Alchian and Demsetz (1972), and S. Cheung (1983). Thus, to them it is essentially misleading to draw a hard line between firms and markets. Although firms constitute legal entities, and although this has important economic consequences (e.g., limited liability, tax deductions for inputs, etc.), firms are merely special kinds of market contracting. What may distinguish them relative to other market contracts stems primarily from the continuity of association among input owners.

Alchian and Demsetz (1972) argue that in consequence of this view the distinction between the authority-based and price-based modes of allocation, emphasized by Coase (1937), is superficial. In practice, there is no difference

⁵ See, for example, Tirole (1997, p. 27).

⁶ See Tirole (1986).

between “firing” one’s grocer and firing one’s employee, and what looks like a long, open-ended employment contract is in reality a cover for a continuous process of implicit negotiation between employers and employees.⁷

The reason that the firm is a special instance of market contracting has to do with the technology of team production, i.e. production with inseparable production functions. As a result, the marginal products are costly to measure, and a free rider problem arises (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 team-members, and the right to the residual income of the team. Thus, the firm is explained in terms of the reduction of post-contractual measurement cost.

From the Williamsonian theory of long-run relationships it follows that firms should write long and detailed contracts wherever feasible and not too costly, and that the incentive to do so increases with the lack of *ex post* outside opportunities and the specificity of investments. Joskow (1985 and 1987) has studied contracts between coal mines and electric utilities in the USA. The specific investments that could be expropriated for the coal mines (suppliers) are investments in mining capacity. In the case of electric utilities (buyers), the specific investments are investments in generating units and in boilers adapted to a particular type of coal.

Two polar geographical cases can be distinguished here. On the one hand, there is a large number of electric utilities and coal mines in the region, and the transportation infrastructure for coal mines (railroads, barges) are numerous and competitive. Furthermore, the coal produced by the various mines is quite homogeneous, so boiler design is basically irrelevant. In such a region, the hazard of expropriation of specific assets is small, so spot markets are relatively efficient. In the other case, there are few coal mines, transportation facilities are limited, and the quality of coal fluctuates. In this region, one would expect complex long-term contracts (or vertical integration) to facilitate trade.

Joskow documents that the regional differences in the USA offer an excellent illustration of Williamsonian predictions. In the East, there are many small mines due to the dominant underground mining characterized by almost none returns to scale. Transportation infrastructure is well developed, and the quality of coal is relatively homogeneous. In the West, there are only few large mines due to the prevalent surface mining with large economies of scale. There is also less competition in transportation than in the East, and the quality of coal fluctuates. Joskow shows that contracts in the West are of a much longer duration than in the East and that the spot market for coal is crucial in the East and practically nonexistent in the West.

Lack of competition and high transportation costs may induce an electric utility to locate near a coal mine, which creates site specificities. Joskow found that mine-mouth plants (i.e., electric-generating plants located near coal mines)

⁷ See Foss (2000).

tend to rely on long-term contracts (or even vertical integration); they sign twenty- to fifty-year agreements that prohibit price renegotiation, give a detailed description of quality of coal, and index costs and the prices of substitutes. This *ex ante* specification of terms of the contract may prevent specific investments from being expropriated.

Another practical application of the theory of long-run relationships is the Williamsonian warning against the hazards of repeated bidding to allocate the rights to natural monopoly. He argues that the incumbent's specific investments, in order to be transferred to an alternative supplier at a fair rate must be observable and verifiable, which is hard to achieve. Even harder to measure are specific investments in human capital of the incumbent's personnel. So a fair reimbursement of investment must be insured by bidding over the equipment. But the incumbent's transferable investments may not be observable by the other bidders. The use of repeated bidding is thus likely to be costly in industries with high specific investments, and more advantageous in industries in which only a small fraction of investments is sunk.



Transaction Cost Economics and the Firm

3.1. *The sources of incomplete contracts*

A contract is complete when all relevant decisions (regarding, for example, trade, transfer, etc.) depend on verifiable variables, including announcements by the parties concerning their valuation, costs etc. However, contracts are usually incomplete due to “transaction costs.” Coase (1937) and Williamson (1975) have distinguished between four types of transaction costs:

- two types of transaction costs at the contracting date:
 - some contingences to be faced by the parties may not be foreseeable at the date of contracting,
 - even if it is possible to foresee everything, there may be too many contingences to include in the contract;⁸
- two types of transaction costs at a later stage:
 - monitoring the contracts (i.e., checking that the other party abides by its terms, may be costly,
 - enforcement of contracts may involve high legal costs.

The approach taken by Coase and Williamson assumes that a major concern of organization design is the minimization of transaction costs.

Formalization of the *ex ante* causes of incompleteness are difficult. There are no generally accepted theories of complexity or of individual decision-making under unforeseeable contingencies. When contingencies unspecified in contracts occur, it is almost certain that there will be conflicts between the concerned parties. It is important to analyse the decision processes that *ex post* handle the unforeseen contingencies. Two polar cases can be distinguished here.

⁸ It should be noticed that in some cases distinction between the types of transaction costs may be hard.

The simplest decision process is the bargaining considered in the previous chapter. Under the assumption of no contract signed *ex ante*, the parties were assumed to bargain over the trading and transfer decisions after investing and learning their valuations and costs. The only legal constraint was that the *ex post* trading be voluntary. The results of bargaining were contrasted with the solution obtained under a complete contract.

Clearly, there should also be considered the intermediate forms of contracting between no *ex ante* contract (and unconstrained *ex post* bargaining) and complete contract. The intermediate forms can allow saving on transaction costs relative to complete contracts, but without the undesirable effects of unconstrained bargaining. Two key possibilities will be considered:

- the two concerned parties seek the intermediation of a third party that is supposed to make the efficient decisions closely resembling the results of complete contracts; it is called arbitration,
- one of the two concerned parties is given the right to determine what happens in an unspecified contingency, it is called authority.

The consequences of each of the possibilities will be analysed in the consecutive sections.

3.2. Arbitration

When two concerned parties seek the intermediation of a third unconcerned party we call the arbitration to be external. For example, a union and a firm may agree to go to binding arbitration if negotiations on labour contract are stalled. Similarly, a seller and a buyer may agree to follow arbitration by outside experts.

External arbitration could be costly, because the outsiders may need to obtain the relevant information in order to make an efficient decision. That is a reason why internal arbitration is viewed as more efficient. Williamson (1975) stresses the superiority of internal arbitration in the setting of disputes. Such arbitration is definitely possible when conflict arises within an organization.

There are at least two important conditions for the arbitration to be efficient. First, an arbitrator must be able to understand the situation at relatively low cost. Not only external arbitration could be costly. Internal arbitration may also involve an expensive use of executives' time. Second, arbitrator must be independent, i.e. he should be able to make decisions that follow joint interests of the two parties, and not favour one at the expense of the other. Arbitrators must be trusted, i.e. they must have a reputation for settling disputes "fairly."

3.3. Authority and its scope

Rather than using an arbitrator, it is possible to give the authority – the power to fill unspecified contingencies – to one of the concerned parties. Assigning the authority does not exclude a possibility that the concerned parties will negotiate *ex post*.⁹ It may be the case when the preferred decision of the party with authority is very costly to the other party, and there exists an alternative mutually beneficial decision. In such cases the party with authority may decide not to exercise its power. In the literature it is stated that authority changes the status-quo point in the bargaining process; the party with authority has more bargaining power. Again, the *ex post* division of the gains from trade will affect *ex ante* investments.

The key issue to be considered then is: how the distribution of authority affects the division of the gains from trade and incentives to invest?

Let us analyse the following example:

- a buyer and a supplier contract to trade tomorrow
- the parties agree to trade in any case
- the only uncertainty is the final specification of the good
- at $t = 1$: basic design is known and contracted for
- at $t = 2$: an opportunity may arise to improve quality that cannot be described at $t = 1$
- the improvement (if any) is learned by both at $t = 2$
- $c (> 0)$ is the cost to the supplier at $t = 2$ (known at $t = 1$, independent of the improvement)
- at $t = 1$: the buyer picks an investment
- at $t = 2$: the value of the improvement to the buyer is
 - $v (> c)$ with prob. x , and
 - 0 with prob. $(1 - x)$
- the cost of investment (l) equals to $x^2/2$
- l is not observable to outsiders
- note: v , and c go beyond the basic design

The following interpretation of the investment technology has been given by Tirole (1997). The buyer invests in flexibility, i.e., a higher investment increases the probability that the improvement (if made) will be used. For example, a power company (buyer) can choose a more costly, but more flexible boiler designed to adapt to various quality of coal from a mine (supplier).

The socially optimal investment is given as a solution to the following program:

$$\max_x [x(v - c) - x^2/2]$$

⁹ See Grossman and Hart (1986); and Hart and Moore (1990).

The f.o.c. are given as: $v - c - x = 0$, thus the optimal investment equals $x^* = v - c$. The total surplus amounts to: $W^* = (v - c)^2/2$.

Three different institutions will be considered:

- 1) unconstrained bargaining (the parties bargain at $t = 2$ over whether to make the improvement; if they do not agree, the improvement is not made, because it was not specified by a contract);
- 2) supplier control (the supplier has the right to decide whether to make the improvement);
- 3) buyer control (the buyer decides whether the improvement is to be made).

For the last two cases, it is assumed that the party who has the authority can also bargain and offer to give this authority away (for some compensation, i.e. transfer). The parties choose the institution that maximizes expected joint surplus, because the gains from changing the institution to a more efficient one can always be redistributed through a transfer at date 1.

First consider the unconstrained bargaining (B):

- the parties trade if $v > c$, and each gets $(v - c)/2$
- the buyer's choice of investment is based on the following program:

$$\max_x \left[\frac{x(v - c)}{2} - \frac{x^2}{2} \right]$$

The f.o.c. are given as: $(v - c)/2 - x^B = 0$, thus $x^B = (v - c)/2 = x^*/2$. It means that there will be underinvestment; joint surplus amounts to: $W^B = x(v - c) - x^2/2 = 3(v - c)^2/8 = 3W^*/4$.

Second institution to be considered is the supplier control (SC). This case is equivalent to unconstrained bargaining. If the parties disagree, the supplier chooses the less costly action for him, i.e. not to invest. His right to order strict compliance with the original contract serves as a threat to get a good bargaining position. Hence, we have: $x^{SC} = x^B$ and $W^{SC} = W^B$, i.e., the buyer's investment is partially expropriated, thus the buyer underinvests.

The last institution to be considered is the buyer control (BC). In this case, the improvement will always be made if the status quo were not renegotiated.

- If the value is v , the status quo is efficient and there is no renegotiation; the buyer gets v by imposing the improvement.
- If the value is 0 , the status quo is inefficient, and renegotiation divides evenly the gain (c) of not making the improvement; the buyer gets $c/2$.

The optimal choice of investment for the buyer is given by the following program:

$$\max_x \left[xv + (1 - x) \frac{c}{2} - \frac{x^2}{2} \right]$$

The f.o.c. are given by: $v - c/2 - x^{BC} = 0$, which yields $x^{BC} = v - c/2 > x^*$. It means that the buyer overinvests, because his authority allows him not to pay

the production cost c if his value is v , but the probability of production increases. The joint surplus amounts to:

$$W^{BC} = x(v - c) - x^2/2 = (v - c/2)(v - 3c/2)/2.$$

Observe that it may be optimal to give the authority to the buyer or to the supplier:

- when $c = 0$, the buyer control is socially efficient, but supplier control is not,
- if $v = c > 0$, then no investment is optimal, and the supplier control (or unconstrained bargaining) is optimal; buyer control is not, because it encourages investment and yields a negative W .

3.4. *Substitutes for contracts*

The key conclusion from the above analyses is that, in order to avoid future hazards, parties should sign complete contracts, or if contracts are too costly or impossible to write, the parties should make a correct use of the authority structure, i.e. to sign a restricted contract. It has been however found long time ago that relations between firms tended to be more informal than predicted by the theory.¹⁰ The long-run relationships are often sustained by the firms' reputation. Cheating creates the risk of losing future profitable deals.¹¹ Reputation allows a firm to save on the costs of writing complete contracts or even on the costs of assigning authority. It is important to remember however that informality exposes the firms to the threat of opportunism. Thus, informality could be expected mostly when specific investments are limited and when trade is sufficiently frequent.

Another way of avoiding the *ex post* hold-up problem is to introduce *ex post* competition whenever feasible. Farrell and Gallini (1986) and Shepard (1986) have analysed Williamson models in which the buyer invests in specific assets and the seller chooses, *ex post*, some variable (a price or a delivery lag) that is noncontractable *ex ante*. *Ex post*, the seller has an incentive to choose low levels of the decision variable, which causes the buyer to invest little in the relationship *ex ante*. Having two or more suppliers who compete *ex post* constitutes a dual sourcing that raises the equilibrium level of the decision variable and the *ex ante* investment. Practical examples of such competition are: Intel licensing its microprocessor technologies, or IBM adopting an "open architecture" policy in regard to its personal computers.

¹⁰ Compare, for example, MaCaulay (1963).

¹¹ See Williamson (1975, chapter 6).

3.5. The property rights approach

The property rights theory of the firm and market organization is a path-breaking contribution by Grossman and Hart (1986), Hart and Moore (1990) and Hart (1995). It builds on and formalizes the institutions of transaction cost economics introduced by Coase and Williamson only in certain respects: complex contracts are incomplete (by reason of bounded rationality, contract as mere promise is not self-enforcing (by reason of opportunism), court ordering of conflicts is limited (by reason of nonverifiability) and the parties are bilaterally dependent (by reason of transaction-specific investments). But whereas transaction cost economics locates the main analytical focus on the governance of ongoing contractual relations, the property rights theory of the firm gets rid of governance issues by assuming common knowledge of payoffs and costless bargaining. Thus, the entire analytical action is focused on the alignment stage of contracting. Since the assumptions of common knowledge and costless bargaining are rather problematic, Williamson (2002) argues that the property rights theory obscures the key interactions instead of spotlighting them.

There is a disagreement between researchers (for example Williamson vs. Hart) about the explanatory power of the property rights approach versus the transaction cost approach. Hart (1995) argues that the property rights framework allows explaining the costs of integration better than the transaction cost economics. Since property rights theory rests only on asset ownership, it disputes all four of the following propositions of transaction cost economics:

- 1) firms enjoy advantages over markets in cooperative adaptation respects,
- 2) incentive intensity is unavoidably compromised by internal organization,
- 3) administrative controls are more numerous and more nuanced in firms,
- 4) the implicit contract law of internal organization is that of forbearance; the firm is its own court for resolving disputes.

Property rights theory makes limited appeal to data, because it yields very few refutable implications and is indeed untestable. By contrast, the transaction cost theory yields numerous refutable implications and encourages empirical testing.

3.6. Other formal approaches to the theory of the firm

Bajari and Tadelis (2001) and Tadelis (2002) offer another approach to the firm in the spirit of transaction cost theory. They provide a formal model of *ex post* adjustments under incomplete contracting. In their basic model, completeness and complexity are chosen simultaneously. The ownership gives the contracting party the right to modify the project design *ex post*. The buyer benefits from integration (internal procurement) as it can request changes to maximize own benefit *ex post*.

It comes at the costs of integration, i.e., weaker incentives for the seller. As a result the internal procurement is an increasing function of complexity.

A key innovation introduced by Baker, Gibbons, and Murphy (2002) was adding a new dimension for characterizing organizational form. Their main proposition is that asset ownership affects parties' temptations to renege on a relational contract. On the one hand, the integration increases the downstream party's incentive to renege on the promise; under non-integration, if the downstream party reneges, the upstream party can sell the good to an alternative user. On the other hand, the non-integration increases the upstream party's incentive to increase the value to the alternative user, increasing its bargaining position should the downstream party renege. Thus we have a trade-off between integration and non-integration.

Considering the three Coasian questions about the firm – existence, boundaries, and internal organization – boundaries received the most attention by economists. The Williamson's theory of the firm is the best-known, but not the only economic approach to vertical boundaries.

Now, we will turn to the discussion of the objectives of the firm. In neoclassical theory, the firm is characterized by a production possibilities set, and it is assumed to maximize profits. The decision makers had perfect knowledge about the production possibilities, and the market situation. There have been several arguments why the firm would seek to maximize profit, but they are not entirely undisputable.

One of the typically cited reasons is that firms maximizing profit will drive out firms that don't by some "natural selection" process. Whether this statement makes sense is not immediately obvious, and requires a detailed model of the natural selection process.

A second argument is based on the assumption that the managers of a publicly traded firm who don't maximize profits create a risk for the firm to be taken over by some corporate raider who will fire the current management and make money by maximizing profit. Since the managers prefer to keep their jobs, they maximize profits.

Formal analysis of the takeover mechanism as a disciplinary device on the incumbent management has been an area of intensive research in microeconomic theory over at least last three decades. The analysis has shown that takeovers have their limits. One of the key issues here is the free-rider problem among the existing shareholders that may impair the incentives of a potential raider to attempt an acquisition of the profit-non-maximizing firm. The literature on the free-riding problem in this context started with the static model of Grossman and Hart (1980), and was further extended to the dynamic setting by Harrington and Prokop (1993).¹²

¹² An excellent overview of this topic may be found in Tirole (2006, chapter 11) or in Prokop (2001), in the Polish-language literature.

The takeover argument for profit-maximization may not be entirely erroneous, but the analyses that have been performed so far do not provide unambiguous support of the justification.

A third argument says that profit maximization is in the best interest of shareholders of the firm. Furthermore, the shareholder can and do create proper incentives for managers to act in the interest of the shareholders, i.e. the managers maximize profits.

There are two main reasons why shareholders might not want to maximize profit (or net market value). Firstly, they may be risk-averse, in which case they may prefer decisions leading the firm's profit to be negatively correlated with the economy's fluctuations (in order to get a less risky portfolio) even when such choices do not quite maximize profit. Secondly, a firm's price for one of its products, for example, may influence the shareholders' welfare not only through the firm's profit but also through their consumption of the firm's good as long as the firm is not perfectly competitive.¹³ These two general-equilibrium effects are important theoretically, but they do not seem to be very strong empirically. The shareholders' portfolio is relatively well diversified, and it is not clear how most decisions of holders' (at least the influential ones') consumption of their firms' products is usually very small, so that price effects are small relative to the income effect generated by the firm's profit levels, as well.

Even when it is in the best interest for the shareholders to maximize profit of the firm, it is not obvious that they can force the managers to follow the shareholders' interests. The inability of shareholders to adequately monitor the managers and discover the firm's cost-and-demand situation has been pointed out as the cause of the principle-agent problem. The principle-agent literature considers how the moral-hazard problem can be reduced through different incentive schemes like direct monetary (in a broad sense) awards, yardstick competition, product-market competition, and supervision. However, we should be aware that there are limits to all of these control mechanisms.

An alternative to profit maximization could be an attempt of those who actually control the firm (usually the managers) to follow some other objectives. The managers are motivated by things other than their salaries. They like prestige that could be derived for example from the size of their organization measured in terms of gross sales or in terms of total capital under management. We might imagine a manager whose objective is to maximize some combination of profits, sales, and capital; or one who seeks to maximize sales or total assets subject to a minimum profit constraint. A manager working in a risky environment may decide to undertake less profitable but safe projects in order to avoid being fired by the angry shareholders in the case of failure to earn some minimum rate of return.

¹³ Shareholders may not even be able to agree on the firm's best choice of action in such cases.

Managerial theories of the firm try to model the firms' decisions along the above lines of consideration. This kind of theory is not much in fashion these days, but it presents a variation on the neoclassical model of the firm – firms as maximizing entities with given production possibilities sets.

A further variation on the neoclassical model is an assumption about the firm acting in a particular way that cannot be described as maximizing anything at all. In this approach, the firm is considered as a behavioural entity that is especially relevant in a dynamic setting that allows capturing certain patterns of behaviour over time.

This is a completely new mode of analysis in economics, and it is clearly one that deserves further development in building models, running simulations, doing more exact analysis, and developing procedures for calibrating and testing models. Certainly, the sort of behavioural patterns that are naturally built into these models are more realistic than those in simple neoclassical models. A true believer in the neoclassical theory would have a ready retort to all this claiming that it is possible to incorporate additional (adjustment, search etc.) costs into the analysis. But to accommodate all these things within the neoclassical framework could be difficult if not impossible. In any event, the notion of a single manager deciding on a production plan for the entire firm does not seem very descriptive of the way firms make decisions.



Principle-Agent Analysis

4.1. The basic incentive problem

The basic incentive problem (the principle-agent problem) in its moral-hazard form stems from a conflict between insurance and incentives. The theory of optimal insurance demonstrates that the optimal division of a pie of a random size (the profit) between a risk-neutral party (the shareholders) and a risk averse one (the manager) has the risk-neutral party bear all the risk, if incentive issues are left aside. The profits from the set $\{\Pi_1, \dots, \Pi_r, \dots, \Pi_n\}$ occur, respectively, with probabilities $\{p_1, \dots, p_r, \dots, p_n\}$, where $p_i > 0$, and $\sum_{i=1}^n p_i = 1$.

Let $\omega(\Pi)$ be the risk-averse party's profit when the realization is Π , and $\Pi - \omega(\Pi)$ be the risk-neutral party's profit when the realization is Π . The expected utilities are as follows:

- for the risk-neutral party:

$$E_{\Pi}[\Pi - \omega(\Pi)] = \sum_i p_i (\Pi_i - \omega_i)$$

- for the risk-averse party:

$$E_{\Pi} u[\omega(\Pi)] = \sum_i p_i u(\omega_i)$$

where $\omega_i = \omega(\Pi_i)$.

A Pareto-optimal contract maximizes the utility of one party given the level of utility of the other party:

$$\max_{\{\omega_i\}} \sum_i p_i (\Pi_i - \omega_i), \text{ s.t. } \sum_i p_i u(\omega_i) \geq U_0$$

The Lagrangian for the above program is:

$$L = \sum_i p_i (\Pi_i - \omega_i) + \lambda \left(\sum_i p_i u(\omega_i) - U_0 \right)$$

The f.o.c. are given as:

$$\frac{\partial L}{\partial \omega_i} = -p_i + \lambda p_i u'(\omega_i) = 0, \text{ or } u'(\omega_i) = 1/\lambda$$

thus ω_i is independent of i , if $u'' < 0$ (risk averse), hence the risk-averse party should get full insurance, i.e. should have a constant income.

Consider a situation in which one party, called the principal, hires a second party, called the agent, to perform some task. For example, the shareholders act as the principal, and the manager serves as an agent. The manager is willing to undertake the task as long as his net utility from performing the task is at least as large as he can get at his next best opportunity, e.g. working outside the firm; we refer to this level of utility as the agent's reservation level of utility. The manager, when he is hired, chooses between two levels of effort: high ("work hard") and low ("shirk"). Hard work is associated with a disutility for the manager, so all other things equal, he would prefer not to work hard. Whether this agent works hard or not determines the value to the principal.

Specifically, suppose that the manager has utility $u(\omega - \Phi)$ when he works and $u(\omega)$ when he shirks, where ω ($\omega > 0$) is the manager's wage, and Φ ($\Phi > 0$) is the monetary disutility of high effort. It is typically assumed that $u(\cdot)$ is an increasing, concave function (with $\lim_{w \rightarrow \infty} u'(w) = +\infty$). The manager's objective function is the expectation of u , denoted $E(u)$. Working outside the firm, he would get $U_0 \equiv u(\omega_0)$, where ω_0 is called the reservation (net) wage.

A firm's profit, Π , may take one of two values: Π_1 or Π_2 (assume $\Pi_1 < \Pi_2$). The shareholders' objective function is the expectation of the net profit, i.e., $E(\Pi - \omega)$.

The technology is as follows:

- if the manager works hard, the profit is Π_2 with probability x , and Π_1 with probability $(1 - x)$;
- if the manager shirks, the profit is Π_2 with probability y , and Π_1 with probability $(1 - y)$; (with the usual assumption $0 < y < x < 1$).

The manager's contract is assumed to be offered by the shareholders. In this contract, let's denote by ω_1 and ω_2 the manager's wage when the profits of Π_1 and Π_2 are earned, respectively.

First, consider the case when the manager's effort is observed by the shareholders, who can choose any level of effort they want and impose on the manager. Since effort is observable, there is no incentive issue. When the shareholders demand a low level of effort, the optimal contract will be achieved at $\omega_1 = \omega_2 = \omega_0$, i.e. the shareholders give the manager exactly his reservation wage. The shareholders' profits are:

$$y \Pi_2 + (1 - y) \Pi_1 - \omega_0.$$

Next, suppose the shareholders demand high level of effort. Now, optimal contract calls for $\omega_1 - \Phi = \omega_2 - \Phi = \omega_0$. The shareholders' expected profits are then:

$$x \Pi_2 + (1 - x) \Pi_1 - (\omega_0 + \Phi)$$

The shareholders decision will be based on the comparison between the above expected profits. Often (but not always) it is the case that high effort is better for shareholders, i.e.:

$$y \Pi_2 + (1 - y) \Pi_1 - \omega_0 < x \Pi_2 + (1 - x) \Pi_1 - (\omega_0 + \Phi)$$

or

$$\Phi < (x - y) (\Pi_2 - \Pi_1)$$

If the increase in expected profits exceeds the monetary disutility of high effort, then it pays for the shareholders to choose the high level of effort.

The more interesting case is when the manager's effort is not observable by the shareholders. The constant wage structure employed in the case of observable actions cannot induce the high level of effort any more. If shareholders still like to induce the high level of effort, they must design a wage structure that satisfies two conditions:

- 1) the "incentive-compatibility" constraint,
- 2) the "individual rationality" or "participation" constraint.

The incentive compatibility constraint guarantees that the agent prefers to work hard rather than shirk, and can be written as:

$$y u(\omega_2) + (1 - y) u(\omega_1) \leq x u(\omega_2 - \Phi) + (1 - x) u(\omega_1 - \Phi)$$

where ω_i is the wage paid when the realized profits are Π_i .

The "individual rationality" or "participation" constraint means that working hard is better than the outside option of the manager, and is written as:

$$u(\omega_0) \leq x u(\omega_2 - \Phi) + (1 - x) u(\omega_1 - \Phi)$$

The shareholders' expected profit is then:

$$x (\Pi_2 - \omega_2) + (1 - x) (\Pi_1 - \omega_1)$$

Observe that in the maximization of the shareholders' profit with respect to the two constraints, they both must be binding. Thus, in our simple case, the optimal wage structure is obtained from two constraints satisfied with equality, i.e.:

$$y u(\omega_2) + (1 - y) u(\omega_1) = x u(\omega_2 - \Phi) + (1 - x) u(\omega_1 - \Phi)$$

$$u(\omega_0) = x u(\omega_2 - \Phi) + (1 - x) u(\omega_1 - \Phi)$$

Notice that the shareholders' profit is lower under unobservability, and the expected wage bill is higher $x \omega_2 + (1 - x) \omega_1 > \omega_0 + \Phi$, i.e., to induce the high effort and obtain the high profit with probability x , the wage must be higher than under effort observability.

On the other hand, if the shareholders wanted to induce the low effort level under observability, the relative desirability of inducing the high effort would be even lower under unobservability; that is the shareholders may be better off with the low effort under unobservability, even though they may prefer high level of effort under observability.

4.2. An example of an optimal incentive scheme

Suppose that the agent's reservation level of utility is $U_0 = 9$. The agent derives utility from how much he is paid, w , and how hard he works, a . The level of a can be "hard" or "high," which we denote $a = 5$, or it can be "not hard" or "low," denoted by $a = 0$. The agent's overall (von Neumann-Morgenstern) utility from w and a is given by:

$$U(w, a) = \sqrt{w} - a$$

If the agent works hard, the accomplished task is worth 0 (no sale) with probability 0.1, \$100 (sale's revenue) with probability 0.3, and \$400 (sale's revenue) with probability 0.6. And if the agent does not work hard, the three probabilities are 0.6, 0.3, and 0.1. The size of the sale is observable, but the agent's effort is not.

(a) What is the optimal way to induce the agent to put in a low level of effort, if you can contract contingent on the size of the sale (only)?

To optimally induce the agent to put in a low level of effort we must:

$$\min\{0.6w_1 + 0.3w_2 + 0.1w_3\}$$

subject to:

$$0.6\sqrt{w_1} + 0.3\sqrt{w_2} + 0.1\sqrt{w_3} \geq 9$$

We solve the above problem using the Lagrangian multipliers:

$$L = 0.6w_1 + 0.3w_2 + 0.1w_3 + \lambda(0.6\sqrt{w_1} + 0.3\sqrt{w_2} + 0.1\sqrt{w_3} - 9)$$

The first order conditions for the above Lagrangian are:

$$\frac{\partial L}{\partial w_1} = 0.6 + \frac{0.6\lambda}{2\sqrt{w_1}} = 0$$

$$\frac{\partial L}{\partial w_2} = 0.3 + \frac{0.3\lambda}{2\sqrt{w_2}} = 0$$

$$\frac{\partial L}{\partial w_3} = 0.1 + \frac{0.1\lambda}{2\sqrt{w_3}} = 0$$

$$\frac{\partial L}{\partial \lambda} = 0.6\sqrt{w_1} + 0.3\sqrt{w_2} + 0.1\sqrt{w_3} - 9 = 0$$

Solving the above system of equations, we obtain $w_1 = w_2 = w_3 = 81$, i.e. the wage is independent of the size of the sales (full insurance of the risk-averse agent).

(b) What is the optimal way to induce the agent to put in a high level of effort?

To optimally induce the agent to put in a high level of effort we must:

$$\min \{0.1w_1 + 0.3w_2 + 0.6w_3\}$$

subject to:

$$0.6\sqrt{w_1} + 0.3\sqrt{w_2} + 0.1\sqrt{w_3} \leq 0.1\sqrt{w_1} + 0.3\sqrt{w_2} + 0.6\sqrt{w_3} - 5$$

and:

$$9 \leq 0.1\sqrt{w_1} + 0.3\sqrt{w_2} + 0.6\sqrt{w_3} - 5$$

We solve the above problem again using the Lagrangian multipliers:

$$L = 0.1w_1 + 0.3w_2 + 0.6w_3 + \lambda_1(-0.5\sqrt{w_1} + 0.5\sqrt{w_3} - 5) + \lambda_2(0.1\sqrt{w_1} + 0.3\sqrt{w_2} + 0.6\sqrt{w_3} - 14)$$

The first order conditions for the above Lagrangian are:

$$\frac{\partial L}{\partial w_1} = 0.1 - \frac{0.5\lambda_1}{2\sqrt{w_1}} + \frac{0.1\lambda_2}{2\sqrt{w_1}} = 0$$

$$\frac{\partial L}{\partial w_2} = 0.3 + \frac{0.3\lambda_2}{2\sqrt{w_2}} = 0$$

$$\frac{\partial L}{\partial w_3} = 0.6 + \frac{0.6\lambda_2}{2\sqrt{w_3}} = 0$$

$$\frac{\partial L}{\partial \lambda_1} = -0.5\sqrt{w_1} + 0.5\sqrt{w_3} - 5 = 0$$

$$\frac{\partial L}{\partial \lambda_2} = 0.1\sqrt{w_1} + 0.3\sqrt{w_2} + 0.6\sqrt{w_3} - 14 = 0$$

Solving the above system of equations, we obtain $w_1 = 29.47$, $w_2 = 196$, $w_3 = 238.04$. The expected wage bill is $(0.1)(29.47) + (0.3)(196) + (0.6)(238.04) = 204.57$.

(c) What is the optimal contract (if you are a risk neutral principal) to offer this agent?

We need to compare the expected gains of the principal from the above two cases. In the case of inducing the low effort level the expected profit of the principal is:

$$0.6(0 - 81) + 0.3(100 - 81) + 0.1(400 - 81) = -11$$

but in the case when the principal induces the high level of effort, his expected profit is:

$$0.1(0 - 29.47) + 0.3(100 - 196) + 0.6(400 - 238.04) = 65.43$$

Clearly, based on the above analysis, the principal should prefer to induce the agent to work hard. Thus the optimal contract offered by the principal should be:

- if no sales are made, the wages are \$29.47,
- if a \$100 sales are made, the wages are \$196,
- if \$400 sales are made, the wages are 238.04.

Despite our great abilities to determine optimal contracts, we should be aware that the modern agency theorists are too optimistic that various governance mechanisms have solved agency problems. Those who are studying the economics of organizations should have grave doubts about the overly optimistic (Chicago School) view that agency problems are solved. The research literature shows that the strong ideological fervour of Chicago School economists in espousing the virtues of the market has lead these economists more to wishful thinking than to coming to terms with agency problems that are well documented in the world of experience.

4.3. Observability, verifiability, authority

We should make a distinction between “observability” that has been discussed above and “verifiability” that relates to the possibility that the principal can observe the agent’s performance but cannot verify his observations to a court, i.e., cannot supply sufficient evidence (the facts speak for themselves, but they tell different things to different people). When performance cannot be verified by a court, contracts that are contingent on performance (e.g., based on profit or sales) cannot be signed, as the courts will be unable to enforce them. An example of such a case is the team work. When an agent is a part of a team, reliable accounting procedures may measure only the team’s performance, not individual contributions. However, an insider (a supervisor, other team workers) may be able to disentangle these contributions, whereas an outsider (a judge) cannot. The same argument applies to the performance of complementary divisions of a firm, like production and marketing.

Suppose that in an agency problem profit Π is observable by the principal but is not verifiable, so the contract cannot depend directly on the agent’s performance. In this case, the principal cannot be trusted to announce truthfully what he observes. When profit is only observable, the principal has an incentive to claim that it is low even when it is high, since the payment to the agent in the case of the lower profit is smaller than in other case.

This picture changes significantly when the principal oversees many agents at the same time. Consider a large number N of agents, each of whom produces an unobservable but unverifiable profit. As in the previous model, the probability that the individual performance is Π_2 rather than Π_1 is x or y , depending on whether or not the agent exerts high effort. The principal

may commit himself to pay higher wage w_2 to x percent of agents (the ones he announces to be the most productive), and as lower wage w_1 to the rest, where w_1 and w_2 are the optimal wages under verifiability. Notice that the total wage bill $N[xw_2 + (1 - x)w_1]$ is fixed, and the principal does not have any incentives to misrepresent the individual performance (as long as he does not collude with some of the agents). On the other hand, if the agents all exert effort, they know that x percent of them will create high profit Π_2 (by the law of large numbers) and will receive wage w_2 . Those generating low profit Π_1 will receive wage w_1 . It means that the incentive-compatibility constraint and the individual-rationality constraint are both satisfied. In the case of many agents, the principal can be given the authority to choose rewards, because the verifiability is obtained indirectly.

4.4. *Limits to discretion – yardstick competition*

Even when an agent's individual performance is verifiable, it is only a garbled measure of the agent's effort. For instance, a profitable investment reduces current profits without reflecting managerial slack or ineptitude. But such factors, which are unverifiable because of accounting manipulations, may be observable by the market and reflected in the firm's valuation. That is why managerial compensation is contingent on the value of the firm as well as on its current profits.

Whether a firm's low profit is due to a decrease in the market demand, or an increase in costs rather than to managerial slack, can be detected, to some extent, by comparing the agent's performance with that of other agents placed in similar conditions.¹⁴

Let us consider an example of yardstick competition. Imagine the shareholders overseeing two managers in charge of two similar divisions. The profits earned by the shareholder are equal to the sum of the profits generated by each manager, net of the expected wage bill. Assume that the probability of generating profit Π_2 is x or y depending on whether the manager works hard or not. We assume that the uncertainties facing both managers are perfectly correlated, so that the same level of effort generates the same amount of profit. It means that if the managers choose to work hard, the realized profit is either Π_2 for both (which occurs with probability x) or Π_1 for both (which happens with probability $1 - x$), and similarly when they both decide to shirk.

In such a case the shareholders can offer the full-information wage $w_1 = w_2 = w_0 + \Phi$ when both managers achieve the same level of profit (no matter whether high or low). If profits differ between managers, the high-profit manager

¹⁴ Original discussion of yardstick competition can be found in Lazear and Rosen (1981), Green and Stokey (1983), Nalebuff and Stiglitz (1983) and Shleifer (1985).

receives $w_2 = w_0 + \Phi$ and the low-profit manager gets heavily punished. Thus each manager's wage depends not only on his own performance, but also on the performance of the other one. Working hard for both managers constitutes equilibrium of this game.¹⁵ The manager who is expected to work hard by earning the high profit automatically reveals that the other manager who gets the low profit did not work hard. The low-profit manager cannot attribute his poor performance to "adverse circumstances," and is heavily punished.

The idea of yardstick competition carries over to situations with imperfect correlation, which is a more reasonable assumption. The applications of yardstick competition are numerous. The CEOs of companies may compare the performances of managers of different divisions facing similar costs or demand conditions. Similarly, the compensation of the managers of one company can be made contingent on the performance of a competitor's management. Alternatively, managerial pay can be connected to the average industry profits.

Yardstick competition also has its limits. Often the comparison between the situations of different units may be hard. In addition their performance may be garbled by accounting procedures or measurement errors. Another issue is the dependence of the managers' performance on the initially inherited assets. All these difficulties explain why there is little yardstick competition in some markets, for example, in the electricity-generation industry.¹⁶

4.5. *The dynamics of managerial incentives*

Another reason why managers may refrain from discretionary behaviour is the concern about their own career not only within the firm, but also outside this firm. The arguments that can be made relate to the interaction between adverse selection (how efficient or trustworthy is the manager?) and moral hazard (how hard does he work?).

A manager who has performed poorly may not be trusted to do well in the future; therefore his prospects for promotion within a firm may be weak. Some authors have focused on the repeated aspects of the relationship between a firm and its employees. To formalize this process suppose that in each period the quality of an employee's performance is either high or low. The quality is observable by the employer, but it is not verifiable. As long as the quality was high before, the employee is offered a rent in each period (a rent induces the employee to strictly prefer staying with the firm rather than quitting). If the employee "cheats" and performs with a low quality, the firm stops offering him the rent (e.g. does not promote him or does not increase his pay). The threat of

¹⁵ Notice that there is a multiplicity of equilibria in this game, e.g., both managers shirking is also an equilibrium; see Mookherjee (1984).

¹⁶ See Joskow and Schmalensee (1986).

losing this rent puts some discipline on the employees. In terms of models we can describe this situation by the use of supergames with asymmetric information.

Another possibility for the manager is to leave the firm. The threat of outside options (offers) may discipline the firm and force it to treat its managers fairly when the quality of their work is observable but not verifiable. Consider a professor at a university. The quality of his work can be measured by the quality of his research and the richness of his interactions with colleagues and students. This quality may be observable by the profession but hard for a court to assess. His threat to move to another university tends to equalize his position within and outside his current workplace.¹⁷

The possibilities of getting good opportunities outside his present firm and being rewarded within the firm give a manager incentive to perform satisfactorily. When only manager's performance is observable, Holmström (1982) shows that a manager at the beginning of his career may work even harder than is socially optimal.

¹⁷ This idea was developed by Fama (1980) and Holmström (1982)



Empirical Research in the Theory of the Firm

5.1. Evidence on boundaries of the firm

According to the modern theory of the firm, the boundaries of the organization depend not only on the productive technology, but also on the costs of transacting business. In this framework, the decision to organize transactions within the firm as opposed to open market depends on the relative costs of internal versus external exchange. Thus, the firm boundaries are determined by the trade-off between the relative transaction costs of “make” or “buy,” i.e. they depend not only on technology, but also on various contracting alternatives.

The empirical literature on the make-or-buy decision usually adopts the transaction cost framework and follows the same basic model. The efficient organizational form (also called governance structure) for a particular business relationship is viewed as a function of the underlying transactional characteristics, such as: asset specificity, uncertainty, frequency of dealings, etc. Organizational form, then, is the dependent variable, while the properties of the transactions serve as independent (explanatory) variables.

Organizational form is mostly modelled as a discrete variable, e.g. “make,” “buy,” or “hybrid,” but it is sometimes measured by a continuous variable. Among the dependent variables, asset specificity has received the most attention, because it plays a central role in the contractual or transaction cost view of the firm. Williamson (1991) named six types of asset specificity:

- site specificity,
- physical asset specificity,
- human asset specificity,
- brand-name capital specificity,

- “dedicated assets,”
- temporal specificity.

Among the typical empirical proxies for asset specificity are technical characteristics. For example, product complexity serves as a proxy for physical asset specificity; technical specifications are qualitatively coded from survey data or quantitatively assigned by inspection.¹⁸

Worker-specific knowledge, also coded from survey data, serves as a proxy for human asset specificity,¹⁹ and physical proximity is used as a proxy for site specificity,²⁰ and special and temporal specificity.²¹

When asset specificity cannot be easily measured, concentration has been used in single-industry studies to capture situations where small-numbers bargaining situations are likely to appear.²² Common proxies for uncertainty include sales variance²³ and some measure of technological uncertainty, such as the frequency of changes in product specification and the probability of technological change.²⁴

The empirical work consists of:

- qualitative case studies,
- quantitative case studies focusing on a single firm or industry, and
- econometric analysis of cross-sectional or panel data from multiple firms or industries.

Among the qualitative case studies, the prominent ones to be mentioned is the study of cable TV franchising in Oakland, California by Williamson (1976) and the reinterpretation of the G.M.-Fisher Body case by Coase (2000).

The investigation of contracting practices in a large aerospace corporation by Masten (1984) is an example of a quantitative case study focused on a single firm. The investigation of electric contracts by Saussier (2000) belongs to the quantitative case studies focusing on a single industry. Other single-industry case studies include Stuckey (1983) on the aluminium, Palay (1984) on rail shipping, Gallick (1984) on tuna processing, Joskow (1985) on coal-burning electric plants, Goldberg and Erickson (1987) on petroleum coke, Pirrong (1993) on ocean shipping, Ohanian (1994) on pulp and paper, Ménard (1996) on poultry, and Martinez (2002) on poultry, eggs, and pork.

Among cross-industry analyses, Levy (1985) focused on manufacturing, and John and Weitz (1988) studied forward integration into distribution.

The studies on the make-or-buy decision are based on cases, because the main variables of interest, i.e. asset specificity, uncertainty, frequency, are dif-

¹⁸ See Masten (1984), or Bigelow (2001).

¹⁹ See Monteverde and Teece (1982a and 1982b), and Masters and Miles (2002).

²⁰ Joskow (1985, 1987, 1988), and Gonzalez-Diaz, Arrunada, and Fernandez (2000).

²¹ Pirrong (1993) and Hubbard (1999).

²² See Ohanian (1994).

²³ See Levy (1985), Anderson and Schmittlein (1984).

²⁴ See Walker and Weber (1984), and Crocker and Reynolds (1993).

difficult to measure. In many earlier works these characteristics were estimated based on surveys or interviews, e.g. a manager might be asked to rate the degree to which an investment has value in outside uses, on a Likert-type 1 to 7 scale. Clearly, such data are subject to the general limits to the survey data, because they are based on the respondents' stated beliefs, rather than on their beliefs or valuations as revealed through choice.

There is also a problem of comparison between industries, since these measurements are based on ordinal rankings. What is ranked as a relatively specialized asset in one firm may be rated differently in another firm or industry. Analogously, a comparatively uncertain production process in one firm may be the standard operating environment in another. Thus, multi-industry studies may therefore contain variables that are labelled similarly, but are really incommensurable, or the other way around, variables that are identical could be labelled differently.

In addition to measurement inconsistencies, there is also a problem with proper classification of discrete variables like "make" or "buy". Moreover, the evidence from individual cases may not be generalizable.

Another problem is that the empirical literature on vertical integration assumes that the actually chosen business arrangements are efficient. Recently, researchers have begun to examine this assumption more closely by comparing the performance of firms. Several papers use a two-step procedure in which organizational form is endogenously chosen in the first stage, and then it is used to explain performance in the second stage. By endogenizing both organizational form and performance this approach also mitigates the selection bias associated with OLS regressions of performance on firm's characteristics.²⁵

One important performance measure used in empirical work is firm's survival. However, survival in many cases may not be the best measurement of performance, compared with profitability or market value. Poorly performing firms may survive due to inefficient competitors, regulatory protection, or legal environment. Efficient alignment between transactions and governance should be expected only if the selection environment is strong.

5.2. Studies on make-or-buy decision

The decision to make components internally or procure them on the open market was the first topic studied within the transaction cost framework. Early research by Monteverde and Teece (1982b) and Masten (1984) use samples of component rated as either "made" or "bought," together with asset specificity such as worker-specific knowledge and component complexity as rated by

²⁵ See Poppo and Zenger (1998), Saussier (2000), Nickerson, Hamilton and Wada (2001), Sampson (2001), Macher (2001), and Yvrande-Billon (2004).

engineers. These papers use a probit model to test the relationship between in-house production and asset specificity, along with uncertainty and other control variables. They found asset specificity to be a statistically significant predictor of vertical integration.

The most commonly cited example of a hold up problem solved by vertical integration is the relationship between Fisher Body and G.M. in the 1920s. Klein, Crawford and Alchian (1978) explain the G.M. buyout of Fisher in terms of the specific assets that accompanied the switch from wooden- to metal-bodied cars. Klein (1988) emphasised specific human capital in this case. Some papers argue that the most important motive for vertical integration was uncertainty.²⁶ Helper, MacDuffie, and Sabel (2000) suggested that vertical integration helped speed up collaborative learning. A different view was presented by Freeland (2000), who maintains that the Fisher brothers successfully held up G.M. after becoming employees.

Statistical significance of asset specificity as a determinant of vertical structure was found, besides autos, in many other industries, including aerospace,²⁷ aluminium,²⁸ forestry,²⁹ chemicals,³⁰ engineering,³¹ trucking,³² offshore oil gathering,³³ information technology,³⁴ electronic components,³⁵ construction,³⁶ and even stock exchanges.³⁷

Most of the papers cited above are focused, single-industry case studies. A few studies have used cross-sectional or panel data to estimate the effects of transactional characteristics on vertical integration using multi-industry data. For example, Levy (1985) used the ratio of value-added to sales as a cross-industry measure of vertical integration, the number of firms and amount of R&D spending as a measure of asset specificity, and the variance of sales as a measure of uncertainty. However, it should be noticed that accounting constructs like the ratio of value-added to sales are problematic because value-added figures are reported inconsistently across firms and industries, and there are several accepted methods for computing value-added ratios.

Another problem is that the cross-sectional studies cannot control for time and for unobserved firm-specific characteristics. This limitation can be overcome by using panel data. For example, Gonzalez-Diaz, Arrunada, and Fernandez

²⁶ See Langlois and Robertson (1989).

²⁷ See Masten (1984).

²⁸ See Stuckey (1983), and Hennart (1988).

²⁹ See, for example, Globerman and Schwindt (1986).

³⁰ See Lieberman (1991).

³¹ Lyons (1995).

³² Nickerson and Silverman (2003).

³³ Hallwood (1991).

³⁴ Ulset (1996).

³⁵ Weiss and Kurland (1997).

³⁶ Gonzalez-Diaz, Arrunada, and Fernandez (2000).

³⁷ Bindseil (1997).

(2000) assemble a panel data of Spanish construction firms over a six-year period and study the use of independent subcontractors. They regress the percentage of subcontracting on a distance-based measure of asset specificity, a measure of uncertainty, time- and firm-fixed effects, and other control variables. Again they found that asset specificity is the only significant determinant of vertical structure. Also other studies using panel data support transaction cost explanations even when fixed effects are included.³⁸

Vertical integration could also mean forward integration into marketing and distribution. Anderson and Schmittlein (1984) consider two marketing alternatives for an electronic component producer (dependent variables): the use of employees as direct sales force, or reliance on independent manufacturers' representatives. The regression is made on managers' perceptions of the importance of specific capital, sales volume uncertainty, and measurement uncertainty obtained from the survey data. It is shown that only specific human capital and measurement uncertainty are statistically significant. Several other marketing studies produced similar results.³⁹

Clearly, marketing and distribution depend on other factors as well. Muris, Scheffman, and Spiller (1992) found that the shift from independent bottlers to captive subsidiaries at the end of the last century was heavily influenced by the emergence of national cola markets, which required greater coordination of advertising and promotional activities.

5.3. Empirical analysis of contracting

Initially, the literature on the make-or-buy decision treated external sourcing and in-house manufacturing as polar opposites. As Williamson (1975) calls it, there was a choice between markets and hierarchies. However, what we observe is a variety of intermediate or hybrid forms of organization, such as long-term contracts, partial ownership agreements, franchises, networks, alliances, and other combinations. It turns out that for certain types of transactions an intermediate form of governance is the most appropriate. For example, in the case of asset specificity and negligible uncertainty, the long-term contracts may be a sufficient means of mitigating opportunism.

A few papers study the choice between contracts and other hybrids such as partial ownership agreements or equity linkages for conducting R&D. For example, Pisano (1990) argued that partial ownership dominates contracts under certain combination of asset specificity, uncertainty, the number of trading partners, and other variables. Oxley (1997 and 1999) showed that the choice

³⁸ See Ohanian (1994), and Laffontaine and Shaw (1999).

³⁹ Compare, e.g., Anderson (1985), John and Weitz (1988), Regan (1997), and Fein and Anderson (1997).

between contractual and equity-based vertical alliances is mainly determined by the costs of contracting.

Joskow (1987) focused on duration of agreements between coal suppliers and coal-burning electrical plants. He examined 277 contracts and found that they tended to be longer, *ceteris paribus*, when relationship-specific investments (in this case site specificity and dedicated assets; mine-mouth as a key independent variable) are at stake. Similar results for the natural gas industry were obtained by Crocker and Masten (1988).

Another investigated aspect of contracts is incompleteness. Crocker and Reynolds (1993) test the relationship between contractual incompleteness and the likelihood of opportunistic behaviour in the case of Air Force engine procurement. They conclude that contracts are more complete when contractor has a history of disputes with purchasers and less complete when there are increases in uncertainty that raise the cost of writing more complete contracts.

In a similar way, Saussier (2000) studied French electricity contracts and Crocker and Masten (1991) analysed incompleteness in natural-gas contracting. As before, the dependent variable is a qualitative measure of the extent to which renegotiation procedures are specified, and the independent variables are assets specificity, uncertainty, etc. The key findings are confirmed: assets specificity and site specificity are positively related to completeness, *ceteris paribus*.

There is also a large literature on other hybrid forms of organization, for example, sharing arrangements such as franchising and agricultural cropsharing, groups of firms organised as networks, clusters, or alliances, and reciprocal-trading arrangements.⁴⁰

5.4. Problems with empirical testing of the theory of the firm

The empirical literature on the make-or-buy decision, including the structure of long-term contracts and hybrid forms of organization, is mostly consistent with the transaction cost theory of the firm, i.e., vertical arrangements are a result of attempts to protect trading partners from hazards of exchange under incomplete contracting. Joskow (1991) pointed out that the literature on the make-or buy decision is in a better shape than much of the empirical work in industrial organization generally. Also Williamson (1996) admits that transaction cost economics is an empirical success story. However, there are still plenty of important issues to be resolved in this area. First, the measurement and definition of asset specificity, uncertainty, and other variables remain inconsistent. Second, many studies do not explicitly compare rival explanations for vertical relationships. Third, correlation between transactional attributes

⁴⁰ See Menard (2004) for a review of the literature on hybrids.

and governance structures is often mistaken for causality. Fourth, the legal and regulatory environments do not always get sufficient attention.

Asset specificity is difficult to measure consistently across industries; that is an important reason for so many single-industry studies of vertical boundaries relatively to cross-industry analyses. Further refinement concerning measurement is necessary. Proxies such as capital intensity or fixed costs are very imperfect and may not capture whether the investment has value outside the transaction for which it was initially made. Another issue is that asset-specificity effects may be confused with market power. Although specific investment may lead to bilateral monopoly, a small-numbers bargaining situation is not by itself evidence of relationship-specific investment. Also, uncertainty is not only hard to measure, but even difficult to define. The analysis is complicated by the fact that the effect of uncertainty depends on competitive conditions.

While the early studies on transaction cost theory emphasized asset specificity as the key to an efficiency explanation for vertical integration, some papers use transaction cost and incomplete contract theory to explain vertical integration without referring to asset specificity. For example Pirrong (1993) argued that long-term contracts and sometimes vertical integration can be efficient in the presence of smaller contracting hazards even without asset specificities. In a study of bulk shipping, he showed that more integrated governance structures can dominate spot trading in the presence of "temporal specificities".⁴¹ Martinez (2002) proved how temporal specificities lead to tight vertical coordination in poultry and egg production.

Several studies of the U.S. trucking market also found long-term contracts in the absence of asset specificity. For example Nickerson and Silverman (2003) showed that tight vertical control (use of company-owned trucks) results from the need for temporal coordination among hauls, and the shipper's desire to protect its brand-name capital. Lafontaine and Masten (2002) showed that the observed variation in contractual arrangements in trucking can best be explained by driver and truck heterogeneity.

None of the studies denies that asset specificities are important determinants of vertical integration; they only stress that in some cases vertical control can be explained without reference to them, or to holdup problems at all. By now, this line of analysis is rather an exception, but accumulation of such evidence could challenge the key structure of the transaction cost approach.

Empirical research on vertical boundaries is also subject to the general problem in this type of work, i.e., alternate hypotheses that could also fit the data are usually not stated. Undoubtedly, studies that explicitly compare competing hypotheses about contractual relationships are needed, because rival theories often posit mutually exclusive outcomes.

⁴¹ This concept was introduced by Masten, Meehan, and Snyder (1991).

For example, Spiller (1985) compares asset specificity and market power as two distinct explanations of vertical mergers. Using site specificity, defined as the proximity of the merging firms, to represent asset specificity, he studied the gains from merger due to unexpected changes in the firms' stock prices at the announcement of the merger. He found that the total gain from merger is smaller where site specificity is lower, but there is no significant relationship between industry concentration and the distance between merging firms. The above findings seem to support the asset-specificity explanation over market-power explanation of vertical integration.

Another comparative study was prepared by Poppo and Zenger (1995), who investigated transaction cost and resourced-based explanations for information-technology (IT) outsourcing. Using a survey of corporate IT managers to measure perceived satisfaction with IT services, they found that asset specificity is negatively related to the performance of market transactions, which is consistent with the transaction cost theory. Contrary to the resource-based approach, they showed that asset specificity does not improve the performance of in-house transactions.⁴²

Most of the comparative researches appear in the strategic management literature, where theories of the firm based on capabilities, power, and trust are important rivals to the transaction cost view. However, in industrial organization theories of vertical boundaries based on market power or technological foundations have not induced much research.

It is worth noting that the evidence presented here is often interpreted as supportive of transaction cost theory as stated by Williamson and Klein, Crawford, and Alchian, and the property rights approach (the formal version) associated with Grossman, Hart, and Moore. However, there are important differences between these two sets of explanations for firm boundaries.⁴³

The property-rights view is focused exclusively on *ex ante* underinvestment in relationship-specific human capital due to inefficient ownership arrangements, while transaction cost theories pay more attention to the *ex post* contract-execution stage.

A more general issue is that most of the empirical studies establish correlations, not causal relations, between asset specificity and internal governance. The researchers test a reduced-form model where the probability of observing a more hierarchical structure of governance increases with the degree of relation-specific investments. When the presence of such investments reduces the costs of internal governance, than the asset specificity could lead to integration, independent of holdup problem or other maladaptation costs.⁴⁴ Masten, Meehan, and Snyder (1991) attempted to distinguish between these two effects in the context of

⁴² Other studies include Poppo and Zenger (1998), Silverman (1999), and Nickerson, Hamilton, and Wada (2001).

⁴³ See Williamson (2000) for a detailed discussion.

⁴⁴ Compare Masten (1994).

human capital showing that the investing in specific human capital reduces the cost of internal organization more than it increases market governance costs. However, we do not have a general theory of how relationship-specific assets might reduce the costs of internal organization. By contrast, the underinvestment problem is fairly well understood, because it is much easier to deal with.

Empirical work on vertical boundaries could be improved by greater attention to the effects of the regulatory and legal environment. The literature has generally focused on relatively unregulated industries operating under a relatively stable regime.⁴⁵ However, differences in regulatory structures and their interpretation by courts can have substantial effects on the performance of alternative vertical arrangements. Neglecting these effects leads to biased estimates of the influence of other factors, such as assets specificity or uncertainty, on the decision about governance structure.

Henisz (2000) started research that focused more closely on the relationship between contractual hazards and political hazards. He showed, for example, that firms tend to prefer joint ventures with foreign partners rather than majority owned plants when political hazards are high, even though majority ownership may better mitigate the contractual hazards associated with asset specificity. The available contracting options may also be limited by regulation, or more generally, by the institutional environment.

Another important area of research is the evolution of contractual relations in rapidly changing environment, such as transition economies.⁴⁶ Comparative analysis, as well as the analysis of the adaptational abilities of various contractual arrangements can be conducted in these settings.

5.5. Opportunities for the future

Despite the problems with empirical work, the transaction cost theory of the firm was successful in explaining the vertical structure of the enterprise, although various researchers differ with respect to the actual level of consistency between the theory and the empirical results.

For example, Shelanski and Klein (1995) concluded that the empirical literature is "remarkably consistent" with the predictions of the theory. As mentioned earlier, a similar assessment was made also by Williamson (1996).

A systematic evaluation of 304 empirical tests reviewed by David and Han (2004) found that the conclusions about Williamson's transaction cost framework were "mixed"; i.e. overall support was at 47 percent even though the sampling was quite conservative.⁴⁷

⁴⁵ A notable exception is research on contracting among public utilities and their suppliers.

⁴⁶ See, e.g., Boger, Hobbs, and Kerr (2001).

⁴⁷ Another interesting study is Carter and Hodgson (2006).

The majority of researchers agree that asset specificity is an important determinant of vertical contractual relations, but it is not the only factor here. Results regarding the role of uncertainty are much less convincing; the majority of tests found uncertainty to be not a significant factor in explaining contractual choices.

Many important aspects of the transaction cost theory have not received much empirical attention at all, for example, frequency and performance. There are actually no tests checking whether hierarchies outperform markets when both asset specificity and uncertainty are high, or equivalently, whether markets perform better when these attributes are both low.

As mentioned earlier, there has been a significant amount of discrepancy and disagreement regarding the operationalization of core constructs and the interpretations of key relationships. More work in this respect would allow the theory to advance more consistently and convincingly across boundaries.

Another important point for future research is the avoidance of methodological pitfalls. Tests about the effects of governance structures on performance are problematic because they are likely to suffer from self-selection issues; unobserved variables may affect both organizational choice and performance, thus potentially biasing estimates. The researchers must be aware of such pitfalls and employ technologies that account for them.⁴⁸

The theory itself could be refined by specifying the conditions under which it works well and under which it does not. It is especially important given some mixed results of empirical tests.

It should also be noticed that empiricists have not taken sufficient advantage of the possibilities for longitudinal work in the theory of the firm. The transaction cost view of the firm can be tested through adjustments of organizational forms over time.

Future empirical research should be firmly grounded, in terms of substance, on core theoretical propositions. On the one hand, some propositions have been loosely interpreted, and on the other hand, some key variables have not received sufficient scrutiny. Explicit attention to the contingent nature of theoretical predictions is needed.

Most of the new theoretical work in economics of firm boundaries builds on the incomplete-contracting framework of Grossman, Hart, and Moore which is not identical with transaction-cost framework of Williamson and Klein, Crawford, and Alchian. The former does not lend itself to empirical testing as easily as the latter one.

Vertical relations are often subtle and complex. It is not enough to distinguish between "make" and "buy"; researchers must recognize a wide variety of contractual and organizational options available. This direction of analysis must be continued.

Comparisons of organizational arrangements across institutional environments may become the next growth area in the transaction cost literature.

⁴⁸ See for example, Poppo and Zenger (1998).

Bibliography

- Alchian, A.A. and H. Demsetz, 1972, *Production, Information Costs, and Economic Organization*, American Economic Review, 62, pp. 772–795.
- Anderson, E., 1985, *The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis*, Marketing Science, 4, pp. 234–254.
- Anderson, E., and D.C. Schmittlein, 1984, *Integration of the Sales Force: An Empirical Examination*, Rand Journal of Economics, 15, pp. 385–395.
- Bajari, P., and S. Tadelis, 2001, *Incentives versus Transaction Costs: A Theory of Procurement Contracts*, Rand Journal of Economics, 32, pp. 387–407.
- Baker, G., R. Gibbons, and K.J. Murphy, 2002, *Relational Contracts and the Theory of the Firm*, Quarterly Journal of Economics, 117, pp. 39–84.
- Barzel, Y., 1997, *Economic Analysis of Property Rights*, 2nd ed., Cambridge: Cambridge University Press.
- Bigelow, L., 2001, *Efficient Alignment and Survival in the U.S. Automobile Industry*, Working Paper, Olin School of Business, Washington, St. Louis.
- Bindseil, U., 1997, *Vertical Integration in the Long Run: The Provision of Physical Assets to the London and New York Stock Exchanges*, Journal of Institutional and Theoretical Economics, 153, pp. 641–656.
- Boger, S., J.E. Hobbs, and W. A. Kerr, 2001, *Supply Chain Relationships in the Polish Pork Sector*, Supply Chain Management, 6, pp. 74–82.
- Carter, R. and G.M. Hodgson, 2006, *The Impact of Empirical Tests of Transaction Cost Economics on the Debate on the Nature of the Firm*, Strategic Management Journal 27, no. 5 (May), pp. 461–76.
- Cheung, S., 1983, *The Contractual Nature of the Firm*, Journal of Law and Economics, 26, pp. 1–22.
- Coase, R.H., 1937, *The Nature of the Firm*, Economica, 4, pp. 386–405.
- Coase, R.H., 1991, *The Nature of the Firm: Origin, Meaning, Influence*, in O. E. Williamson and S. G. Winter, eds. *The Nature of the Firm*, Oxford: Basil Blackwell.
- Conlisk, J., 1996, *Why Bounded Rationality?* Journal of Economic Literature, 34, pp. 669–700.
- Crocker, K.J., and K.J. Reynolds, 1993, *The Efficiency of Incomplete Contracts: An Empirical Analysis of Air Force Engine Procurement*, Rand Journal of Economics, 24, pp. 126–146.
- Crocker, K.J., and S.E. Masten, 1988, *Mitigating Contractual Hazards: Unilateral Options and Contract Length*, Rand Journal of Economics, 19, pp. 327–343.
- Crocker, K.J., and S.E. Masten, 1991, *Pretia Ex Machina? Prices and Process in Long-Term Contracts*, Journal of Law and Economics, 24, pp. 69–99.

David, R.J. and S.-K. Han, 2004, *A Systematic Assessment of the Empirical Support for Transaction Cost Economics*, *Strategic Management Journal*, vol. 25, no. 1, pp. 39–58.

Demsetz, H., 1988, *The Theory of the Firm Revisited*, *Journal of Law, Economics, and Organization*, 4(1), pp. 141–161.

Fama, E., 1980, *Agency Problems and the Theory of the Firm*, *Journal of Political Economy*, 88, pp. 288–307.

Farrell, J., and N. Gallini, 1986, *Second-Sourcing as a Commitment: Monopoly Incentives to Attract Competition*, Working Paper 8618, University of California, Berkeley.

Fein, A.J., and E. Anderson, 1997, *Patterns of Credible Commitments: Territory and Brand Selectivity in Industrial Distribution Channels*, *Journal of Marketing*, 61, pp. 19–34.

Foss, N.J., 2000, *The Theory of the Firm: an Introduction to Themes and Contributions*, in idem. *The Theory of the Firm: Critical Perspectives in Economic Organization*, London: Routledge.

Freeland, R.F., 2000, *Creating Holdup Through Vertical Integration: Fisher Body Revisited*, *Journal of Law and Economics*, 43, pp. 33–66.

Gallick, E.C., 1984, *Exclusive dealing and Vertical Integration: The Efficiency of Contracts in the Tuna Industry*, Federal Trade Commission Bureau of Economics Staff Report, Washington, D.C.: Federal Trade Commission.

Gibbons, R., 2004, *Four Formal(izable) Theories of the Firm?* Working Paper, MIT, September.

Globerman, S., and R. Schwindt, 1986, *The Organization of Vertically Related Transactions in the Canadian Forest Products Industries*, *Journal of Economic Behavior and Organization*, 7, pp. 199–212.

Goldberg, V., and J.R. Erickson, 1987, *Quantity and Price Adjustment in Long-Term Contracts: A Case Study of Petroleum Coke*, *Journal of Law and Economics*, 30, pp. 369–398.

Gonzalez-Diaz, M., B. Arrunada, and A. Fernandez, 2000, *Causes of Subcontracting: Evidence from Panel Data on Construction Firms*, *Journal of Economic Behavior and Organization*, 42, pp. 167–187.

Green, J. and N. Stokey, 1983, *A Comparison of Tournaments and Contests*, *Journal of Political Economy*, 91, pp. 349–364.

Grossman, S. and O. Hart, 1980, *Takeover Bids, the Free Rider Problem and the Theory of the Corporation*, *Bell Journal of Economics*, 11, pp. 42–64.

Grossman, S. and O. Hart, 1986, *The Costs and Benefits of Ownership: A Theory of Vertical Integration*, *Journal of Political Economy*, 94, pp. 691–719.

Hallwood, P. C., 1991, *On Choosing Organizational Arrangements: The Example of Offshore Oil Gathering*, *Scottish Journal of Political Economy*, 38, pp. 227–241.

Harrington, J. and J. Prokop, 1993, *The Dynamics of the Free-Rider Problem in Takeovers*, *Review of Financial Studies*, 6, pp. 851–882.

- Hart, O., 1995, *Firms, Contracts and Financial Structure*, Oxford: Clarendon Press.
- Hart, O., 1989, *An Economist's Perspective on the Theory of the Firm*, *Columbia Law Review*, 89(7), pp. 1757–1774.
- Hart, O., and B. Holmström, 1987, *The Theory of Contracts*, in T. F. Bewley, ed., *Advances in Economic Theory. Fifth World Congress*, Cambridge: Cambridge University Press.
- Hart, O., and J. Moore, 1990, *Property rights and the Nature of the Firm*, *Journal of Political Economy*, 98, pp. 1119–1158.
- Helper, S., J. P. MacDuffie, and C. F. Sabel, 2000, *Pragmatic Collaborations: Advancing Knowledge While Controlling Opportunism*, *Industrial and Corporate Change*, 9, pp. 443–483.
- Henisz, W. J., 2000, *The Institutional Environment for Multinational Investment*, *Journal of Law, Economics and Organization*, 16, pp. 334–364.
- Hennart, J.-F., *Upstream Vertical Integration in the Aluminum and Tin Industries*, *Journal of Economic Behavior and Organization*, 9, pp. 281–299.
- Holmström, B., 1982, *Managerial Incentive Problems: A Dynamic Perspective*, in *Essays in Economics and Management in Honor of Lars Wahlbeck*, Helsinki: Swedish School of Economics.
- Holmström, B. and J. Tirole, 1989, *The Theory of the Firm*, in R. Schmalensee and R. D. Willig, eds., *Handbook of Industrial Organization*, Vol. 1, Amsterdam: North-Holland.
- Hubbard, T.N., 1999, *How Wide is the Scope of Hold-Up Based Theories of Governance? Shipper-Carrier Relations in Trucking*, Working Paper, Department of Economics, 34, pp. 25–67.
- Jensen, M. and W. Meckling, 1976, *The Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, *Journal of Financial Economics*, 3, pp. 305–360.
- John, G., and B. A. Weitz, 1988, *Forward Integration into Distribution: An Empirical Test of Transaction Cost Analysis*, *Journal of Law, Economics and Organization*, 4, pp. 337–355.
- Joskow, P., 1985, *Vertical Integration and Long Term Contracts: The Case of Coal-Burning Electric Generating Plants*, *Journal of Law, Economics and Organization*, 1, pp. 33–79.
- Joskow, P., 1987, *Contract Duration and Relationship-Specific Investments: The Case of Coal*, *American Economic Review*, 77, pp. 168–185.
- Joskow, P., 1988, *Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence*, *Journal of Law, Economics and Organization*, 4, pp. 95–117.
- Joskow, P. and R. Schmalensee, 1986, *Incentive Regulation for Electric Utilities*, *Yale Journal on Regulation*, 4, pp. 1–49.
- Kihlström, R.E. and J.-J. Laffont, 1979, *A General Equilibrium Entrepreneurial Theory of Firm Formation based on Risk Aversion*, *Journal of Law, Economics, and Organization*, 4, pp. 199–213.

Klein, P.G., 1988, *Vertical Integration as Organized Ownership: The Fisher Body-General Motors Relationship Revisited*, *Journal of Law, Economics, and Organization*, 4, pp. 199–213.

Klein, P.G., 2005, *The Make-or-Buy Decision: Lessons from Empirical Studies*, in C. Ménard and M. Shirley, eds. *Handbook of New Institutional Economics*, Springer, pp. 435–64.

Klein, B., R.A. Crawford, and A. A. Alchian, 1978, *Vertical Integration, Appropriable Rents, and the Competitive Contracting Process*, *Journal of Law and Economics*, 21, pp. 297–326.

Knight, F.H., 1921, *Risk, Uncertainty, and Profit*, New York: Augustus M. Kelley.

Kreps, D., *A course in Microeconomic Theory*, Princeton University Press, chapters: 16, 19, 20.

Laffontaine, F., and K. Shaw, 1999, *The Dynamics of Franchise Contracting: Evidence from Panel Data*, *Journal of Political Economy*, 107, pp. 1041–1080.

Langlois, R.N. and M. Cosgel, 1993, *Frank Knight on Risk, Uncertainty and the Firm: a New Interpretation*, *Economic Enquiry*, 31, pp. 456–465.

Langlois, R.N. and N. J. Foss, 1998, *Capabilities and Governance: the Rebirth of Production in the Theory of the Firm*, *KYKLOS*, 52, pp. 201–218.

Lazear, E. and S. Rosen, 1981, *Rank-Order Tournaments as Optimal Labor Contracts*, *Journal of Political Economy*, 89, pp. 841–864.

Levy, D., 1985, *The Transaction Cost Approach to Vertical Integration: An Empirical Examination*, *Review of Economics and Statistics*, 67, pp. 438–445.

Lieberman, M., B., 1991, *Determinants of Vertical Integration: An Empirical Test*, *Journal of Industrial Economics*, 39, pp. 451–466.

Lyons, B.R., 1995, *Specific Investment, Economies of Scale, and the Make-or-Buy Decision: A Test of Transaction Cost Theory*, *Journal of Economic Behavior and Organization*, 26, pp. 431–443.

MaCaulay, S., 1963, *Non-Contractual Relations in Business*, *American Sociological Review*, 28, pp. 55–70.

Macher, J. T., 2001, *Vertical Disintegration and Process Innovation in Semiconductor Manufacturing: Foundries vs. Integrated Device Manufacturers*, Working Paper, McDonough School of Business, Georgetown University.

Mahoney, J., 2005, *Economic Foundations of Strategy*, London: Sage.

Marschak, T., 1965, *Economic Theories of Organization: The Role of the Firm Reconsidered*, in J. G. March, ed., *The Handbook of Organizations*, Chicago: Rand McNally and Co.

Marschak, J., and R. Radner, 1972, *The Theory of Teams*, New Haven: Yale University Press.

Martinez, S., 2002, *A Comparison of Vertical Coordination in the U.S. Poultry, Egg, and Pork Industries*, *Current Issues in Economics of Food Markets, Agriculture Information Bulletin No. 747–05*, U.S. Department of Agriculture, Economic Research Service.

- Masten, S.E., 1984, *The Organization of Production: Evidence from the Aerospace Industry*, *Journal of Law and Economics*, 27, pp. 403–417.
- Masten, S.E., J. W. Meehan, and E. A. Snyder, 1991, *The Costs of Organization*, *Journal of Law, Economics and Organization*, 7, pp. 1–25.
- Masten, S.E., 1994, *Empirical Research in Transaction-cost Economics: Challenges, Progress, Directions*, Mimeo, University of Michigan Business School.
- Masters, J.K. and G. Miles, 2002, *Predicting the Use of External Labor Arrangements: A Test of the Transaction Cost Perspective*, *Academy of Management Journal* 45, pp. 431–442.
- Ménard, C., 1996, *On Clusters, Hybrids and other Strange Forms. The Case of the French Poultry Industry*, *Journal of Institutional and Theoretical Economics*, 152, pp. 154–183.
- Ménard, C., 2004, *The Economics of Hybrid Organizations*, *Journal of Institutional and Theoretical Economics*, 160, pp. 345–376.
- Milgrom, P.J., and J. D. Roberts, 1988, *Economic Theories of the Firm: Past, Present, and Future*, *Canadian Journal of Economics* 21, pp. 444–458.
- Monteverde, K. and D.J. Teece, 1982a, *Appropriable Rents and Quasi-Vertical Integration*, *Journal of Law and Economics*, 25, pp. 321–328.
- Monteverde, K. and D.J. Teece, 1982b, *Supplier Switching Costs and Vertical Integration in the Automobile Industry*, *Bell Journal of Economics*, 13, pp. 206–213.
- Mookherjee, D., 1984, *Optimal Incentive Schemes with Many Agents*, *Review of Economic Studies*, 51, pp. 433–446.
- Muris, T.J., D. Scheffman, and P. T. Spiller, 1992, *Strategy and Transaction Costs: The Organization of Distribution in the Carbonated Soft Drink Industry*, *Journal of Economics and Management Strategy*, 1, pp. 83–128.
- Nalebuff, B. and J. Stiglitz, 1983, *Prices and Incentives: Towards a General Theory of Compensation and Competition*, *Bell Journal of Economics*, 14, pp. 21–43.
- Nelson, R. and S. Winter, 1973, *Toward an Evolutionary View of Economic Capabilities*, *American Economic Review*, 63, pp. 440–449.
- Nelson, R. and S. Winter, 1982, *An Evolutionary Theory of Economic Change*, Cambridge, Mass.: Harvard University Press.
- Nickerson, J.A., B.H. Hamilton, and T. Wada, 2001, *Market Position, Resource Profile, and Governance: Linking Porter and Williamson in the Context of International Courier and Small Package Services in Japan*, *Strategic Management Journal*, 22, pp. 251–273.
- Ohanian, N.K., 1994, *Vertical Integration in the U.S. Pulp and Paper Industry, 1900-1940*, *Review of Economics and Statistics*, 74, pp. 202–207.
- Oxley, J.E., 1997, *Appropriability Hazards and Governance in Strategic Alliances: A Transaction Cost Approach*, *Journal of Law, Economics, and Organization*, 13, pp. 387–409.

Oxley, J.E., 1999, *Institutional Environment and the Mechanisms of Governance: The Impact of Intellectual Property Protection on the Structure of Inter-firm Alliances*, *Journal of Economic Behavior and Organization*, 38, pp. 283–309.

Palay, T.M., 1984, *Comparative Institutional Economics: The Governance of Rail Freight Contracting*, *Journal of Legal Studies* 13, pp. 265–287.

Pirrong, S.C., 1993, *Contracting Practices in Bulk Shipping Markets: A Transaction Cost Explanation*, *Journal of Law and Economics*, 36, pp. 937–976.

Pisano, G.P., 1990, *Using Equity Participation to Support Exchange: Evidence from the Biotechnology Industry*, *Journal of Law, Economics and Organization*, 5, pp. 109–126.

Poppo, L., and T. Zenger, 1998, *Testing Alternative Theories of the Firm: Transaction Cost, Knowledge-Based, and Measurement Explanations for Make-or-Buy Decisions in Information Services*, *Strategic Management Journal*, 19, pp. 853–877.

Prokop, J., 2001, *Przyczynek do teorii przejmowania kontroli nad spółkami akcyjnymi*, Monografie i Opracowania, Warszawa, Oficyna wydawnicza SGH.

Regan, L., 1997, *Vertical Integration in the Property-Liability Insurance Industry: A Transaction Cost Approach*, *Journal of Risk and Insurance*, 64, pp. 41–62.

Ross, S., 1973, *The Economic Theory of Agency: The Principal's Problem*, *American Economic Review*, 63, pp. 134–139.

Sampson, R.C., 2001, *The Cost of Inappropriate Governance in R&D Alliances*, Working Paper, Stern School of Business, New York University.

Saussier, S., 2000, *Transaction Costs and Contractual Incompleteness: The Case of Electricite de France*, *Journal of Economic Behavior and Organization*, 42, pp. 189–206.

Scherer, F., 1980, *Industrial Market Structure and Economic Performance*, second edition, Chicago: Rand-McNally.

Shelanski, H.A., and P. G. Klein, 1995, *Empirical Research in Transaction Cost Economics: A Review and Assessment*, *Journal of Law, Economics and Organization*, 11, pp. 335–361.

Shepard, A., 1986, *Licensing to Enhance Demand for New Technologies*, Mimeo, Yale University.

Shleifer, A., 1985, *A Theory of Yardstick Competition*, *Rand Journal of Economics*, 16, pp. 319–327.

Silverman, B.S., 1999, *Technological Resources and the Direction of Corporate Diversification: Alignment, and Organizational Mortality in the U.S. Trucking Industry*, *Strategic Management Journal*, 18, pp. 31–52.

Spiller, P., 1985, *On Vertical Mergers*, *Journal of Law, Economics and Organization*, 1, pp. 285–312.

Stuckey, J., 1983, *Vertical Integration and Joint Ventures in the Aluminium Industry*, Cambridge, Mass.: Harvard University Press.

Tadelis, S., 2002, *Complexity, Flexibility and the Make-or-Buy Decision*, American

Economic Review, vol. 92, no. 2, pp. 433–37.

Tirole, J., 1986, *Hierarchies and Bureaucracies*, Journal of Law, Economics and Organization, 2, pp. 181–214.

Tirole, J., 1997, *The Theory of Industrial Organization*, The MIT Press.

Tirole, J., 2006, *The Theory of Corporate Finance*, Princeton University Press, Princeton and Oxford.

Ulset, S., 1996, *R&D Outsourcing and Contractual Governance: An Empirical Study of Commercial R&D Projects*, Journal of Economic Behavior and Organization, 30, pp. 63–82.

Walker, G., and D. Weber, 1984, *A Transaction Cost Approach to Make-or-Buy Decisions*, Administrative Science Quarterly, 29, pp. 373–391.

Weiss, A.M., and N. Kurland, 1997, *Holding Distribution Channel Relationships Together: The Role of Transaction-Specific Assets and Length of Prior Relationship*, Organization Science, 8, pp. 612–623.

Williamson, O.E., 1964, *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm*, Englewood Cliffs: Prentice-Hall.

Williamson, O.E., 1970, *Corporate Control and Business Behavior*, Englewood Cliffs: Prentice-Hall.

Williamson, O.E., 1971, *The Vertical Integration of Production: Market Failure Considerations*, American Economic Review, 61, pp. 112–123.

Williamson, O.E., 1975, *Markets and Hierarchies*, New York: Free Press.

Williamson, O.E., 1985, *The Economic Institutions of Capitalism*, New York: Free Press.

Williamson, O.E., 1991, *Comparative Economic Organization: The Analysis of Discrete Structural Alternatives*, Administrative Science Quarterly, 36, pp. 269–296.

Williamson, O.E., 1996, *The Mechanisms of Governance*, Oxford: Oxford University Press.

Williamson, O.E., 1998, *Strategy Research: Governance and Competence Perspectives*, Mimeo.

Williamson, O.E., 2002, *The Theory of the Firm as Governance Structure: From Choice to Contract*, Journal of Economic Perspectives, vol. 16, no. 3, pp. 171–95.

Williamson, O.E. and S. Masten, eds., 1995, *Transaction Cost Economics*, Aldershot: Edward Elgar.

Winter, S.G., 1988, *On Coase, Competence, and the Corporation*, in O. E. Williamson and S. Winter, eds. *The Nature of the Firm*, Oxford: Basil Blackwell.

Yrande-Billon, A., 2004, *Contractual Choices and Performances: Evidence from the British railways*, in G. Hendrikse, J. Windsperger, G. Cliquet, and M. Tuunanen, eds., *Economics and Management of Franchising Networks*, Heidelberg: Physica/Springer.