Notes to chapter 1

* Paper prepared for Liberty Fund Symposium on 'German Ordnungstheorie and American Constitutional Economics' Bonn, 3-6 June 1989.

** James M. Buchanan was born in Murfreesboro, Tennessee, October 3, 1919. Educational background: B.A., Middle Tennessee State College (1940); M.S., University of Tennessee (1941); Ph.D., University of Chicago (1948). Is currently Advisory General Director, Center for Study of Public Choice; Harris University Professor, George Mason University, Fairfax, Virginia and holder of the Nobel Prize in Economic Sciences, 1986. The following is a short list of his most well-known major works: *Ethics and Economic Progress* (1994), *The Return to Increasing Returns* (with Yong J. Yoon (1994)); *Better than Plowing: And Other Personal Essays* (1992); *The Economics and the Ethics of Constitutional Order* (1991); *Constitutional Economics* (1991); *Explorations into Constitutional Economics* (1989); *Economics: Between Predictive Science and Moral Philosophy* (1987); *The Reason of Rules* (with Geoffrey Brennan, (1985)); *Freedom in Constitutional Contract* (1978); *The Limits of Liberty* (1975); *The Calculus of Consent* (with Gordon Tullock, (1962)).

CHAPTER 2
TRANSACTION COST ECONOMICS*

*Trond Petersen***

2.1 Introduction

The central objective of the research programme known as Transaction Cost Economics (henceforth TCE) is to explain why some transactions are done in markets whereas other are done within firms. This is done by analyzing the relative costs of the two ways of executing transactions. Below I discuss the TCE framework. I focus on the part of TCE that has become associated with Williamson (1975, 1985).

In TCE one distinguishes between two broad classes of exchanges or transactions: market transactions and hierarchical transactions. In market exchanges, also referred to as contractual *exchanges,* everything relevant to the transaction or exchange is agreed upon before the exchange occurs. Every contingency is spelled out in the contract. There is no uncertainty, no scope for renegotiation, etc.

In hierarchies or hierarchical exchanges, in contrast, not everything relevant to the exchange is spelled out in advance. Instead, one leaves scope for unforeseen events to occur, for renegotiation, and for bargaining. There will typically also be authority systems where someone has the right to make decisions and impose these on others as events unfold.

The concept of transaction costs (TC) refers to the costs associated with executing market or contractual exchanges. These are the costs of writing the initial contract, which may be high in complex situations with lots of uncertainty, and the costs of bargaining and renegotiating contracts that are under execution. Take the example of an automobile producer contracting with a supplier for some custommade part to be used in automobiles. Producing such parts often requires development. Unforeseen events may then occur, which in turn may require renegotiation of the contract. So, the transaction costs will here be first in terms of writing the initial contract, which tries to foresee later complications, as well as possible later renegotiations of the contract. When the transaction costs of a market exchange are high, alternative exchange structures might be sought, in particular those of a hierarchy.

The TCE framework tries to explain why some economic transactions are organized in markets whereas others in hierarchies or within firms. It attempts to explain the various governance structures – market, hierarchy, long-term contract – under which transactions are carried out. Some of the specific institutional arrangements it purports to explain are internal labour markets in firms, vertical integration between firms, the growth of larger firms, and the emergence of the multi-divisional form for firms. In the automobile example above, a market relationship exists if the automobile manufacturer buys the parts from an independent supplier. A hierarchy comes into existence if the automobile manufacturer buys up the supplier and then produces the parts inhouse.

The claim made in TCE is that transactions often will be carried out in hierarchies, that is, under an authority system rather than under a fully specified contract, when there are high costs of carrying out the transactions as regular contractual exchanges.

Williamson's framework has several antecedents (see, e.g., Williamson, 1981, pp. 550–552, 1985, ch. 1). The idea of transaction costs itself originated with Coase ([1937] 1988). He attempted to explain first why not all transactions are carried out in markets. He did so by identifying the costs of executing market transactions. Having done that, Coase ([1937] 1988) next addressed why not all economic transactions are executed within a single large firm. He identified three factors, two of which were decreasing returns to the managerial function with size of firm and that managers of large firms might not use resources in the most productive ways, which is loss in incentives in large firms. Coase (1987) gives an interesting overview of ideas, their origin, meaning, and impact, as well as several comments on his own intellectual and professional development and the social and intellectual context in which the ideas arose.

The 'Carnegie School' with March and Simon (1958) and Cyert and March (1963) was another antecedent. Here, the idea of bounded rationality originated, a concept that is distinct from the more controversial concept of 'satisficing,' which means that an agent tries to find a satisfactory solution, not necessarily the one thought best. Cyert and March's work on the 'behavioural theory of the firm' aimed at getting closer to the actual decision processes managers took part in, and the assumption of boundedly rational behaviour was claimed to provide a better description of actual behaviour than more standard economic theory.

The legal literature on contract law was a third antecedent (see, e.g., Macneil, 1974). Here, one finds the distinction between classical contract law versus neoclassical contract law and excuse doctrine. Classical contract law applies to transactions that are quite formal and rule governed, where there is no room for renegotiation when unanticipated events occur, where the identities of the parties do not matter. Neoclassical contract law applies to transactions where the legalistic features are less prevalent, where the parties are quite dependent upon each other and where litigation in courts would be costly, as it could ruin relationships built up over a longer period. Some of this is quite well known in the sociology of law literature (Macaulay, 1963; Macneil, 1974).

The remainder of the paper is organized in five sections. Section 2.2 outlines the basic ideas. Section 2.3 applies those ideas to three specific substantive domains. Section 2.4 discusses some empirical studies based on ideas from TCE in each of the three substantive domains discussed in Section 2.3. Section 2.5 discusses extensions and criticisms of TCE. Section 2.6 concludes the paper. Two appendices elaborate on some issues. My viewpoint is that of an organizational sociologist, and, when relevant, I use that viewpoint to evaluate some of the ideas.[1]

2.2 Elements of the theory

The basic terms

I concentrate on the revised version of TCE put forth in Williamson (1979, 1981, 1985).[2] In explicating the basic terms of the theory, Williamson (e.g., 1981, pp. 554–55) distinguishes between behavioural assumptions on the one hand and characteristics of transactions on the other. The latter are referred to as dimensionalizing. One could say that the behavioural assumptions deal with the psychology of the relevant actors, whereas the characteristics of the transactions analyzed are more akin to the technology faced in various transactions.

There are two behavioural assumptions. *First,* Williamson operates with the assumption of *bounded rationality.* This refers to the inability of human actors to compute correctly, to gather the correct information, to express themselves correctly, etc. Williamson, following Herbert Simon (e.g., 1976, [1945]), says that human behaviour is intendedly rational, but only limitedly so.

Second, Williamson operates with the assumption of *opportunism.* This refers to the claim that humans act not only upon self-interest, they also act with guile. They strategically misrepresent information and intentions in order to attempt securing a better outcome for themselves.

Regarding the characteristics of the transactions, three characteristics are in focus. *First,* Williamson stresses that *uncertainty* is important: one does not always know which state of the world has obtained at a given point in time; one does not know which states will obtain in the future; one is partially ignorant about own needs; one is partially ignorant about costs of doing something; and one is uncertain about the preferences, information, and goals of other actors.

The *second* characteristic of transactions is the *frequency* with which they recur. At the one extreme one has a one-shot deal, where one transacts only once. At the other extreme one has a transaction that occurs frequently and regularly, say, once a week.

The *third* characteristic of transactions is called *asset specificity*. It refers to the degree to which durable, transaction-specific investments are required to realize least cost supply. For example some firms might invest in training of employees that can be used for producing only specific goods, training (i.e., investments) with a low value were one to cease producing those goods.

Of the five terms presented above, that of asset specificity is especially important in Williamson's analysis and requires some elaboration. It can be described as follows. Suppose a principal contracts with an agent. The agent then makes an investment. This investment is an asset. If the use of the asset has high value within the relationship between the principal and the agent, but only a low value outside the relationship, then that asset is specific to the relationship. As Williamson (1981, p. 555) makes clear, asset specificity is both the most important dimension for describing transactions and the most neglected in previous studies.

Asset specificity can arise in one of four ways (Williamson, 1985, p. 55): *site* specificity, *physical asset* specificity, *human asset* specificity, and *dedicated asset* specificity. Site specificity occurs, for example, when a supplier locates her firm close to the buyer's, so as to minimize transportation costs. The value of the site is much higher within the relationship than outside. Physical asset specificity occurs when investments are made in machines producing specialized products for a customer that cannot easily be used for producing for other potential customers. Human asset specificity arises from learning by doing. For example, operators of machines may have knowledge about idiosyncrasies of the machines in the factory they work that are not in the operating manuals for the machine but that are essential for operating the machines well. Dedicated assets arise when general investments are made that would not take place but for the prospect of selling a significant amount to a particular customer. It usually involves expanding an existing plant on behalf of a particular buyer, that is, to increase capacity of production (see Williamson, 1985, p. 96). If the contract is terminated prematurely, it would leave the supplier with significant excess capacity (e.g., Joskow, 1985, p. 38).

Effects of basic terms

Using the basic terms – the two behavioural assumptions and the three characteristics of transactions – Williamson derives what one somewhat simplistically could call a 'simple causal chain' model. The exogenous or explanatory variables in the chain are the basic terms explained above. In the first step in the chain, these are used to derive the concept of transaction costs, which then becomes the first dependent variable. In the second step in the chain, the concept of transaction costs is turned into an explanatory or independent variable, which is used, then, to explain the second or most important dependent variable in the framework. That variable is the governance structure for transactions, markets versus hierarchies. In this subsection I outline how the first dependent variable in the chain, transaction costs, gets derived from the basic terms.[3]

I start by considering the first characteristic of transactions, uncertainty. Under uncertainty about future conditions one needs, in specifying contracts, to anticipate as many of the possible future conditions as possible. For those conditions that can be specified in advance, one can then write a contingent contract: For each state of affairs that might realize in the future, a separate contractual outcome is specified. So uncertainty by itself is not a great problem.

When bounded rationality is added to uncertainty, things become more difficult. Actors that are boundedly rational, will not be able to deal explicitly, at the time of entering a contract, with all the contractually relevant aspects. They cannot foresee all possible contingencies that might arise in the future, and even if they could, they need not be able to stipulate how to deal with these. As a consequence, therefore, incomplete contracting is the best that can be achieved. Nevertheless, uncertainty and bounded rationality, which lead to incomplete contracts, need not lead to breakdown or problems for contracting (or market-type transactions). As long as the parties are fully trustworthy, contracting would still be feasible. Fully trustworthy parties could promise to behave responsibly and in good faith when unanticipated events occur (see Dore, 1986, 1987).[4]

When opportunism is added to uncertainty and bounded rationality, contracting becomes one step more difficult. Then, when unanticipated consequences occur, the actors may misrepresent information and interests. They will act self-interestedly with guile. They may attempt to renegotiate the contract to their own advantage, and in doing so, they may attempt to take unreasonable or unfair advantage of the situation

But, even if opportunism creates problems when unanticipated events occur during the execution of a contract (e.g., contract to build a weapons system, contract to document the level of living in a country, etc.), it need not make execution of the contract impossible. If the parties costlessly can withdraw from the contract and reenter it with some other party, opportunism will not create severe problems.

It is at this point that the concept of asset specificity, that is, the degree to which there are relationship-specific investments, becomes important. The idea is this. During the execution of the contract, relationship-specific investments are often made. These are investments that have a much higher value within the relationship than outside. So, if one party were to breach, that is, to withdraw from the contract, the value of the relationship-specific investments would fall. There are therefore so-called lock-in effects, where much can be lost if the relationship dissolves. Note that the investments can be physical, such as in plants, machines, etc., but they can also be site specific (as when a certain location has been chosen), as well as in human assets, when, for example, some people have been trained for tasks that have a high value in the relationship but not outside it. Moreover, some investments are less tangible, but still important, such as specialized language, which facilitates communication economies, and institutional and personal trust, which alleviates opportunism (Williamson, 1985, p. 62).

With such lock-in effects, which appear in the presence of asset specificity or relationship-specific investments, the problem of opportunism arises with full force when unanticipated events occur. There is an incentive for either party to engage in bargaining to redesign the contract in their best interest to deal with the unanticipated events. This is sometimes referred to as small-numbers bargaining.

To summarize, uncertainty and bounded rationality lead to incomplete contracts. Opportunism under incomplete contracts makes contract execution between the two parties difficult when unanticipated events occur. However, if the parties costlessly can recontract with other parties, opportunism need not pose a big problem, because each party can just find some different partner to complete the contract. In most cases there are relationship-specific investments, called asset specificity. These have a much higher value within the relationship than outside. Therefore, the parties rarely can costlessly recontract with other parties. This opens up the possibility of small-numbers bargaining: Each party will try to bargain so as to better their own position when the unanticipated events occur. Such bargaining is costly. Contractual exchanges are therefore susceptible to incurring such additional renegotiation and bargaining costs when unanticipated events occur.

Solutions to the problem of transaction costs

The analysis above yielded the first step in a causal chain, where the first dependent variable was the transaction costs of market exchanges. Below I outline the solution to the problem of transaction costs, namely the second step in the causal chain, where the central dependent variable is the governance form for transactions, markets versus hierarchies. The governance form is explained by the amount of potential or anticipated transaction costs.

As detailed above, uncertainty, bounded rationality, opportunism, and asset specificity lead to high costs of conducting market transactions (i.e., contractual exchanges), called transaction costs. These are due to the cost of writing contracts themselves, which are likely to be complex when there is a vast array of possible contingencies that may arise in the future. Therefore, contracts are usually incomplete. When contracts are incomplete, the additional problem of small numbers bargaining arise when unanticipated events occur. Then one incurs the bargaining and renegotiation costs.

The question posed, then, is: When it is difficult to write contracts and there are dangers of small-numbers bargaining, should one rely on a contractual exchange or should one seek some other governance structures for the transaction? The alternative governance structure is customarily a hierarchy: One party gives up their autonomy and agrees, within limits, to follow the instructions of the other party, also when unanticipated events occur. Exchanges or transactions do no longer occur according to prespecified contractual rules. Instead, some party has the authority to impose decisions and other parties are obliged to obey these within reasonable limits. In this manner one can deal with uncertainty as it occurs along the road. Opportunism becomes less important because transactions in hierarchies are repeated. Experience-ratings develop and opportunistic parties can be penalized and non-opportunistic parties can be rewarded.

In order to answer the question about the choice between markets and hierarchies, one needs to know first the costs of writing the contracts for the contractual exchanges, as well as the potential costs of resolving small-numbers bargaining when unanticipated events occur. Second, one needs to know the savings of the alternative governance structure, hierarchy. To assess the savings, one needs to consider the costs of the hierarchy.

There are at least two costs of a hierarchy. First there are the costs of setting it up, rather than relying on a market-type transaction. At this point, the remaining characteristic of the transaction enters, namely, the frequency with which it occurs. The more frequently the transaction is to be made in the future, the lower the setting-up costs will be per transaction: There are more transactions over which to spread the costs of setting up the hierarchy. So, high frequency of transaction tends to favour hierarchies when transaction costs are high. Second, incentives are weakened under a hierarchy. The latter may lead to inefficiencies. One must, therefore, trade off the savings in transaction costs for the lowered incentives when hierarchies are chosen.

The two extreme solutions, then, are markets (i.e., contractual exchanges) and hierarchies (i.e., non-contractual exchanges, based on authority systems, etc.). Intermediate governance forms are mixtures of markets and hierarchies, as described by Williamson (e.g., 1991).

Examples of these kinds of situations are many. The marriage contract is one. There is great uncertainty about future conditions. There is clearly small numbers bargaining. Bounded rationality is present. There may also be opportunism. Typically, marriage contracts are not very specific. Another example is the employment relationship. Bounded rationality is present. Opportunism is present. Employers and employees may both misrepresent information and intentions. There is clearly uncertainty, both about present and future conditions. Finally, there is small-numbers bargaining. Some employees need not be easy to replace. In some geographic locations, employees can choose only between a few employers, which put the latter in strong positions. Therefore, the claim is, authority systems rather than contracts prevail.

2.3 Applications of theory

Application 1: work organization and hierarchies

The treatment of work organization in Williamson (1985, ch. 10; also, 1981, pp. 562-566) keeps bounded rationality, opportunism, and frequency of transaction constant, and to some extent also uncertainty, but varies two factors: (1) human asset specificity and (2) the degree to which performance or productivity can be measured, which is the internal organizational counterpart to uncertainty (Williamson, 1981, p. 564). Assuming only two possible values on each dimension, low and high, a fourfold classification of governance structures for employment relationships is proposed.[5]

Internal spot market occurs when there is low human asset specificity and a high degree of measurement of performance. In such a case there is no reason for the employer or the worker to maintain the relationship. Workers can move between employers without losing skills and employers can costlessly replace these workers. Examples include migrant farm workers and custodial employees. Some professionals also fall into this group (certain engineers). There is no need for internal labour markets or other special governance structures to develop.

Primitive team occurs when there is low human asset specificity and individual performance is not easy to measure. In such teams, the membership may change without any loss of productivity, but individuals cannot be compensated on basis of their individual performance. Supervision of work effort is often introduced to regulate the work-pace of team members. So in this case, the governance structure of supervision may be introduced. Thus, one rewards workers on the basis of their work effort, say, that it attains a minimum level, rather than on their productivity, as in internal spot market.[6]

Obligational market obtains when there is a high degree of human asset specificity and individual performance is easy to measure. Now, both the firm and the workers have an interest in maintaining the employment relationship. The governance structure will be various safeguards and monetary penalties, such as severance pay, to discourage arbitrary dismissal. Non-vested retirement benefits for workers will discourage unwanted quits. In such cases the worker must remain with a firm for some period before accruing all rights to firm-paid pension benefits.

Relational team obtains when there is a high degree of human asset specificity and a low degree of measurement of performance. Special governance structures are then needed. The firm will engage in a considerable degree of social conditioning to help assure that workers understand and are dedicated to the purposes of the firm, and workers will be provided with considerable employment security. It is often said that many Japanese organizations are of this kind (see, e.g., the description in Aoki, 1988, ch. 3). Williamson (1985, ch. 10) claims that the relational team

is unusual, at least in contemporary Western organizations. Russell (1985) discusses how relational teams in fact may be more common than Williamson claims. Russell finds such teams in professional firms of lawyers, accountants, etc., and in a variety of worker-owned companies.

The obligational market is the kind of situation that usually calls for the development of internal labour markets, although this point is not stressed in Williamson (1985). The explanation of internal labour markets, where employees are shielded from competition of workers in the external market, face long promotion ladders, etc., focuses heavily on the human asset specificities that arise in some organizations and the transaction costs in dealing with these. The most explicit discussion of these asset specificities was given in Williamson (1975, ch. 4), where he draws on Doeringer and Piore (1971, pp. 15–16). Williamson (1975, p. 62) discussed four so-called task idiosyncrasies, which in Williamson (1985, ch. 9) were renamed asset specificities, namely: '(1) *equipment idiosyncrasies*, due to incompletely standardized, albeit common, equipment, the unique characteristics of which become known through experience; (2) *process idiosyncrasies*, which are fashioned or 'adopted' by the worker and his associates in specific operating contexts; (3) *informal team accommodations*, attributable to mutual adaptation among parties engaged in recurrent contact but which are upset, to the possible detriment of group performance, when the membership is altered; and (4) *communication idiosyncrasies* with respect to information channels and codes that are of value only within the firm. Because 'technology is [partly] unwritten and that part of the specificity derives from improvements which the work force itself introduces, *workers are in a position to perfect their monopoly over the knowledge of the technology should there be an incentive to do so*' (Doeringer and Piore, 1971, p. 84).' These asset specificities (called idiosyncrasies above) can account for the firm internal governance structure of internal labour markets.

Williamson's discussion of internal labour markets have received attention in the sociology literature. There are several alternative explanations of the emergence and existence of internal labour markets. I shall mention two.

In the so-called radical approach, in contrast to Williamson, internal labour markets are claimed to have no efficiency aspects. Rather, they were developed with the purpose of fragmenting and controlling the working class. The earlier statements were made by Stone (1973) and Marglin (1974), but the most sophisticated account is probably that of Edwards (1979, esp. ch. 1), who proposes an interesting theory for the development of organizational structure over the last 100 years, as now outlined.

In early capitalism, each employer employed a few workers and the machines used were simple. In this period, the organizational structure was one of *simple control*. The employer rewarded and punished workers as he wished.

As firms grew, more and more workers were employed under the same roof, and simple control became difficult. In this period employers developed machines to increase productivity. These machines in turn controlled the pace of work. This is referred to as *technical control*. The predominant 20th century example would be Ford and what later became termed Fordism.

Technical control turned out not to be fully efficient because workers could revolt against it. Employers therefore developed a third form of control. In this form, workers were given opportunities for advancement within an organization. The workers therefore came to identify with the organization and were hence less likely to revolt. This is referred to as *bureaucratic control*. Burawoy (1979, ch. 6) discussed how this type of bureaucratic control, relying on internal labour markets, induces lateral conflict between workers, since they now compete for promotions, and reduces hierarchical conflict between management and workers.

The story for why bureaucratic control was invented, as given by Edwards, has been disputed by many. Specifically, Jacoby (1985) claims the bureaucratization of the manufacturing employment relationship occurred as much from pressures of unions and workers as from the initiatives of employers. Elbaum (1984), in a study of labour relations in the U.S. steel industry, documents how demands from workers were important in developing internal labour markets.

Application 2: vertical integration

A firm stands in relationships both to suppliers of raw materials and to customers of their products. For example, producers of computer chips buy materials for the chips from suppliers all over the world, and sell the finished product, the chips, to computer producers, the customers. There is a market-relationship between supplier and customer.

Sometimes a firm buys up a supplier or a customer. Instead of standing in a market relationship to the customer or the supplier, the firm then stands in a hierarchical relationship to either of those two. The firm has authority over the previous supplier or buyer. There is no longer a need to write a contract. This is called *vertical integration*. When a firm buys up a supplier, it is more specifically called *backward integration*. Buying up a customer is called *forward integration*.[7]

Williamson (e.g., 1975, ch. 5, 1985, ch. 4) claims that vertical integration, which is widespread, can best be explained by concepts from transaction cost economics. Attempts to minimize transaction costs are not the only reasons for vertical integration, but according to Williamson, they are the most important.

The argument goes as follows (Williamson, 1985, ch. 4). Under uncertainty, bounded rationality, and opportunism, it will, as always, be difficult to write complete contracts with suppliers and customers. This is so because a complete contract probably would, due to bounded rationality, be costly and complex both to write and understand, even if one could anticipate all possible future contingencies. When the latter is not possible, then one can no longer write a complete contract. That in itself would not be a problem were it not for opportunism. However, when opportunism is

present then the parties will have the incentive to renegotiate to their own advantage when unanticipated events occur. There are then the costs of renegotiation. Again, the threat of renegotiation need not pose a problem. If the parties costlessly could withdraw from the contract and reestablish contracts with other parties, opportunism would in itself not pose a problem.

It is, however, rarely the case that the parties costlessly can withdraw from a contract, opening it up for competitive bidding, and reenter it with some other parties. The reason is that during contract execution some relationship-specific investments are usually made, of the types considered in Section 2.2 above: in site, physical, human, or dedicated assets. Moreover, there will usually be communication economics (private language between the parties) and trust relationships that need to be rebuilt.

When such relationship-specific investments are present, that is, there is asset specificity, the parties cannot costlessly withdraw from a contract. In the presence of unanticipated consequences that need to be dealt with in order to complete the transaction, opportunistic parties then have an interest in renegotiating the terms of the contract to their own best interest. There are costs of such renegotiation.

So, as before, there are two costs: the costs of writing the contract ex ante, before the contract execution, and the costs of renegotiation, ex post, during contract execution. Both are transaction costs. Williamson's claim is that vertical integration occurs in order to save on these transaction costs.

In deciding whether to vertically integrate or not, a company must weigh the weaker incentives and bureaucratic dysfunctions of a larger company against the savings in writing and renegotiating contracts.

Williamson (1985, ch. 4, esp. pp. 86–90) contrasts his own account of vertical integration with more standard accounts found in economic theory and U.S. antitrust legislation. I focus on the alternative account that stresses economies of scope.

Economies of scope occur when there are technological interdependencies between successive stages of production. Suppose product x_1 is an input into the production of product x_2. There are technological interdependencies when it is cheaper, assuming zero transaction costs, to produce both x_1 and x_2 in the same firm. The idea is that the machinery and human assets used to produce x_1 can also partly be used to produce x_2. So, when both products are produced in the same firm, there are some savings from the fact that some part of the machinery and human assets can be utilized in both production processes, thereby minimizing for example idle time.

Economies of scope have been a traditional argument for vertical integration, an argument also acknowledged by U.S. courts in antitrust cases. The most elaborate statement of these ideas is in Baumol, Panzar, and Willig (1982), ideas that were used extensively in two U.S. Supreme Court cases, the U.S. Government versus IBM and the U.S. Government versus AT&T and Bell Telephone. The former case was abandoned. The U.S. Government won the latter.

Williamson (1985, pp. 86–90, 92–95) argues that such economies of scope or technological interdependencies between successive stages of production are rare. Therefore, vertical integration due to economies of scope is not widespread. Instead, vertical integration is more often caused by attempts to save the transaction costs with which market-type relationships are plagued.[8]

Application 3: multidivisional form of organizations

Contemporary organizations are often organized according to three different organizational forms: the U-form, the M-form, and the H-form (see Williamson, 1975, ch. 8).[9]

In the *U-form*, that is, the *unitary form*, the company is organized along functional lines. The divisions of the company correspond to the functions they serve: Sales, finance, and manufacturing.

In the *M-form*, that is, the *multidivisional form*, the company is organized along product, brand, or geographical lines. Each such division is semi-autonomous. Typically, each division is organized as a U-form. The divisions are held together by a general office or management. The latter is in charge of strategy, planning, auditing, and resource allocation. Each division is in charge of operations.

The *H-form*, that is the *holding company*, is organized along product, brand, or geographical lines, as in the M-form. However, it lacks the general office that coordinates the divisions. Each division is more or less autonomous.

In the 20th century companies have changed from U-form to M-form. The M-form was developed more or less simultaneously in the 1920s by Pierre S. du Pont of the du Pont Company and Alfred P. Sloan of General Motors. du Pont used to be organized as a U-form company. General Motors used to be organized as an H-form company (see Williamson, 1985, pp. 279-280).[10] Williamson claims that transaction cost economics can explain this change.

More specifically he claims the following. When the U-form company grows in size two problems arise. *First*, there is a loss of control on the part of management of the company. This loss of control is to a large extent due to bounded rationality. With increasing scale and scope of operations, it is no longer possible to have full oversight and control of the operations. As the size increases, more hierarchical levels are required, which in turn therefore increases transaction costs, leads to lowering of incentives and bureaucratic dysfunctions. *Second*, management within various divisions of the company have less incentives to pursue profit maximization. It is harder as the company grows, to detect deviations from profit maximization and to reward those that do well. This is particularly the case because in a large company it is almost impossible for stockholders to collect the information about internal operations that is relevant for profitability. The deviations from profit maximization is caused by opportunism. Each manager might pursue her own goals instead of those of the company without being caught.

The solution to these two problems is to create the M-form. Each division in the company typically operates as a profit-centre. The general office or management audits each centre, collects information, awards incentives, and allocates funds. This is much easier to do for the general office than for stockholders.

In the M-form company, therefore, there is a separation of strategic planning and resource allocation on the one hand and daily operations on the other hand. The general office or management is responsible for strategic planning and resource allocation. It also monitors and controls the divisions. The division managers are responsible for operating duties and tactical decisions.

In order successfully to implement the M-form, it is required that the general office is in charge of the following activities (Williamson, 1985, p. 284): (1) the identification of separable economic activities within the firm; (2) according quasi-autonomous standing (usually of a profit-centre nature) to each; (3) monitoring the efficiency performance of each division; (4) awarding incentives; (5) allocating cash flows to high-yield uses; and (6) performing strategic planning (diversification, acquisition, divestiture, and related activities) in other respects. The M-form structure is thus one that *combines* the divisionalization concept with an internal control and strategic decision-making capability.[11]

Note that in this analysis, the concept of asset specificity is absent. The causes of the M-form of an organization are (1) bounded rationality which leads to loss of control when the organization grows, and (2) opportunism which leads to managers pursuing goals different from profit maximization when control is being diffused.

In summary, Williamson claims that the M-form replaced the U-form or the H-form replaced the U-form for large companies for efficiency reasons: The M-form economized better on the transaction costs that were incurred as company size grew.[12]

2.4 Empirical investigations

Williamson's TCE framework has generated a large number of empirical investigations. Many of these are summarized and discussed in Shelanski (1992). Rather than discussing a large number of studies, I focus on a few select studies and discuss these in more detail.[13]

Empirical investigation 1: work organization and hierarchy

The aspect of Williamson's work that seems to have received the least attention in empirical investigations is what he has to say about work organization and work hierarchy. His ideas on this topic have influenced a considerable number of empirical studies in the sociology literature, but I am not aware of any studies in that literature that directly address Williamson's concerns.[14]

One interesting and rather successful application of TCE to employment relationship is that of Galanter and Palay (1991). They discuss several aspects of corporate law firms in the U.S. Corporate law firms typically hire graduates directly out of law schools. The more prestigious the firm, the more prestigious law schools it hires from. The newly hired law school graduate works as an associate in the firm. This is a position without security of employment. The salary as an associate depends mostly on his or her seniority in the company. It increases with seniority. Most corporate law firms pay the same going rate for associates with same seniority. Then, after a period of 5–10 years, the associate is up for promotion to a partner. If he or she receives the promotion, then he or she receives security of employment, a so-called tenured position. A partner usually receives a baseline salary, which can be quite high, plus a share of profits from the business, which can be even higher. The partners run the firm more or less as a club. If the associate is denied the promotion, he or she usually has to leave. It is an 'up-or-out' system. Accounting firms use a similar system. In the 1960s, about 1 out of 5 to 1 out of 15, depending on the firm, of associates made the promotion to partner after some 5–7 years (see Galanter and Palay, 1991, pp. 26–32).

A quick look at the earnings of associates and partners in large corporate law firms gives a sense of the salience which the promotion from associate-to-partner holds for most associates. In 1986, the top corporate law firm, Cravath, Swaine & Moore, based in New York City, paid a first-year associate an annual salary of $65,000. In 1989, Cravath paid its 67 partners incomes at an average of $1,595,000 (see *The American Lawyer,* July-August, 1989, p. 34). The five highest-paying corporate law firms paid their 438 partners average incomes ranging from $1.15 million to almost $1.6 million. Partners at the top of these firms earned considerably more (see Jensen and Murphy, 1991, p. 13).

Galanter and Palay (1991, ch. 5) proposed a TCE theory for why this promotion-to-partnership system developed. The explanation goes as follows. According to Galanter and Palay (1991, pp. 89–98), partners with lots of experience possess excess human capital, which means that there is more demand for their services than they can satisfy, due to built-up reputation, long-standing ties with some customers, etc. The partners therefore take on associates to deal with this excess demand. The associates would not be able to generate demand on their own. But, in the relationship with the partner, they can get business. In this relationship, there is ample opportunity for opportunism on the part of the associates in the use of the assets (i.e., excess business) the associates get turned over to them. To alleviate this opportunism the incentive and monitoring device of 'promotion-to-partnership' developed. A fixed percentage of eligible associates is promoted each year. And, once promoted, the rewards can be staggering. Therefore, it may be in the best interest of an associate to act responsibly, not to take advantage of the opportunities for opportunism that exist, otherwise, he or she may be denied the promotion and its rewards. This is a TCE explanation for the promotion-to-partnership system. The emphasis is on the behavioural assumption of opportunism on part of associates, and to a lesser extent on the human asset specificity these associates develop prior to potentially becoming a partner. It is the opportunity to get promoted that keeps the associates working hard and prevents them from acting opportunistically.[15]

Empirical investigation 2: vertical integration

Monteverde and Teece (1982) attempted to test the hypothesis that vertical integration in automobile production would occur for those automobile parts for which the highest amount of engineering effort is required. The ideã is this. Some automobile parts require a lot of research and experimentation to develop. Engineers will do this. In those cases, there will be transaction costs from having a supplier produce the parts, because there may be need for renegotiation, bargaining etc., as the part is being developed. Thus, automobile parts that require a lot of engineering effort are more likely to be produced in-house than those that do not.

They collected data from GM and Ford on the degree of in-house production of a selection of automobile parts and on the amount of engineering effort required per part. The data show that in-house production is more widespread for parts that require a lot of engineering effort. This is consistent with a TCE interpretation.

A functional substitute for vertical integration can be a long-term contract. Joskow (1985) studied the electrical utility industry in the U.S. Most utilities use coal as the basic raw material for producing electricity. A utility plant may choose between locating itself close to its customers of electricity or close to a coal supplier, that is, a coal mine. If the former is done, coal can be bought from any supplier. If the latter is done, the utility has a clear interest in buying from the coal mine that it has located its plant next to. In that case, the utility's investment in the plant has a lot of asset specificity, namely site specificity. The utility has higher profit rates if it can buy the coal directly from the mine next to which it is located because transportation costs are then saved.

Under this scenario, the theory of TCE would predict that a utility that built next to a mine would try to vertically integrate the mine. Joskow shows that this rarely occurs. Instead of vertical integration, the coal mine and the utility relies on a functional substitute, namely a long term contract (e.g., 25–30 years) regulating the sale of coal to the utility. Such long-term contracts ensure that an open-mouth utility (i.e., a utility located next to a coal mine) can capitalize on their asset-specific investment. The contracts are typically signed prior to the utility building next to the coal mine. Joskow shows that long-term contracts are prevalent among open-mouth utilities, located next to a coal mine, whereas they are not prevalent among utilities located next to customers, that is, not located next to coal mines. This is consistent with predictions from the theory of TCE.

I make two comments on Joskow. First, one can argue that there need not be a strong incentive for the open-mouth utility to vertically integrate the coal mine. The reason is that the utility knows that the coal mine has an interest in selling coal to the utility because that saves on sales effort and transportation costs for the coal mine. The utility still needs a long-term contract, because without it, the coal mine could start asking higher prices once the utility is built. The utility is probably not economically feasible if it must revert to buying on the open market because then it will incur the transportation costs in addition to the costs of getting the electricity to the customers who may be located far away. This can be exploited by the coal mine unless there is some kind of long-term contract regulating what goes on. But there is not necessarily a need for vertical integration.

Second, if most utilities are open-mouth utilities, as in Western U.S., and most coal is sold to utilities, the problems of opportunism under site-specific investments might disappear altogether. The reason then is that the mine need not have an alternative customer for its coal and threats of withdrawing supply would hence be less credible.

Another functional substitute for vertical integration is incomplete contracts executed under trust and commitment. It is often claimed that Japanese business relationships between a firm and its customers and suppliers are different from those in the U.S. In particular, it is claimed that there is considerable amounts of trust and commitment in the relationships, as described, for example, by Dore (1986, esp. ch. 3). He claims that relationships of trust and moral obligation are important in regulating interaction between firms (Dore, 1986, p. 77). These relationships are also supported by social sanctions (Dore, 1986, p. 80). In the presence of such relationships of trust and commitment the dangers of opportunism in transactions with relationship-specific investments are mitigated. Thus, one economizes on transaction costs. At the same time one does not incur the costs of lower incentives and bureaucratic dysfunction following from vertical integration. See, however, Williamson's (1993) discussion of trust.

Dore (1987, esp. ch. 9) claims that such 'relational' contracting, as he calls it, based on trust and commitment, serves many functions, namely providing risk-sharing and security, providing trust, providing friendships, but above all, providing economic efficiency. If transaction costs due to opportunism when relationship-specific investments are made is as big a problem as Williamson claims, then the Japanese seem to have developed a solution superior to vertical integration, a solution other countries may attempt to imitate.

Empirical investigation 3: the multidivisional form

Armour and Teece (1978) studied a sample of petroleum firms in the 1955–1973 period, some of which were organized as M-form and others as U-form. They assessed how the companies' profit rates depended on their forms.

They found that in the early period, 1955–1968, the profit rates of M-form companies on the average were 33% higher than those of U-form companies, 8 and 6

percent respectively. In the later period, 1969–1973, no differential performance could be observed. The explanation for this is that in the earlier period the M-form was being diffused. All companies that could benefit from the M-form adopted it. In the latter period, the remaining U-form companies were probably best organized as U-form companies. Hence, no differences in performance between M-form and U-form were observed.

I make one comment on this study. What would have constituted a more crucial test of the M-form hypothesis, would have been to compare the intra-firm performance of companies before and after the switch from U-form to M-form.

Fligstein (1985) studied the causes of adoption of M-form of the 100 largest U.S. companies in the period 1919–1979. He claims his evidence is not consistent with Williamson's argument. The size of a company is not a major determinant of adoption of M-form. Size was measured by the value of assets. The problem with this conclusion is that Fligstein does not present data on distribution of assets in his sample. Therefore, one cannot assess whether assets is a strong determinant or not of adoption of M-form because one has no idea about what the relevant range of the asset variable is. Also, there is a selection problem in that only the 100 largest firms were studied.

Fligstein reports two additional important findings. First, the strategy of a company has a strong effect on whether M-form is adopted. Three strategies were considered: product dominant, where a single product accounts for at least 70% of a firm's output; product related, where the firm produces several related products and no one product line accounts for more than 70% of output; and product unrelated, where the firm engages in several unrelated businesses and no one product line accounts for more than 70% of revenue (Fligstein, 1985, p. 383). Firms with product dominant strategies are less likely to adopt the M-form.

Second, Fligstein reports that the background of the president of the company has strong effects on adoption of M-form. If the president is from the finance or sales division of the company, adoption of the M-form is much more likely. This finding is interpreted as follows. The adoption of the M-form to a large extent reflects the interests of those who are in power, and those with a background in the sales or finance division had larger interests in adopting the M-form than those with other backgrounds.

2.5 Criticisms and elaborations

Stinchcombe's elaboration of TCE

In many situations where bounded rationality, opportunism, uncertainty, and small-numbers bargaining are present, we still see that markets or contracts are used (Stinchcombe, 1985). Some examples are these. Research universities build quite a bit. Still they do not buy up building companies, that is, vertically integrate, but rely on outside contractors. The transaction costs are high and there are lots of uncertainty in all construction projects. Another example is Research and Development. Governmental agencies often put out requests for research proposals and then give the grants to whoever they find best. There are tremendous transaction costs in writing the contracts for a research project. Still the government does not vertically integrate.

Stinchcombe (1985) discusses how to minimize the transaction costs in these kinds of market-type relationships, in situations where vertical integration is not chosen. Instead of vertical integration, what happens in these cases, according to Stinchcombe, is that hierarchical elements are added to the contracts. What looks like a market relationship, that is, a contract, has elements of hierarchy or authority. Specifically, many contracts contain clauses about five features typically thought to characterize hierarchies. *First*, there are authority systems. These specify who can decide on changes under which conditions. For example, in research, the contractor cannot make changes, though it can propose changes, which the client has the authority to accept or turn down. The client does not always have the authority to require changes, because they are bound by the contract, but typically the contractor will yield to the demands for changes the client might make. *Second*, there will often be incentive systems, where rewards for performance are given. The client has the authority, often, to terminate a contract if certain goals are not met within specified dates. *Third*, there will often be provision for conflict resolution. *Fourth*, there may be explicit standard operating procedures, such as deadlines, etc. *Fifth*, there may be administered price systems. These are rules for determining prices in the future, for adjusting costs, etc. In conclusion, these are elements of hierarchy, but are found in many long-term contracts.

Including these hierarchical features in a contract allow the parties to be flexible about the future, in particular to adjust to uncertainty and allow risk-sharing. For example, if costs go up, the client and the contractor may share the burdens of this.

Joskow (1985) documents how several of these hierarchical features are included in long-term contracts between open-mouth electrical utilities and the coal mines located next to these utilities. These are complex rules about how prices may be changed, as well as procedures for conflict resolution.

Stinchcombe's conclusion is that the dichotomy between contracts (i.e., markets) and hierarchies (i.e., large firms) is too ideal-typic. It covers only a few of the existing cases. Instead, what we often see is that long-term contracts have hierarchical elements (as in construction and research). Conversely, many hierarchies have contractual elements built into them (such as piece-rates in employment relationships).[16]

The critics

There are several critical treatments in sociology of Williamson's framework. One that has become well known is that of Perrow (1986, ch. 7).[17] He makes three points, aimed against Williamson's discussion of the growth of large firms. First, he claims that there are transaction costs also in hierarchies. Second, incentives get distorted in hierarchies. Third, the reasons for large firms are not transaction costs, but primarily these: (a) market power, that is, the elimination of competitors, and (b) government support and regulation. As has been discussed several times above, Williamson accommodates the two first points. They figure prominently in his explanation of the M-form. The third point can only be settled by empirical research.

Granovetter (1985), in an important article, argues that most economic transactions are embedded in social relationships – in particular, interpersonal networks – that are important for how transactions are executed. He does not want to get rid of the rationality assumption, or the methodological individualism framework, but argues that (Granovetter, 1985, pp. 504–510): (1) one must include in the description of the constraints (as well as resources) that actors face, the relevant interpersonal networks, the information available, and the reputation actors enjoy, and (2) one needs a more broad definition of what a rational actor is, namely an actor that also seeks sociability, approval, that is, one whose utility function consists of more arguments than the utility function found in economic theory.

This framework he uses to criticize transaction cost economics explanations of vertical integration (Granovetter, 1985, pp. 493–504). His main point is that most markets are sufficiently integrated by personal relationships so that hierarchy (i.e., vertical integration) is not needed to safeguard against opportunism when unanticipated events occur in on-going market exchanges with relationship-specific investments. Vertical integration may be needed to safeguard against opportunism only in the absence of personal relationships.

The reasons that personal relationships can alleviate opportunism are first that trust develops between the partners, and second that social rewards develop when business partners have relationships that are social as well as for business purposes. These two factors mitigate against opportunism.

Granovetter's framework I find persuasive. It even suggests some researchable hypotheses, namely that vertical integration should develop more often in settings where interpersonal networks are lacking, whereas where those are present, market-type relationships should be more prevalent, even in the presence of asset specificity.

Williamson (1988) responds both to Perrow and Granovetter. To Perrow he basically responds that the fictitious example which Perrow bases his discussion on hardly warrants attention. Moreover, he points out that Perrow does not offer any alternative theory for vertical integration, nor does he examine the empirical literature on vertical integration. Williamson (1988, p. 183) sees the value in the embeddedness argument but finds it too underdeveloped to accept it as an alternative viewpoint.

2.6 Concluding remarks

In the field of organization theory, transaction cost economics is currently probably the most comprehensive theoretical framework available. It is comprehensive because it deals with a wide range of empirical phenomena: the internal organization of firms through authority systems and internal labour markets, vertical integration between firms, the multidivisional versus the unitary form of an organization, long-term contracts versus spot contracts, and more. In organizational sociology, in contrast, population ecology (Hannan and Freeman, 1989) and institutional theory (e.g., Scott, 1987), two leading frameworks, deal only with some limited aspects of organizational behaviour, be it the births and deaths rates of organizations or imitation of other firms and adaptation to external pressures. TCE has furthermore generated a rather large empirical literature, quantitative as well as qualitative, a literature that assesses the fruitfulness of the perspective in a wide array of substantive settings, often showing it to yield insights.

In the economics literature the other main theory of organization is that of principal-agent theory (see, e.g., Milgrom and Roberts, 1992, chs. 5–6). Some comparative remarks are in order. In TCE the central dependent variable is the governance structure for transactions. The framework has been most successful for explaining exchanges of physical goods, such as automobile parts, computer chips, and so forth. But it applies also to the exchange of services, that is, to the governance of personnel and its associated motivational problems. The so-called principal-agent literature also explains governance structures for transactions, but it explains, in contrast, almost exclusively the exchange of services, that is, it deals with how to contract for labour. The focus is therefore, in that literature almost exclusively on selection and motivation of agents, not on how to organize exchange of physical goods.[18]

There are areas where the transaction cost economics approach can fruitfully be supplemented with or contrasted to ideas from sociology. To date, the most successful attempts to do so have come from the network literature, where the claim is that the social networks that business leaders are embedded in mitigate against opportunism and hence may operate as functional substitutes for the solutions Williamson outlines, such as vertical integration (i.e., Granovetter, 1985). Another area where sociology has made some contributions is in cross-national and cross-cultural studies of organizations (e.g., Hamilton and Biggart, 1988).

Appendix 1: Relationship between Williamson 1975 and 1985

Above, I discussed the theory as set out in Williamson (1985). Below, I relate that account to the earlier account as set out in Williamson (1975).

The theory, as set out in Williamson (1975, esp. ch. 2), has four elements. Economic and social exchanges are troubled with

1. *Bounded rationality*, as in Williamson (1985),
2. *Opportunism*, as in Williamson (1985),
3. *Uncertainty*, as in Williamson (1985),
4. *Small numbers bargaining*, which was not a basic term in Williamson (1985), but rather a derived term. See section 2.2 above.

There is a slight change in presentation of ideas and a sharper definition of terms between Williamson (1975) and Williamson (1985). I believe Williamson (1979) to be the place where these changes were first articulated.

Williamson (1985) maintains three of his original four basic terms, bounded rationality, opportunism, uncertainty, but not the fourth, small numbers bargaining. He now regroups these three terms by distinguishing between *behavioural assumptions* on the one hand and characteristics of the transactions in question on the other hand. The characteristics of the transactions he refers to as *dimensionalizing* (e.g., Williamson, 1981, pp. 554–555).

In Williamson (1985) the *behavioural assumptions*, then, are these

1. *Bounded rationality*, as in Williamson (1975),
2. *Opportunism*, as in Williamson (1975),

while the *characteristics* of the *transactions* are

1. *uncertainty*, as in Williamson (1975),
2. the *frequency* with which *transactions recur*, which was not in Williamson (1975), but is new,
3. *asset specificity*, which was not in Williamson (1975), but is new.

So, the change since 1975 is threefold. First the basic terms have been regrouped into those that constitute behavioural assumptions and those that characterize transactions. Second, the term small-numbers bargaining has been dropped and replaced with the frequency with which transactions recur. Third, the term asset specificity has been added to the list of basic terms. Small-numbers bargaining is still important in the theory, but is no longer one of the basic terms. Rather, it is a term that gets derived from the basic terms, as explained in Section 2.2 above.

Appendix 2: Opportunism and the breakdown of trust in incomplete contracts

Michael Gerlach (see Gerlach, 1990) was an expert witness in a California Supreme Court case between Kikkoman, a major producer of soy sauce in Japan, and Sugihara, the sole distributor of Kikkoman soy sauce in the U.S. The case is based on his work as a witness in the case. It illustrates clearly the problems arising from relying on trust, as opposed to some of the claims made by Dore (1986, 1987). See also Williamson's discussion of trust.

Kikkoman entered into a sole distributor agency contract with Sugihara in 1951, but their relationship dated back to about 1939. Sugihara, a U.S. based firm with a U.S.-Japanese owner, was to distribute Kikkoman products. The contract between the parties was intentionally left incomplete, according to Sugihara. The rationale behind the incomplete contract was that the parties could not foresee all the events that might occur as the relationship would unfold and that an incomplete contract would allow the parties to adapt to unanticipated events in an appropriate way.

Note that this way of conducting business appears to be standard among Japanese corporations (e.g., Dore, 1986, 1987), where considerable amounts of trust and goodwill are claimed to be present.

The relationship unfolded and became quite successful. Kikkoman and Sugihara developed a joint venture in 1957 and expanded distribution from California to the entire U.S. During the period 1957–1984 (in particular 1957–1969) the venture was based on the logic of long-term reciprocity (Gerlach, 1990).

During the joint venture period, Sugihara made a number of relationship-specific investments, in monetary terms, perhaps not so large, but in human asset specificity (i.e., knowledge and expertise) enormous. Among other things, Sugihara sold his family business to Kikkoman at discounted prices in order to strengthen the joint venture.

This venture was held together by quite close ties between family members of the two companies. However, as older family members in Kikkoman started to get replaced by younger ones in the late 1970s and early 1980s, the joint venture started to break down. The younger members had not been party to the original informal agreements. They started to act opportunistically.

Then in the 1980s Sugihara decided to withdraw from the relationship. It was no longer tenable. Kikkoman's response when the proposal of withdrawal was made, was to treat Sugihara not as a joint partner, but as a second- or third-order subsidiary. Kikkoman consequently offered to pay Sugihara a small sum of money, about $170,000, which would be equivalent to a return on investment of about 5% annually on Sugihara's original investment of about $70,000.

The case was brought to court by Sugihara in 1984 and was settled in favor of Sugihara against Kikkoman in 1989. An out-of-court monetary settlement was reached in December 1989. The numbers are confidential, but it appears that Sugihara did quite well, leaving the relationship with a large sum of money.

The argument of the plaintiff Sugihara in the court case was that the joint venture with Kikkoman was intentionally left incomplete in contractual terms so as to allow the relationship to evolve in a manner best suited to their joint interest when unanticipated events would occur. This, they claimed, was the way business routinely is conducted in Japan. The incomplete contract worked well for almost 40 years.

Kikkoman, in contrast, claimed that the joint venture was nothing more than a standard business relationship that did not extend beyond the terms of the explicit parts of the contract by the parties.

In the court proceedings, Sugihara argued (1) that they would not have acted as they actually did unless the contract between the parties extended well beyond the terms of the explicit contract, for example, they would not have sold the family business at discounted prices, and (2) that their actions were in full congruence with norms for doing business in Japan, whereas Kikkoman had violated those norms in the early 1980s.

Kikkoman defended themselves by arguing that the relationship did not extend beyond the explicit contract and that they had acted in full compliance with California and U.S. Corporate law.

The California Supreme Court ruled (October, 1989) in favor of Sugihara. Two principles of legal reasoning were brought to bear. First, there clearly was an unwritten or implicit contract between the parties that extended beyond the explicit contract. All of the actions of Sugihara were consistent with such verbal agreement. In California law, verbal agreements are as binding as written; their main problem is one of verification. Second, it was not unreasonable to hold the two parties to norms for business practices prevalent in Japan, as long as both parties understand that those norms are in force and the norms are not counter to U.S. law. After the first verdict against Kikkoman, the parties came to an out-of-court settlement as the trial was to continue in December 1989.

In more conceptual terms, the case illustrates three things. First, it shows the enormous strength of an incomplete contract in allowing the parties to adapt to unanticipated consequences. Under bounded rationality such incomplete contracts can be quite advantageous. Second, it shows how trust and interpersonal relationships mitigate against opportunism when unanticipated events occur. Finally, it shows the weaknesses of incomplete contracts and the breakdown of trust when some of the members to the original agreement no longer are in charge. So, even potentially quite powerful, incomplete contracts are also fragile. The crux of the problem was that the transaction or relationship was executed in the U.S., where different norms for conducting business prevail. Sugihara, who was located in California, felt bound by Japanese business norms while conducting business in the U.S. with a Japan-based corporation. The latter felt bound by the same norms initially. However, as leadership in Kikkoman was slowly transferred to a new generation, the younger managers did not feel constrained by Japanese business norms while conducting business with the U.S.-Japanese Sugihara in California.

Finally, it is quite instructive to discuss the case in light of legal theory and sociology of law. One often distinguishes, in an ideal-typic description of law, between two extreme positions. First, there is law as a *closed* system in which all decisions follow from applying logic to legal rules or laws. This is sometimes referred to as the formal law position or legal positivism (see, e.g., Murphy and Coleman, 1990, pp. 19–33). Second, there is the position known as free judicial discretion, where judges are not bound by legal rules nor by precedent. This is a position taken by the Critical Legal Studies group at the Harvard University Law School (see e.g., Kelman, 1987).

According to the influential legal theorist Karl Llewellyn, whose position usually is called legal realism, the U.S. case law system lies somewhere in between these two extremes. The case law system has two central features: (1) what judges think is regarded as more centrally important than what legislators or scholars think, and (2) no systematic code of law exists.

Llewellyn ([1933] 1989) claims that in order to understand legal reasoning in the U.S. one must not only analyze formal law and precedent but also the *social context* within which decisions occur. Judges abide by three principles in reaching decisions. First, the 'operating techniques' of judges and lawyers constitute a set of internalized norms in the legal profession for reaching decisions. Second, the facts of the situation, rather than the legal rules, guide decisions and produce a 'sense of justice in the individual case.' Third, legal decisions and legal rules tend to be in accordance with 'real-life' norms. Therefore, they are in constant change as 'real-life' norms change.

There are several debatable features of Llewellyn's analysis. With respect to the second point, facts themselves are often constructed, and rarely they tell us what is the just decision. With respect to the third point, norms are not always shared by all relevant groups, and even if they were, norms are not always just and sometimes are contradictory (see Gewirtz, 1989).

Nevertheless, Llewellyn's analysis can usefully be applied to the case at hand. It is clear that the judges in the case did not slavishly follow formal law, relying instead on a different operating procedure, probably the operating procedure of the legal profession. It is equally clear that the 'facts' of the situation guided the decision and made it fair, thereby serving justice, even when formally, as seen from the narrow viewpoint of law, Kikkoman appears to have formal law on their side. Finally, the decision was in line with 'real-life' norms, in this case business norms in Japan. The judges decided to let the parties be bound by Japanese norms, which constituted the relevant benchmark for both Kikkoman and Sugihara.

References

Anderson, Erin (1988), 'Transaction Costs as Determinants of Opportunism in Integrated and Independent Sales Forces,' *Journal of Economic Behaviour and Organization,* 9(3), pp. 247–262.

Aoki, Masahiko (1988), *Information, Incentives, and Bargaining in the Japanese Economy,* Cambridge University Press, New York.

Armour, Henry, and David Teece (1978), 'Organizational Structure and Economic Performance,' *Bell Journal of Economics,* 9(1), pp. 106–122, Spring 1978.

Baumol, W.J., J.C. Panzar, and R.D. Willig (1982), *Contestable Markets and the Theory of Industry Structure,* Harcourt Brace Jovanovich Inc., New York.

Becker, Gary S. (1975), (2nd ed.), *Human Capital,* University of Chicago Press, Chicago.

Blau, Peter M., and Marshall W. Meyer. 1987 (3rd ed.), *Bureaucracy in Modern Society,* Random House, New York.

Burawoy, Michael (1979), *Manufacturing Consent,* University of Chicago Press, Chicago.

Chandler, A.D., Jr. (1962), *Strategy and Structure,* MIT Press, Cambridge, MA, subsequently published in New York, Doubleday & Co., 1966.

Chandler, A.D., Jr. (1977), *The Visible Hand: The Managerial Revolution in American Business,* Harvard University Press, Cambridge, MA.

Coase, R.H. [1937] (1988), 'The Nature of the Firm,' *Economica N.S.,* 4, 386–405, Reprinted as ch. 2 in R.H. Coase 1988, *The Firm, The Market, and the Law,* University of Chicago Press, Chicago.

Coase, R.H. (1987), 'R.H. Coase Lectures,' *Journal of Law, Economics, and Organization* 4(1), pp. 3–47, Spring.

Cyert, R.M., and J.G. March (1963), *A Behavioural Theory of the Firm,* Prentice-Hall, Englewood Cliffs, NJ.

Doeringer, P., and M. Piore. (1971), *Internal Labour Markets and Manpower Analysis,* D.C. Health, Lexington, MA.

Dore, Ronald (1986), *Flexible Rigidities: Industrial Policy and Structural Adjustments in Japan 1970–1980,* Stanford University Press, Palo Alto.

Dore, Ronald (1987), *Taking Japan Seriously. A Confucian Perspective on Leading Economic Issues,* Stanford University Press, Stanford, CA.

Eccles, Robert G., and Harrison C. White (1988), 'Price and Authority in Inter-Profit Center Transactions,' *American Journal of Sociology,* 95, Supplement, pp. S17–S51.

Edwards, Richard (1979), *Contested Terrain. The Transformation of the Workplace in the Twentieth Century,* Basic Books, Inc, New York.

Elbaum, Bernard (1984), 'The Making and Shaping of Job and Pay Structures in the Iron and Steel Industry,' ch. 3, pp. 71–107 in Paul Osterman (ed.), *Internal Labour Markets*, MIT Press, Cambridge, MA.

Elster, Jon (1983), *Explaining Technical Change*, Cambridge University Press, Cambridge.

Fligstein, Neil (1985), 'The Spread of the Multidivisional Form Among Large Firms, 1919–1979,' *American Sociological Review*, 50(3), April 1985, pp. 377–391.

Fligstein, Neil (1990), *The Transformation of Corporate Control*, Harvard University Press, Cambridge, MA.

Galanter, Marc, and Thomas Palay (1991), *Tournament of Lawyers: The Transformation of the Big Law Firm*, University of Chicago Press, Chicago.

Gerlach, Michael (1990), 'Trust is Not Enough: Cooperation and Conflict in Kikkoman's American Development,' *Journal of Japanese Studies*, 16(2), pp. 389–424.

Gewirtz, Paul (1989), 'Editor's Introduction,' pp. ix–xxiii in Llewellyn, Karl [1933] 1989, *The Case Law System in America*, Translated from original 1933 German edition, University of Chicago Press, Chicago.

Granovetter, Mark (1985), 'Economic Action and Social Structure: The Problem of Embeddedness,' *American Journal of Sociology*, 91(3), pp. 481–508.

Hamilton, Gary G., and Nicole Woolsey Biggart (1988), 'Market, Culture, and Authority: Analysis of Management and Organization in Far East,' *American Journal of Sociology*, 95, Supplement, pp. S52–S94.

Hannan, Michael T., and John Freeman (1989), *Organizational Ecology*, Harvard University Press, Cambridge, MA.

Jacoby, Sanford M. (1985), *Employing Bureaucracy. Managers, Unions, and the Transformation of Work in American Industry, 1900–1945*, Columbia University Press, New York.

Jensen, Michael C., and Kevin F. Murphy (1991), 'CEO Incentives – It's Not How Much You Pay, But How,' pp. 3–18 in *Compensation. Fair Pay for Executives and Employees*, Harvard Business Review, Harvard Business School, Boston, MA.

Joskow, Paul L. (1985), 'Vertical Integration and Long-term Contracts: The Case of Coal-burning Electric Generating Plants,' *Journal of Law, Economics, and Organization*, 1(1), Fall 1985, pp. 33–80.

Kelman, Mark (1987), *A Guide to Critical Legal Studies*, Harvard University Press, Cambridge, MA.

Llewellyn, Karl [1933] (1989), *The Case Law System in America*, Translated from original 1933 German edition, University of Chicago Press, Chicago.

Lorsch, Jay W. (1967), 'The Texana Petroleum Corporation,' Case Study 6–413–056, Harvard Business School, Harvard University, Boston, MA.

Macaulay, S. (1963), 'Non-contractual Relations in Business,' *American Sociological Review*, 28(1), pp. 55–70.

Macneil, I.R. (1974), 'The Many Futures of Contracts,' *Southern California Law Review*, 47, May, pp. 691–816.

March, James G., and Herbert D. Simon (1958), *Organizations*, Wiley, New York.

Marglin, Stephen D. (1974), 'What Do Bosses Do? The Origins and Functions of Hierarchy in Capitalist Production,' *Review of Radical Political Economics*, 6, pp. 33–60.

Milgrom, Paul, and John Roberts (1992), *Economics, Organization, and Management*, Prentice Hall, Englewood Cliffs, NJ.

Moe, Terry (1984), 'The New Economics of Organization,' *American Journal of Political Science*, 28(4), November 1984, pp. 739–777.

Monteverde, Kirk, and David Teece (1982), 'Supplier Switching Costs and Vertical Integration in the Automobile Industry,' *Bell Journal of Economics*, 13 (1), Spring 1982, pp. 206–213.

Murphy, Jeffrie G., and Jules L. Coleman (1990), *Philosophy of Law. An Introduction to Jurisprudence*, Westview Press, Boulder, CO.

Perrow, Charles (1986), (3rd ed.), *Complex Organizations. A Critical Essay*, Random House, New York.

Petersen, Trond (1992a), 'Individual, Collective, and Systems Rationality in Work Groups: Dilemmas and Market-type Solutions,' *American Journal of Sociology*, 98(3), November 1992, pp. 469–510.

Petersen, Trond (1992b), 'Individual, Collective, and Systems Rationality in Work Groups: Dilemmas and Nonmarket Solutions,' *Rationality and Society*, 4(3), July, pp. 332–355.

Powell, Walter, W. 1987. 'Hybrid Organizational Arrangements: New Form or Transitional Development?' *California Management Review*, 30(1), Fall 1987, pp. 67–87.

Russell, Raymond (1985), 'Employee Ownership and Internal Governance,' *Journal of Economic Behaviour and Organization*, 6(3), September 1985, pp. 217–241.

Scott, W. Richard (1987), 'The Adolescence of Institutional Theory,' *Administrative Science Quarterly*, 32(4), pp. 493–511.

Shelanski, Howard A. (1992), 'Empirical Research in Transaction Cost Economics,' Unpublished manuscript, Haas School of Business, University of California at Berkeley, Berkeley, California.

Simon, Herbert A. (1976), [1945], *Administrative Behaviour*, The Free Press, New York.

Stinchcombe, Arthur L. (1985), 'Contracts as Hierarchical Documents,' ch. 2, pp. 121–171, in Arthur L. Stinchcombe and Carol A. Heimer, *Organizations and Project Management*, Norwegian University Press, Oslo.

Stone, K. (1974), 'The Origins of Job Structures in the Steel Industry,' *Review of Radical Political Economics*, 6, Summer, pp. 61–97.

Swedberg, Richard (1990), *Economics and Sociology. Redefining Their Boundaries: Conversations with Economists and Sociologists*, ch. 6, pp. 115–129, Interview with Oliver E. Williamson, Princeton University Press, Princeton, NJ.

Williamson, Oliver E. (1975), *Markets and Hierarchies. Analysis and Antitrust Implications*, The Free Press, New York.

Williamson, Oliver E. (1979), 'Transaction-Cost Economics: The Governance of Contractual Relations,' *Journal of Law and Economics*, 22(2), October 1979, pp. 233–261.

Williamson, Oliver E.(1981), 'The Economics of Organization: The Transaction Cost Approach,' *American Journal of Sociology*, 87(3), November 1981, pp. 548–577.

Williamson, Oliver E. (1985), *The Economic Institutions of Capitalism*, The Free Press, New York.

Williamson, Oliver E. (1988), 'The Economics and Sociology of Organization. Promoting a Dialogue,' ch. 8, pp. 159–185, in *Industries, Firms, and Jobs. Sociological and Economic Approaches*, edited by George Farkas and Paula England, Plenum Press, New York.

Williamson, Oliver E. (1991), 'Comparative Economic Organization: The Analysis of Discrete Structural Alternatives,' *Administrative Science Quarterly*, 36(2), June 1991, pp. 269–292.

Williamson, Oliver E. (1993), 'Calculativeness, Trust, and Economic Organization,' *Journal of Law and Economics*, 36 (1), Pt. 2, April 1993, pp. 453–86.

Notes to chapter 2

* This paper is based on a set of lectures given in the Department of Sociology, University of Oslo on April 29, 1991, on October 9, 1992, and on February 19, 1993. I thank seminar participants and Pål Foss for useful comments and discussions. In particular, I thank Vemund Snartland for research assistance.

** Trond Petersen, University of California, Berkeley and University of Oslo, is a sociologist who does research on reward systems in formal organizations and on quantitative methods of social research.

1 Other accounts of TCE in the sociology literature are Perrow (1986, ch. 7) and Blau and Meyer (1987, ch. 8). See also the interview with Oliver Williamson in Swedberg (1990, ch. 6). See Moe (1984) for an account from the viewpoint of a political scientist. See also, from the viewpoint of economics, the broad account in Milgrom and Roberts (1992, ch. 9), emphasizing in particular Ronald Coase's contribution.

2 The earlier version in Williamson (1975) is discussed in Appendix A and the relationship between the two versions is spelled out there.

3 The causal-chain account of TCE given here should be viewed as a device for exposition of the ideas, how they are related to each other, rather than as depicting a causal structure. Most explanations in TCE are intentional in that behaviour is explained by showing that the actor did what he did for a reason. Williamson (1988, pp. 178-182) argues that one particular TCE explanation also satisfies the criteria for a full functionalist explanation in Elster's (1983, p. 57) sense, namely that for the appearance and diffusion of the multi-divisional form of organizations.

4 In the appendix I discuss at length a case study by Gerlach (1990) that illustrates both the strengths and pitfalls of economic transactions executed under trust rather than fully specified contracts.

5 The treatment in Williamson (1975, chs. 3–4) drew considerably on Doeringer and Piore (1971). See below. It also debated features of Becker's (1975) analysis of firm-specific human capital.

6 Petersen (1992a, 1992b) discusses alternative governance structures to supervision in such teams, in particular, one that relies on a market-type relationship, as in the case of internal spot contract above.

7 Vertical integration differs from *horizontal* integration. In the latter, a firm buys up another firm that is neither a supplier nor a customer, but is a firm producing the same product.

8 Another account which first of all can explain horizontal integration, focuses on *economies of scale*. These occur when the cost per unit of a given product decreases with the number of units produced. In such a situation, a natural monopoly is said to exist. It is usually, or at least historically, thought that telecommunications and postal services are of this kind. In most industries this has been a justification for only having one provider of those kinds of services. Often, when a natural monopoly exists, entry into the relevant market is barred by law, so that the incumbent firm neither faces competition nor the threat of competition. The idea is that a single firm can produce the product at lower costs than two or more firms, due to the economies of scale. Even without legal barriers to entry, it should, if a natural monopoly exists and produces efficiently, not be possible for potential competitors to enter, simply because they will not be able to produce at the same scale and will hence have to charge higher prices. Williamson (1985, pp. 92–95) argues that economies of scale, although important, are not the main reason for the growth of large firms, or vertical integration. Rather, the main reason is the attempt to save on transaction costs.

9 With respect to the analysis of the M-form, Williamson (1985, ch. 11) is more or less identical to Williamson (1975, ch. 8).

10 Much of this is described in detail by Chandler (1962); for du Pont and General Motors in chapters 2 and 3, and for Standard Oil Company and Sears, Roebuck, and Company in chapters 4 and 5.

11 Lorsch (1967) provides an interesting case study that illustrates many of the issues that arise in M-form companies in connection with awarding internal incentives and setting internal transfer prices between divisions of the organization.

12 Fligstein (1985; 1990, esp. ch. 7) presents a different account of why the M-form developed, proposing that it was related to the internal power struggle within large corporations. I discuss his 1985 paper below.

13 In sociology there are several empirical studies that are critical of TCE, among others, Fligstein (1985, 1990), Dore (1986, 1987), Eccles and White (1988), and Hamilton and Biggart (1988). Fligstein (1985) is discussed below, whereas Dore's (1986, 1987) more conceptual points were discussed above.

14 One study, drawing on the sociology literature is that of Russell (1985), who analyzes governance structures and ownership forms for taxi-cab companies, professional firms (e.g., accounting, law), scavenger companies for refuse collection, and other worker-owned organizations.

15 One study addresses employee opportunism directly. Anderson (1988) shows that a salesforce that is in-house, as employees of a company, is less opportunistic than one that is out-house, as representatives, not employees of a company.

16 Powell (1987) makes a similar claim, showing that many transactions between firms have the character neither of market nor hierarchy. Instead, many relationships between firms are closer to those found in Japan, as described by Dore (1986, 1987) and discussed in Section 2.4 above. Powell (1987) calls these relationships for hybrid organizational forms. These are analyzed in Williamson (1991).

17 There are also several critiques in the economics literature. Milgrom and Roberts (1992, pp. 33-35) bring up two issues. The first is that production and transaction costs cannot be as easily separated as claimed in TCE literature. The second issue they discuss is the claim made in the TCE literature that organizations must minimize total transaction costs. Rather, according to Milgrom and Roberts (1992, p. 34), an organization should minimize the categories of costs it has to bear itself.

18 Williamson (1985, pp. 50-52) briefly compares his framework to other frameworks, such as principal-agent theory. He notes that his term opportunism covers both 'moral hazard' and 'adverse selection,' terms used in the principal-agent literature. Moreover, he claims that the analysis of moral hazard is too narrow in that literature.