Abstract

This chapter surveys the new institutional economics, a rapidly growing literature combining economics, law, organization theory, political science, sociology and anthropology to understand social, political and commercial institutions. This literature tries to explain what institutions are, how they arise, what purposes they serve, how they change and how they may be reformed. Following convention, I distinguish between the institutional environment (the background constraints, or ‘rules of the game’, that guide individuals’ behavior) and institutional arrangements (specific guidelines designed by trading partners to facilitate particular exchanges). In both cases, the discussion here focuses on applications, evidence and policy implications.

**JEL classification:** D23, D72, L22, L42, O17

**Keywords:** Institutions, Firms, Transaction Costs, Specific Assets, Governance Structures

1. Introduction

The new institutional economics (NIE) is an interdisciplinary enterprise combining economics, law, organization theory, political science, sociology and anthropology to understand the institutions of social, political and commercial life. It borrows liberally from various social-science disciplines, but its primary language is economics. Its goal is to explain what institutions are, how they arise, what purposes they serve, how they change and how - if at all - they should be reformed. This essay surveys the wide-ranging and rapidly growing literature on the economics of institutions, with an emphasis on applications and evidence. The survey is divided into eight sections: the institutional environment; institutional arrangements and the theory of the firm; moral hazard and agency; transaction cost economics; capabilities and the core competence of the firm; evidence on contracts, organizations and institutions; public policy implications and influence; and a brief summary.

Until recently, ‘institutional economics’ usually referred to the writings of Thorstein Veblen, John R. Commons, Wesley C. Mitchell, Clarence Ayres and their followers. This is a diverse group, but their work reflects several common
themes, mostly criticisms of orthodox economics: (1) a focus on collective rather than individual action; (2) a preference for an 'evolutionary' rather than mechanistic approach to the economy; and (3) an emphasis on empirical observation over deductive reasoning. (For a sampling of the secondary literature see Seckler, 1975; Gruchy, 1972; Gruchy, 1987; Rutherford, 1983; Langlois, 1989; and Hodgson, 1998. On the German roots of American institutionalism, see Richter, 1996.) Whatever their contributions, the older institutionalists are little known to most contemporary economists. Coase's (1984, p. 230) dismissal is typical: ‘Without a theory they had nothing to pass on except a mass of descriptive material waiting for a theory, or a fire’. Still, this tradition (broadly defined) continues in such outlets as the *Journal of Economic Issues*, the *Cambridge Journal of Economics* and the *Review of Political Economy*.


Like its older counterpart, the new institutional economics is interested in the social, economic and political institutions that govern everyday life. However, the new institutional economics eschews the holism of the older school. NIE follows strict methodological individualism, always couching its explanations in terms of the goals, plans and actions of individuals. Of course, NIE appreciates social phenomena like corporate culture, organizational memory, and so on. Still, NIE takes these as *explananda*, not the *explanans*.

NIE differs from mainstream neoclassical economics, however, in insisting that policy analysis be guided by what Coase (1964) calls ‘comparative institutional analysis’. Orthodox welfare analysis typically compares real-world outcomes with the hypothetical benchmark of perfectly competitive general equilibrium. It is unsurprising, then, that actual market outcomes will come up short. The relevant question, Coase explains, is whether a feasible alternative can be devised:

Contemplation of an optimal system may provide techniques of analysis that would otherwise have been missed and, in certain special cases, it may go far to providing a solution. But in general its influence has been pernicious. It has directed economists’ attention away from the main question, which is how alternative arrangements will actually work in practice. It has led economists to derive
conclusions for economic policy from a study of an abstract of a market situation. It is no accident that in the literature ... we find a category ‘market failure’ but no category ‘government failure’. Until we realize that we are choosing between social arrangements which are all more or less failures, we are not likely to make much headway. (Coase, 1964, p. 195)

Coase’s own investigation of British lighthouses (Coase, 1974) is a well-known comparative-institutional study. Coase discovered that before the 1830s, the typical British lighthouse - the classic, textbook example of a ‘public good’, which presumably the market cannot supply and therefore the state must provide - was privately owned and operated. Coase pointed out that here, public ownership does not overcome the ‘free-rider problem’ any better than does private ownership. Thus we should not be surprised to see private entrepreneurs providing this public good - at least until it was nationalized and provided thereafter by the state. (For other examples of public goods that were privately provided, but later nationalized - typically to raise revenue for the sovereign - see Benson, 1994 on police services and public highways, Benson, 1992 on criminal law and Selgin and White, forthcoming on money.)

To organize the various strands of the NIE, it is useful to begin with Davis and North’s (1971) distinction between the ‘institutional environment’ and ‘institutional arrangements’. The former refers to the background constraints, or ‘rules of the game’, that guide individuals’ behavior. These can be both formal, explicit rules (constitutions, laws, property rights) and informal, often implicit rules (social conventions, norms). While these background rules are the product of - and can be explained in terms of - the goals, beliefs and choices of individual actors, the social result (the rule itself) is typically not known or ‘designed’ by anyone. Institutional arrangements, by contrast, are specific guidelines - what Williamson (1985, 1996b) calls ‘governance structures’ - designed by trading partners to mediate particular economic relationships. Business firms, long-term contracts, public bureaucracies, nonprofit organizations and other contractual agreements are examples of institutional arrangements.

2. The Institutional Environment

The institutional environment forms the framework in which human action takes place. ‘Institutions reduce uncertainty by providing a structure to everyday life’, writes North (1990, p. 3). ‘In the jargon of the economist, institutions define and limit the set of choices of individuals. Institutional constraints include both what individuals are prohibited from doing and, sometimes, under what conditions some individuals are permitted to undertake certain activities. ... They are perfectly analogous to the rules of the game in a competitive team sport’ (North, 1990, pp. 3-4). Unlike the rules in team sports, however, these guidelines often arise ‘spontaneously’, as by-products of

*The Legal Environment and Property Rights*

Of these sets of rules, the legal environment has received the most attention. Economists have long been interested in the economic effects of laws (for instance, the effects of a price ceiling on equilibrium price and quantity), but only in the last few decades has economics been applied to the design of legal rules and the legal system itself. Beginning with the early literature on the efficiency of the common law (Rubin, 1977; Priest, 1977), economics has been used to study not only the character and effects of law but the mechanisms by which legal rules change. In this sense, law and economics may therefore be considered a part of NIE, although it is customary to speak of law and economics and NIE as separate movements. (See the exchange between Posner, 1993, and Williamson, 1993, for contrasting views on the relationship between these two literatures. See also Williamson, 1996c, on the relationship between NIE and legal realism.)

NIE has been particularly interested in contract law (Llewellyn, 1931; Macneil, 1974, 1978; Langbein, 1987) and property law (Alchian, 1961; Demsetz, 1967; Furubotn and Pejovich, 1972, 1974; De Alessi, 1980; Barzel, 1989). However, unlike the ‘legal centralism’ tradition, which holds that disputes are primarily settled by the courts as official agents of the state, NIE often focuses on private solutions, holding that ‘in many instances the participants can devise more satisfactory solutions to their disputes than can professionals constrained to apply general rules on the basis of limited knowledge of the dispute’ (Galanter, 1981, p. 4). The recent studies on decentralized law and its evolution by Benson (1990), Ellickson (1991) and Cooter (1994), for example, are examples of this ‘private ordering’ tradition.

*Norms and Social Conventions*

Equally important are the informal and often tacit, rules that structure social conduct. ‘[F]ormal rules ... make up a small ... part of the sum of constraints that shape choices; ... the governing structure is overwhelmingly defined by codes of conduct, norms of behavior and conventions’ (North, 1990, p. 36).

Such rules, once established, form constraints for individual actors. Yet how can the rules themselves be explained in terms of purposeful individual choices? In Menger’s (1883, p. 146) words: ‘How can it be that institutions which serve the common welfare and are extremely significant for its development come into being without a common will directed toward establishing them?’

One approach is to interpret social conventions as noncooperative Nash-equilibrium solutions to a variety of repeated games (‘supergames’) faced by individuals in social settings. An example is the coordination game made famous by Schelling (1960). Two friends arrange to meet one day at 5:00 p.m. in New York City. As the time of the meeting approaches, however, neither can
remember where the meeting was to take place. Furthermore, the friends cannot contact each other to verify the location of the meeting; each must guess, independently, a likely meeting place. What can they do? Obviously, this game has multiple Nash equilibria: any outcome in which both friends choose the same location - say, the corner of 34th Street and 5th Avenue - is a Nash equilibrium to the game. According to Schelling, when faced with this kind of problem, agents rely on cultural information outside the structure of the game. Everyone simply knows, for example, that the logical place to meet in New York City is beneath the clock in the main terminal of Grand Central Station. This equilibrium is what Schelling called a ‘focal point.’ Over time, he argued, behavioral regularities develop so agents can solve these kinds of coordination problems.

Ullman-Margalit (1977) calls these equilibria ‘norms’; Sugden (1986) calls them ‘conventions’; Schotter (1981) calls them ‘social institutions’. In Schotter’s (1981, p. 11) words, a social institution is ‘a regularity in social behavior that is agreed to by all members of society, specifies behavior in specific recurrent situations and is either self-policed or policed by some external authority’. These regularities are presumed to arise over time as agents interact repeatedly. Game theory itself, however, usually says little about how a particular convention is chosen; it only identifies combinations of strategies that are mutual best responses. More recently, some explicitly evolutionary models (Witt, 1989; Wärnereyd, 1990; Boyer and Orlean, 1992) have tried to explain the dynamic process by which particular equilibria are chosen. Axelrod (1984) has shown experimentally that strategies of repeated cooperation tend to be established relatively quickly.

Ellickson (1991) explains that social norms, as ‘customary law’, can be superior to administrative or judicial dispute resolution among people with close social ties. Ellickson studied disputes between cattle ranchers and farmers in Shasta County, California and found that these disputes were usually resolved by appeal to generally accepted social rules, not by bargaining over legal rights (as the Coase Theorem would predict). ‘[M]embers of a close-knit group develop and maintain norms whose content serves to maximize the aggregate welfare that members obtain in their workaday affairs with one another’ (Ellickson, 1991, p. 167). That is, through repeated play, agents tend to converge on strategies of cooperation that improve joint well being. These strategies replace traditional legal remedies. ‘Law solves the problem of cooperation by altering the payoff structure in each game; relationships solve the problem by repeating the game. In Shasta County, where both solutions are available, relationships prevail over law’ (Cooter, 1993, p. 423). Informal norms, in these cases, replace law.

Norms and law are not necessarily substitutes, however. Law can shape the outcome of private bargaining by serving as a backup mechanism for resolving disputes that cannot be resolved privately. If the alternative to private dispute resolution is resolution in court, then the expected outcome at trial determines the parties’ ‘threat values’ in bargaining. Bargaining typically takes place ‘in
the shadow of the law’ (Cooter, Marks and Mnookin, 1982). Moreover, norms can help shape the law, if judges look to social norms as guidelines for legal decisions. The traditional account of the medieval law merchant illustrates this phenomenon. During the commercial revolution merchants developed a system of private courts to resolve disputes among themselves. The rules of these courts became general merchant practice, enforced by the threat of ostracism. As the English legal system developed, judges began to hear commercial disputes once handled privately. In resolving these disputes, English common-law judges tended to enforce the merchant customs already in place. In this way the common law came to embody the principles that already existed, principles developed through private interaction among merchants. (On the law merchant see Trakman, 1983 and Benson, 1989). Today, many commercial disputes are resolved privately, through organizations such as the VISA Arbitration Committee (Solove, 1986; Cooter, 1994).

**Economic History and Economic Growth**

Attention to the institutional environment has become increasingly common in economic history and it has deeply enriched our understanding of how economies develop through time (North and Thomas, 1973; North, 1990; Drobak and Nye, 1997). Economic development is no longer regarded as a gradual, inevitable transformation from local autarky to specialization and the division of labor. Instead, development is seen as a response to the evolution of institutions that support social and commercial relationships. Economic growth thus depends on the degree to which the potential hazards of trade (shirking, opportunism and the like) can be controlled by institutions, which reduce information costs, encourage capital formation and capital mobility, allow risks to be priced and shared and otherwise facilitate cooperation.

In early societies, agency problems were typically solved through kinship or other close social ties. Greif (1989), for example, has shown how eleventh-century Jewish traders in the Mediterranean trade enforced codes of conduct by maintaining close social relationships, using the threat of ostracism as a disciplinary device. Later, standardized weights and measures, units of account, media of exchange and procedures to resolve disputes (such as merchant law courts) supported the expansion of trade by lowering information costs. Capital markets could flourish only in societies where rulers could credibly commit not to expropriate private wealth; North and Weingast (1989) show how capital markets emerged in Britain after the Glorious Revolution of 1688 placed parliamentary limits on the authority of the Crown. The growth of product and factor markets depends similarly on establishing secure property rights. Furthermore, as an economy industrializes, more and more commercial activity involves ‘transacting’: trade, finance, banking, insurance and management (Wallis and North, 1986). Industrialization requires institutions to mitigate the costs associated with these transactions.
Economic development, then, is institutional development. ‘The central issue of economic history and of economic development is to account for the evolution of political and economic institutions that create an economic environment that induces increasing productivity’ (North, 1991, p. 98).

Positive Political Theory
Political institutions have also received much attention in NIE. The rational-choice approach to politics, as outlined in public choice (Buchanan and Tullock, 1962; Mueller, 1979, 1989) and positive political theory (McKelvey, 1976; Riker, 1981; Enelow and Hinich, 1984), holds that political institutions can be explained in terms of purposeful human choice. This framework has been applied to constitutions, legislatures, executives, bureaucracies, courts and elections. Spatial models of voting, for example, show how different voting rules (such as which party can set the agenda) affect the outcome. Among the better-known applications of the spatial model are studies of the committee structure in Congress, under which committees have agenda-setting power (Denzau and Mackay, 1983; Shepsle and Weingast, 1987). The rational-choice perspective is also used to explain the effects of political institutions on public policy, including macroeconomic policy, welfare policy, budgets, regulation and technology policy (see Weingast, 1996, for an overview).

Why, however, do political institutions take one form or another? One approach is to identify particular political institutions that are self-perpetuating, meaning that those individuals or groups which can modify the institution have no incentive to do so (Oreshek, 1993; Weingast, 1995). Antebellum American federalism is one example; it survived, arguably, because it was supported by institutions such as the territorial ‘balance rule’ established by the Missouri Compromise of 1820 (Weingast, 1994). Another approach identifies institutions that allow bureaucrats to make their policy choices last beyond their own tenures (Moe, 1989).

Complexity and Cognitive Science
A few recent papers have focused on the relationship between the institutional environment and cognitive processes in forming a framework for decision making under uncertainty (Denzau and North, 1994; Clark, 1997). Denzau and North (1994) argue that ideology, along with institutions, helps agents cope with complex decisions. They define ideology as a shared set of mental models possessed by groups of individuals. These mental models are ‘the internal representations that individual cognitive systems create to interpret the environment’; institutions are ‘the external (to the mind) mechanisms individuals create to structure and order the environment’ (Denzau and North, 1994, p. 4). Together, ideology and institutions form a framework for economic activity under conditions of uncertainty. If social learning is path-dependent, as they maintain, then economic development will be gradual and uneven. This may explain why some economies continue to perform poorly for long periods.
3. Institutional Arrangements and the Theory of the Firm

These rules and customs that make up the institutional environment are primarily economy-side phenomena. Another aspect of the new institutional economics focuses on agreements made by specific individuals to govern their own relationships. Such institutional arrangements—what Williamson (1996b, p. 5) calls the institutions of governance—include contracts and organizations and in particular, the business firm. The study of governance is more prosaic than the study of the institutional environment. ‘Mundane questions of whether to make or buy a component to be used in the manufacture of an automobile or whether to expand the hospital into outpatient and home health services are ones that arise at the level of governance. By contrast, composite economic growth and income distribution are more apt to be the objects of interest in an inquiry into the institutional environment’ (Williamson, 1996b, p. 5). However, the study of governance—in particular, the theory of the firm—is arguably more developed than the study of the institutional environment.


The Conventional Theory of The Firm

What economists usually mean by ‘the theory of the firm’ is the theory of production, not the theory of the firm as a legal entity. In economics textbooks, the ‘firm’ is a production function or production possibilities set, a ‘black box’ that transforms inputs into outputs. Given technology, input prices and a demand schedule, the firm maximizes money profits subject to the constraint that its production plans must be technologically feasible. The firm is modeled as a single actor, facing a series of straightforward decisions: what level of output to produce, how much of each factor to hire, and so on. Similarly, the firm’s size and product range are usually explained in terms of production costs: economies of scale imply larger firms, while economies of scope justify the multiproduct firm (Spulber, 1989, pp. 113-20).

The conventional theory has proved highly useful for understanding pricing and output decisions and how these vary with competitive conditions. It also has the appeal of analytical tractability along with its elegant parallel to neoclassical consumer theory (profit maximization is like utility maximization, isoquants are indifference curves, and so on). However, the production-function approach provides little insight into the boundaries of the firm. For example, cost subadditivity (as reflected in economies of scale and scope) implies that certain quantities of output can be produced more efficiently when they are
produced together. Yet this does not explain why the joint production must take place in a single firm; absent transactional difficulties, two independent firms could simply contract to share the same plant or facility and jointly produce the efficient level of output (Teece, 1980, 1982). Whether the firms will integrate thus depends on the cost of writing and enforcing contracts, not only on the underlying productive technology.

The black-box model is really a theory about a plant or production process, not a firm. Textbook treatments frequently blur the distinction between firm and plant (for a welcome exception, see Sharkey, 1982, pp. 73-83), but the two are quite distinct. A single firm can own and operate multiple production processes, just as two or more firms can contract to operate jointly a single production process (as in a research joint venture). For this reason, the production-function approach cannot fully explain such real-world business practices as vertical and lateral integration, acquisitions, geographic and product-line diversification, franchising, long-term commercial contracting, transfer pricing and joint ventures, nor is it an adequate guide for antitrust and regulatory policy. Instead, the new institutional economics sees the firm as a set of arrangements - as an organization - itself worthy of economic analysis.

Coase and Transaction Costs

The new institutional approach to the firm is usually traced to Coase’s celebrated 1937 paper on ‘The Nature of the Firm’. Coase was the first to explain that the boundaries of the organization depend not only on the productive technology, but on the costs of transacting business. In the Coasian framework, as developed and expanded by Williamson (1975, 1985, 1996b), Klein, Crawford and Alchian (1978), Grossman and Hart (1986) and Hart and Moore (1990), the decision to organize transactions within the firm as opposed to on the open market - the ‘make or buy decision’ - depends on the relative costs of internal and external exchange. The market mechanism entails certain costs: discovering the relevant prices, negotiating and enforcing contracts, and so on. Within the firm, the entrepreneur can reduce these ‘transaction costs’ by coordinating these activities himself. However, internal organization brings another kind of transaction costs, namely problems of information flows, incentives, monitoring and performance evaluation. More generally, all feasible modes of economic organization incur costs. The nature of the firm, then, is determined by the relative costs of organizing transactions under alternative institutional arrangements.

This transformation of economists’ thinking about the firm is nicely summarized by Roe (1994, p. vii):

Economic theory once treated the firm as a collection of machinery, technology, inventory, workers and capital. Dump these inputs into a black box, stir them up and one got outputs of products and profits. Today, theory sees the firm as more, as a management structure. The firm succeeds if managers can successfully coordinate
the firm’s activities; it fails if managers cannot effectively coordinate and match people and inputs to current technologies and markets. At the very top of the firm are the relationships among the firm’s shareholders, its directors and its senior managers. If those relationships are dysfunctional, the firm is more likely to stumble.

With this new orientation, economic theory is playing an increasingly visible role in finance, accounting, management and other areas once thought to be beyond the purview of economics.

4. Moral Hazard and Agency

The modern theory of the firm comprises several approaches. The moral-hazard or agency-theoretic approach begins with Berle and Means’s (1932) identification of the ‘separation of ownership and control’ in the large firm. The modern corporation, they claimed, is run not by owners (shareholders), but by salaried managers, whose goals often differ from those of the owners. Managers may use their discretion to ‘shirk’ or otherwise pursue personal objectives (firm growth, personal power, entrenchment, perquisites) at the expense of shareholder value.

The Berle-Means account omits the possibility that competition might impose discipline on shirking managers. Product-market competition, competition in the internal market for managers (Fama, 1980) and competition in the market for corporate control (Manne, 1965) all place limits on managerial discretion. Still, their basic model of conflict between shareholders and managers - what we would now call a principal-agent problem - remains a powerful lens for viewing the internal organization of the firm. Agency theory, as developed by Jensen and Meckling (1976), Fama (1980), Fama and Jensen (1983) and Jensen (1986), has become the standard language of corporate finance.

Agency theory studies the design of ex-ante incentive-compatible mechanisms to reduce agency costs in the face of potential moral hazard (malfeasance) by agents. Agency costs are defined by Jensen and Meckling (1976, p. 308) as the sum of ‘(1) the monitoring expenditures of the principal, (2) the bonding expenditures by the agent and (3) the residual loss’. The residual loss represents the potential gains from trade not realized because principals cannot provide perfect incentives for agents when the agents’ actions are unobservable. In a typical agency model, a principal assigns an agent to do some task (producing output, for instance), but has only an imperfect signal of the agent’s performance (for example, effort). The agency problem resembles the signal-extraction problem popularized in macroeconomics by Lucas (1972): how much of the observable outcome (output) is due to the agent’s effort and how much is due to factors beyond the agent’s control? The optimal incentive contract balances the principal’s desire to give the agent incentives to increase effort (for example, by basing compensation on the outcome) with the agent’s
desire to be insured from the fluctuations in compensation that come from these factors beyond his control.

In the agency literature, the firm itself is not the subject of attention. According to Alchian and Demsetz (1972) and Jensen and Meckling (1976), the ‘firm’ is simply a convenient label for the collection of contracts between owners and managers, managers and employees and the firm and its customers and suppliers. The firm is a nexus of a set of contracting relationships. ... The firm is a legal fiction which serves as a focus for a complete process in which the conflicting objectives of individuals ... are brought into equilibrium within a ‘framework of contractual relations’ (Jensen and Meckling, 1976, pp. 311-12).

The question of interest is thus the degree to which various contracts can mitigate these conflicts; the boundary of the firm is a secondary issue.

5. Transaction Cost Economics

Transaction cost economics (TCE) represents another approach to studying institutional arrangements. Here, the emphasis is on governing transactions. TCE holds that all but the simplest transactions require some kind of mechanism - what Williamson (1985) calls a governance structure - to protect the transacting parties from various hazards associated with exchange. The appropriate governance structure depends on the characteristics of the transaction, so TCE implies an applied research program of comparative contractual analysis: how do different forms of governance work in various circumstances? For this reason, TCE (associated mainly with Williamson) is sometimes described as the ‘governance’ branch of the NIE, as opposed to the ‘measurement’ branch (associated with Alchian and Demsetz, 1972).

The governance approach is distinguished by its emphasis on incomplete contracts. In the transaction cost framework, economic organization imposes costs because complex contracts are usually incomplete. A complete contract specifies a course of action, a decision, or terms of trade contingent on every possible future state of affairs. In the textbook model of competitive general equilibrium, all contracts are assumed to be complete. The future is not known with certainty, but the probability distributions of all possible future events are known (what Knight, 1921, would call ‘risk’ rather than ‘uncertainty’). In an important sense, the model is ‘timeless’: all relevant future contingencies are considered in the ex ante contracting stage, so there are no decisions to be made - no actions to be taken at all, really - as the future plays itself out.

TCE relaxes this assumption and holds that all complex contracts are unavoidably incomplete. In a world of ‘true’ uncertainty, the future holds genuine surprises and this limits the available contracting options. In simple transactions - for instance, procuring an off-the-shelf component - uncertainty may be relatively unimportant and spot-market contracting works well. For more complex transactions, such as the purchase and installation of specialized
equipment, a more sophisticated contract is needed. However, such a contract will typically be incomplete - it will provide remedies for only some possible future contingencies. One example is a relational contract, an agreement that describes shared goals and a set of general principles that govern the relationship (Goldberg, 1980). Another is an implicit contract - an agreement that while unstated, is assumed to be understood by all sides.

Williamson attributes contractual incompleteness to cognitive limits or ‘bounded rationality’, following Simon’s (1961, p. xxiv) interpretation of human action as ‘intendedly rational, but only limitedly so’. Other NIE economists are more agnostic, assuming only that some quantities or outcomes are unobservable (or not verifiable to third parties, such as the courts), in which case contracts cannot be made contingent on these variables or outcomes.

Specific Investments and the Holdup Problem
Contractual incompleteness exposes the contracting parties to certain risks. Primarily, if circumstances change unexpectedly, the original governing agreement may no longer be effective. The need to adapt to unforeseen contingencies constitutes an additional cost of contracting; failure to adapt imposes what Williamson (1991a) calls ‘maladaptation costs’. The most-often-discussed example of maladaptation is the ‘holdup’ problem associated with relationship-specific investments. Investment in such assets exposes agents to a potential hazard: If circumstances change, their trading partners may try to expropriate the rents accruing to the specific assets. Suppose an upstream supplier tailors its equipment for a particular customer. After the equipment is in place, the customer may demand a lower price, knowing that the salvage value of the specialized equipment is lower than the net payment it offers. This creates an underinvestment problem: Anticipating the customer’s behavior, the supplier will be unwilling to install the custom machinery without protection for such a contingency, even if the specialized technology would make the relationship more profitable for both sides.

One way to safeguard rents accruing to specific assets is vertical (or lateral) integration, where a merger eliminates any adversarial interests. Less extreme options include long-term contracts (Joskow, 1985, 1987, 1988, 1990), partial ownership agreements (Pisano, Russo and Teece, 1988; Pisano, 1990), or agreements for both parties to invest in offsetting relationship-specific investments (Heide and John, 1988). Overall, several governance structures may be employed. TCE holds that parties tend to choose the governance structure that best controls the underinvestment problem, given the particulars of the relationship.

The holdup problem is the best-known example of a contractual hazard. More generally, contractual difficulties can arise from several sources: (1) bilateral dependence; (2) weak property rights; (3) measurement difficulties and/or oversearching; (4) intertemporal issues that can take the form of disequilibrium contracting, real-time responsiveness, long latency and strategic
abuse; and (5) weaknesses in the institutional environment’ (Williamson, 1996b, p. 14). Each of these has the potential to impose maladaptation costs. Foreseeing this possibility, agents seek to reduce the potential costs of maladaptation by matching the appropriate governance structure with the particular characteristics of the transaction.

In this way, TCE may be considered the study of alternative institutions of governance. Its working hypothesis, as expressed by Williamson (1991c, p. 79), is that economic organization is mainly an effort to ‘align transactions, which differ in their attributes, with governance structures, which differ in their costs and competencies, in a discriminating (mainly, transaction cost economizing) way’. Simply put, TCE tries to explain how trading partners choose, from the set of feasible institutional alternatives, the arrangement that protects their relationship-specific investments at the least cost.

Institutions as Governance Structures

Transactions differ in several ways: the degree to which relationship-specific assets are involved, the amount of uncertainty about the future and about other parties’ actions, the complexity of the trading arrangement and the frequency with which the transaction occurs. Each matters for the preferred institution of governance, although the first - asset specificity - is particularly important. Williamson (1985, p. 55) defines asset specificity as ‘durable investments that are undertaken in support of particular transactions, the opportunity cost of which investments is much lower in best alternative uses or by alternative users should the original transaction be prematurely terminated’. This could describe a variety of relationship-specific investments, including both specialized physical and human capital, along with intangibles such as R&D and firm-specific knowledge or capabilities.

Governance structures can be described along a spectrum, with ‘market’ and ‘hierarchy’ at the poles. At one end lies the pure anonymous spot market, which suffices for simple transactions such as basic commodity sales. Market prices provide powerful incentives for exploiting profit opportunities and market participants are quick to adapt to changing circumstances as information is revealed through prices. When relationship-specific assets are at stake, however and when product or input markets are thin, bilateral coordination of investment decisions may be desirable and combined ownership of these assets may be efficient. At the other end of the spectrum from the simple, anonymous spot market thus lies the fully integrated firm, where trading parties are under unified ownership and control. TCE posits that such hierarchies offer greater protection for specific investments and provide relatively efficient mechanisms for responding to change where coordinated adaptation is necessary. Compared with decentralized structures, however, hierarchies provide managers with weaker incentives to maximize profits and normally incur additional bureaucratic costs. Between the two poles of market and hierarchy are a variety of ‘hybrid’ modes, such as complex contracts and partial ownership arrangements. The movement from market to hierarchy thus
entails a tradeoff between the high-powered incentives and adaptive properties of the market and the safeguards and central coordinating properties of the firm.

The general theoretical framework of TCE is now sufficiently accepted to have been incorporated in several textbook treatments (Kreps, 1990, pp. 744-90; Rubin, 1990; Milgrom and Roberts, 1992; Acs and Gerlowski, 1996; and Besanko, Dranove and Shanley, 1990).

6. Capabilities and the Core Competence of the Firm

An alternative approach to explaining institutional arrangements focuses on the ‘core competence’ or ‘capabilities’ of the firm. The capabilities view, which traces its roots to Alfred Marshall and Joseph Schumpeter, was first stated explicitly by Edith Penrose in her *Theory of the Growth of the Firm* (1959). It has been developed further by Teece (1980, 1982) and by Nelson and Winter in their *Evolutionary Theory of Economic Change* (1982). The capabilities view regards the firm not as a nexus of contracts (as in Alchian and Demsetz, 1972), or as a set of residual control rights (as in Grossman and Hart, 1986), but as a stock of knowledge. The firm’s capabilities depend on the tacit knowledge it contains, as manifested in organizational memory or routines. This knowledge is considered technologically inseparable; it is firm-specific, not transaction-specific and thus helps determine the boundary of the firm (Chandler, 1992). In this sense, ‘firms exist because they are superior institutional arrangements for accumulating specialized productive knowledge, quite independently of considerations of opportunism, incentive alignment and the like’ (Foss, 1996, p. 2).

In the field of strategic management, organizational capabilities have been examined from within the ‘resource-based’ view of the firm. In the resource-based view, competitive advantage comes from having unique factors of production, resources that are not easily imitable or transferable (and thus cannot be purchased in factor markets). Excess profits or supernormal returns are seen as rents accruing to these unique resources (Rumelt, 1984; Wernerfelt, 1984). Firm-specific resources may include organizational capabilities, managerial skill, technological innovation and reputational capital. When firms have excess capacity in these unique factors, they expand and diversify into product lines whose manufacture employs similar capabilities or routines. The resource-based view is thus concerned as much about economic change, or ‘evolution’, as it is about the design and use of optimal contracts. For this reason, the capabilities approach is often associated with the ‘evolutionary’ theory of the firm. (See Langlois and Robertson, 1995, for an overview.)

The capabilities perspective is intriguing and offers many useful insights into firm organization and behavior. However, research in this area often proceeds at a very high level of abstraction. While there have been several applied studies (Kogut and Chang, 1991; Langlois and Robertson, 1989;
Langlois, 1992a, 1992b; Teece et al., 1994; Argyres, 1996) further empirical work is needed to develop the economics of firm capabilities.

7. Evidence on Contracts, Organizations and Institutions

The same criticism has been leveled at the theory of the firm more generally. Simon (1991, p. 27), for example, has charged the new institutional economics and transaction cost economics in particular, with lacking sufficient empirical support. Until the relevant empirical studies have been done, he says, ‘the new institutional economics and related approaches are acts of faith, or perhaps of piety’. However, much empirical work has already been carried out. (Shelanski and Klein, 1995, provide a comprehensive survey; Masten, 1996, collects many of the important articles.) On balance, a remarkable amount of this empirical work is consistent with TCE - much more so, perhaps, than is the case with most of industrial organization (Joskow, 1991, p. 81). As Williamson (1996a, p. 55) puts it: ‘TCE is an empirical success story’.

Much of the empirical research in TCE follows the same basic model. The efficient form of organization for a given economic relationship - and, therefore, the likelihood of observing a particular organizational form or governance structure - is a function of certain properties of the underlying transaction or transactions: asset specificity, uncertainty, complexity and frequency. Organizational form is the dependent variable, while asset specificity, uncertainty, complexity and frequency are independent variables. Specifically, the probability of observing a more integrated governance structure depends positively on the amount or value of the relationship-specific assets involved and, for significant levels of asset specificity, on the degree of uncertainty about the future of the relationship, on the complexity of the transaction and on the frequency of trade.

Empirical work in TCE implicitly assumes that market forces work to cause an ‘efficient sort’ between transactions and governance structures, so that exchange relationships observed in practice can be explained in terms of transaction cost economizing. Williamson (1988, p. 174) acknowledges this, while recognizing that the process of transaction cost economizing is not automatic:

The [transaction cost] argument relies in a general, background way on the efficacy of competition to perform a sort between more and less efficient modes and to shift resources in favor of the former. This seems plausible, especially if the relevant outcomes are those that appear over intervals of five and ten years rather than in the very near term. This intuition would nevertheless benefit from a more fully developed theory of the selection process. Transaction cost arguments are thus open to some of the same objections that evolutionary economists [for example, Nelson and Winter] have made of orthodoxy.
Still, he maintains that the efficiency presumption is reasonable, offering the argument - analogous to Friedman’s famous (1953) statement on the selection process - that inefficient governance arrangements will tend to be discovered and undone. Concerning vertical integration, for example, Williamson (1985, pp. 119-20) writes that ‘backward integration that lacks a transaction cost rationale or serves no strategic purposes will presumably be recognized and will be undone’, adding that mistakes will be corrected more quickly ‘if the firm is confronted with an active rivalry’. Silverman, Nickerson and Freeman (1997) have shown that transaction cost efficiency is positively correlated with firm survival in the for-hire trucking industry, though evidence from other industries is scant.

Despite such concerns, most empirical literature inspired by TCE takes as given an economizing framework, assuming that we can draw inferences about the efficiency of organizational forms by observing what organizations actually do. Unlike earlier traditions in industrial organization, which presumed that complex contracts and similar deviations from perfect competition are usually attempts to gain monopoly power, TCE follows the common-law presumption that such contracts ‘serve affirmative economic purposes’ (Williamson, 1985, p. 200). Furthermore, such contracts are objectionable only if some feasible alternative exists (Coase, 1964).

Organizational form is often modeled as a binary variable - ‘make’ or ‘buy’, for example - though it can sometimes be represented by a continuous variable. Of the independent variables, asset specificity is the most difficult to measure. Williamson (1991a) distinguishes among six types of asset specificity. The first is site specificity, in which parties are in a ‘cheek-by-jowl’ relationship to reduce transportation and inventory costs and assets are highly immobile. The second, physical asset specificity, refers to relationship-specific equipment and machinery. The third is human asset specificity, describing transaction-specific knowledge or human capital, achieved through specialized training or learning-by-doing. The fourth is brand-name capital, reflected in intangible assets reflected in consumer perceptions. The fifth is ‘dedicated assets’, referring to substantial, general-purpose investments that would not have been made outside a particular transaction, the commitment of which is necessary to serve a large customer. The sixth is temporal specificity, describing assets which must be used in a particular sequence.

Among the common empirical proxies for asset specificity are component ‘complexity’, qualitatively coded from survey data, as a proxy for physical asset specificity (Masten, 1984); worker-specific knowledge, again coded from survey data, as a proxy for human asset specificity (Monteverde and Teece, 1982); physical proximity of contracting firms, as a proxy for site specificity (Joskow, 1985, 1987, 1988, 1990; Spiller, 1985); and R&D expenditure, as a proxy for physical asset specificity. Other proxies, such as fixed costs or ‘capital intensity’, have more obvious limitations and are rarely used.
Vertical Integration

Vertical integration, or the ‘make-or-buy’ decision, was the first topic studied extensively within the TCE framework. Traditionally, economists viewed vertical integration as an attempt to earn monopoly rents by gaining control of input markets or distribution channels. But in the early 1980s a few authors began to investigate transaction-cost (that is, efficiency) rationales for vertical integration. Monteverde and Teece (1982) made one of the first systematic efforts to test a contractual interpretation of vertical integration. They examined the effects of asset specificity, defined as worker-specific knowledge or ‘applications engineering effort’, on the decision to produce components in-house or to obtain them from outside suppliers. They found applications engineering effort to be a statistically significant determinant of backwards integration. The results are consistent with case-study evidence from Globerman (1980) on firm-specific technical knowledge and integration in the Canadian telecommunications industry. Globerman studied evidence from public hearings and found a tendency toward common ownership of telephone lines and equipment as the research and development demands of a carrier on its equipment suppliers become more complex and uncertain and require more relationship-specific investments.

Other studies of component procurement have found similar support for transactional explanations of vertical relationships. Two studies by Walker and Weber (1984, 1987) focus on uncertainty as a determinant of vertical integration in the auto industry. Like Monteverde and Teece, they worked with a list of automobile components, coded as made or bought, as the dependent variable. They found that greater uncertainty about production volume raises the probability that a component is made in-house, but that ‘technological uncertainty’, measured as the frequency of changes in product specification and the probability of technological improvements, has little effect. Their second (1987) study included measures of market competition, testing the interactive effects of both uncertainty in production and competition among suppliers and added the qualification that volume uncertainty matters only when supply markets are thin.

In a further refinement, Masten, Meehan and Snyder (1989) distinguished among types of specific assets, comparing the relative importance of relationship-specific human and physical capital. They also studied automobile component production, finding that engineering effort, as a proxy for human asset specificity, appears to affect the integration decision more than physical or site specificity. Klein (1988), in a discussion of the G.M.-Fisher Body case, also suggests that specific human capital in the form of technical knowledge was a major determinant of G.M.’s decision to buy out Fisher.

The relationship between G.M. and Fisher Body in the 1920s is a frequently discussed application of TCE. Both Klein, Crawford and Alchian (1978) and Williamson (1985, pp. 114-15) explain G.M.’s buyout of Fisher in terms of the specific physical assets that accompanied the switch from wooden-
metal-bodied cars. The account in Klein (1988) is somewhat different, emphasizing specific human capital. Langlois and Robertson (1989) also criticize the earlier TCE account of the G.M.-Fisher relationship, arguing that systemic uncertainty, rather than asset specificity, accounts for the failure of long-term contracting there. Helper, MacDuffie and Sabel (1997) propose another alternative: G.M. acquired Fisher to promote collaborative learning, not to avoid hold-up. Obviously, this case continues to stimulate interest.

Other studies have documented a similar link between integration and R&D, which usually involves specific human capital (Armour and Teece, 1980; Joskow, 1985; Pisano, 1990). Site specificity, dedicated assets and the need for specifically tailored products or production facilities have been shown to increase vertical integration in a variety of industries, including electricity generation (Joskow, 1985), aerospace (Masten, 1984), aluminum (Stuckey, 1983; Hennart, 1988), forestry (Globerman and Schwindt, 1986), chemicals (Lieberman, 1991) and offshore oil gathering (Hallwood, 1991).

These papers are case studies of particular industries or production processes. As such, they avoid the problem of inconsistent measurement across industries, but have measurement difficulties of their own. The classification of dichotomous variables like ‘make-or-buy’, for example, is typically based on survey data, requiring more discretion by the researcher than economists are comfortable with. Nonetheless, most of the empirical work in TCE on vertical integration has been of this type. While generalizing the results is of course difficult, the cumulative evidence from different studies and industries is quite consistent with the basic theory. Also, there do exist some cross-sectional studies on transactional determinants of vertical integration using multi-industry data and most have been supportive (Levy, 1985; MacMillan, Hambrick and Pennings, 1986).

Long-term Contracts and ‘Hybrid’ Forms

Long-term contracts can be interpreted as intermediate or ‘hybrid’ forms of organization, neither market nor hierarchy. A series of papers by Joskow (1985, 1987, 1988, 1990) investigates the effects of asset specificity on contract duration and price adjustment in agreements between coal suppliers and coal-burning electrical plants. He examined a large sample of coal contracts and found that contracts tended to be longer, all else equal, when relationship-specific investments (here, site specificity and dedicated assets) are at stake. Crocker and Masten (1988) found the same result for the natural gas industry. More generally, they argue that efficient contract duration depends on the costs of contracting; contract terms become shorter, for example, as uncertainty increases. Goldberg and Erickson (1987) analysed contracts for petroleum coke and concluded that many provisions of the contracts can best be interpreted as efforts by the parties to protect themselves against expropriation of specialized investments. Other relevant studies on natural gas contracts have been done by Crocker and Masten (1991) and Hubbard and Weiner (1991).
DeCanio and Frech (1993) tried to measure more precisely the efficiency gains from long-term contracts in natural gas. Relationship-specific investments are critical for transactions between wellhead owners and pipelines. For that reason, ‘take-or-pay’ contracts, in which the buyer must pay for some minimum quantity even if delivery is not taken, are often used to safeguard against buyer (pipeline) opportunism. In 1987, the Federal Energy Regulatory Commission (FERC) outlawed take-or-pay contracts. The authors used data from before and after the FERC order to test its effect on spot gas prices and prices at the wellhead. They found that FERC’s interference with parties’ ability to craft long-term governance mechanisms raised natural gas prices between 21 percent and 31 percent in the year following FERC’s order. The results support TCE explanations for the relative efficiency of long-term contracts where asset specificity is required, while representing an effort to quantify that efficiency gain. (Mulherin, 1986, and Masten and Crocker, 1985, also examine ‘take-or-pay’ contracts.)

Another hybrid form of organization is a partial ownership agreement or ‘equity linkage’. Pisano (1990) asks why firms may rely on equity linkages instead of contracts to support certain transactions. He argues that partial ownership will dominate contractual governance when a relationship involves uncertainty, transaction-specific capital and other variables. He hypothesizes that equity linkages are more likely when R&D is to be done during collaboration and when collaboration encompasses multiple projects and less likely when there are more potential collaborators. His study of collaborative arrangements in the biotechnology industry supported all these claims. Pisano, Russo and Teece (1988) applied a similar analysis to the telecommunications equipment business and found that the same basic framework can explain the choice between equity linkages and other forms of cooperative ventures (joint ventures, consortiums, or other non-equity linkages).

Allen and Lueck (1993) studied ‘cropshare’ contracts between farmers and landowners. Such contracts specify sharing rules for both inputs and outputs; in doing so, they pool enforcement costs by making both the farmer and the landowner residual claimants. This reduces the farmer’s incentive to deplete the capital value of the soil. Allen and Lueck show that optimal sharing rules will involve either full payment of inputs by farmers or sharing input costs in proportion to the output-sharing rule. Nee (1992) studied hybrid governance structures in China’s transitional economy such as small, family-owned firms run by peasant entrepreneurs (‘cadre-entrepreneurs’) and collectively owned enterprises leased to private operators (‘marketized firms’). On hybrids see also Gallick (1984), Masten and Snyder (1993), Lafontaine and Masten (1995) and Menard (1996).

Informal Agreements
Like other parts of the new institutional economics, TCE pays special attention to the importance of private solutions to resolve disputes, in contrast to the
older tradition of legal centralism. Several studies have investigated whether informal trade arrangements, which are not legally enforceable, may also be motivated by the desire to make exchange more efficient. Important work in this area has been done by Palay (1984, 1985). In two closely related papers, Palay studied the role of informal, legally unenforceable agreements between rail-freight carriers and shippers. He argues that ICC regulation of the industry, which prohibits vertical integration of carriers and shippers, was geared to ‘classical contracting’ (Macneil, 1978) but is inappropriate for transactions requiring more complex agreements. Shipment of items like automobile parts and chemicals, for example, requires specially designed rail cars and equipment that cannot be easily redeployed for other uses. Palay claims that informal agreements, substituting for combined ownership, would emerge both to encourage and to protect these relationship-specific investments. Furthermore, he argues that the underlying characteristics of a transaction predict whether it will be supported by an informal agreement. Evidence from case studies of shipper-carrier transactions reveals a pattern of informal agreements highly consistent with TCE. Equipment tailored for particular users - custom carrier racks for automobile parts, tank and covered hopper cars for specific volatile chemicals, and so on - was owned by individual shippers. For more standardized shipments, these would be owned by rail carriers. The informal agreements also provided handling procedures for unusual circumstances related to shipment. The transactions that did not use informal contracting all involved non-specialized capital such as standard box cars. All of this suggests the importance of asset specificity for complex contracting.

Two studies of New England fishing industries also examined the role of transaction costs in determining trade agreements and market structure. Wilson (1980) conducted an intensive study of the New England fresh-fish market. He found that underlying the smooth functioning of the market was a system of mutual dependence created by the particular trade arrangements there; reputation effects provided an enforcement mechanism. Acheson’s (1985) study of the Maine lobster market reached similar conclusions, finding the lobster market to be characterized by long-term, informal relationships between fishermen and lobster-pound operators. Fisherman and pound operators typically crafted agreements to reduce the costs of information and the possibility of opportunistic use of informational asymmetries. The agreements were reinforced by reputation considerations and interdependencies arising from sharing scarce resources, such as market information, boat fuel and bait. Informal agreements and norms in eighteenth- and nineteenth-century whaling have been studied similarly by Ellickson (1989) and Gifford (1993).

Finally, in an interesting application of TCE to the context of personal relationships, Brinig (1990) employed transaction cost reasoning to explain the sudden increase in the demand for diamond engagement rings in the mid-1930s. The increase, she argues, can be traced to the abolition in several states of the ‘breach of promise to marry action’ around the same time. Before
this action was abolished, a broken engagement could trigger a lawsuit, because a woman in this situation faced considerable loss of reputation. Once the cause of action was eliminated, however, another arrangement was needed to ensure the credibility of the marriage commitment. Diamond engagement rings filled that role. In this way, rings may be seen as a governance structure: They safeguard the future bride’s relationship-specific investment - her good reputation.

Franchise Contracting
Williamson’s (1976) case study of the Oakland, California Cable TV (CATV) franchise was an early empirical study using transactional reasoning. Responding to the Posner-Demsetz argument that competitive bidding for monopoly franchises would result in competitive prices, Williamson claimed that once idiosyncratic investments are in place, what was a large-numbers bargaining situation during the bidding process is transformed into a bilateral monopoly. Because of this change, the terms of the original contract may no longer be suitable. Williamson outlined the difficulties faced by Oakland in the early 1970s over its CATV franchise. The franchise was awarded to the lowest bidder in 1970. After the franchise was awarded, however, the construction process went more slowly than expected, fewer households signed up than predicted and costs escalated. Consequently, the franchisee requested a renegotiation of the contract. A complex dual-source agreement was eventually reached, but this outcome was far different from that specified in the initial agreement.

Two later studies of CATV have looked for similar problems, with mixed results. Zupan (1989) examined a series of public cable franchise agreements, comparing the terms of trade struck during the original franchise agreement with those prevailing at the time of renewal, after relationship-specific investments had been made; he found no significant differences in those terms. Prager (1990), however, found that opportunistic behavior by the franchisee, as perceived by cable customers, was higher for franchises awarded through competitive bidding.

Of course, it is not always the franchisee who is opportunistic; the franchiser may be as well. Grandy’s (1989) examination of nineteenth-century railroad regulation in New Jersey found that the railroads in that state were willing to make large specialized investments only when they were protected by ‘special corporation charters’ limiting state action against them. Levy and Spiller’s (1994) comparative study of telecommunications regulation in Argentina, Chile, Jamaica, the Philippines and the UK shows that private investment is forthcoming only when regulators can commit not to pursue arbitrary administrative actions. Furthermore, many private franchise contracts can also be explained through TCE (Norton, 1989; Dnes, 1992).

Besides these contractual phenomena, TCE has been brought to bear on such diverse topics as labor market contracts and regulation (Barker and Chapman, 1989), tie-ins and ‘block booking’ (Kenney and Klein, 1983),
international trade and the multinational corporation (Yarbrough and Yarbrough, 1987; Gatignon and Anderson, 1988; Hennart, 1989; Klein, Frazier and Roth, 1990), company towns and company stores (Fishback, 1986, 1992), land tenure agreements (Roumasset and Uy, 1980; Alston and Higgs, 1982; Alston, Datta and Nugent, 1984; Datta, O’Hara and Nugent, 1986) and even indentured prostitution (Ramseyer, 1991). These and other ‘non-standard’ contracting practices, when viewed through a transaction cost lens, often turn out to have efficiency properties, particularly in offering safeguards for specific investments.

Is the Empirical Evidence Reliable?
The discussion here presents only a sampling of the empirical literature on contracts, organizations and institutions (see the Appendix to Shelanski and Klein, 1995, for a more complete list). While the vast majority of these studies are consistent with transaction cost reasoning, some difficulties remain. Besides the measurement difficulties discussed above, empirical research on the institutions of governance is often hampered by confusion about the definitions of and therefore the empirical proxies for, key variables, especially uncertainty. Asset specificity has been more successfully treated in the empirical literature; relationship-specific physical, site and human capital investments have all been studied, both independently and comparatively. However, further refinement and analysis need to be done here, particularly concerning measurement. 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 concern is that asset-specificity effects may be confused with market power. While specific investment may lead to bilateral monopoly, a small-numbers bargaining situation is not by itself evidence of relationship-specific investment.

Besides these difficulties of measurement and definition, empirical research on the institutions of governance is also subject to the problem found in empirical work generally: alternate hypotheses that could also fit the data are rarely stated and compared. Usually, the data are found only consistent or inconsistent with the hypothesis at hand. Undoubtedly, studies that explicitly compare competing, observationally distinct hypotheses about contractual relationships are needed, because rival theories commonly posit mutually exclusive outcomes. One example is Spiller’s (1985) comparison of asset-specificity and market-power explanations for vertical mergers, explanations that have rival predictions about the size of the gains from mergers under various competitive conditions. Another prototype for such a project might be MacDonald’s (1985) cross-sectional study of vertical integration, which incorporated elements of both TCE and Stigler’s theory of the vertical ‘life-cycle’ of the firm (though it did not attempt to distinguish between them). Poppo and Zenger (1997) compare transaction-cost and resource-based explanations for the make-or-buy decision, finding greater empirical support for the former.
A more general concern is that most of the empirical studies discussed here establish correlations, not causal relations, between asset specificity and internal governance. These studies typically test a reduced-form model where the probability of observing a more hierarchical form of governance increases with the degree of relationship-specific investments. Plausibly, if the presence of such investments reduces the costs of internal organization, then asset specificity could lead to integration, independent of the holdup problem or other maladaptation costs (Masten, 1994, p. 10). Masten, Meehan and Snyder (1991) attempt to distinguish these two effects in the context of human capital. They find that specific human capital investments appear to reduce internal governance costs more than they increase market governance costs. Further studies of this type would be valuable in assessing the implications of the evidence for the reduced-form version of the basic theory. However, we do not yet have a general theory of how relationship-specific assets might reduce the costs of internal organization. By contrast, the underinvestment problem associated with specific assets and market governance is fairly well understood.

8. Public Policy Implications and Influence

Theoretical and empirical research in the NIE has strong implications for antitrust, regulation and other aspects of public policy. This is particularly true for the studies of institutional arrangements discussed in the previous section. A basic conclusion of transaction cost economics is that vertical mergers, even when there are no obvious technological synergies, may enhance efficiency by reducing governance costs. Hence Williamson (1985, p. 19) takes issue with what he calls the ‘inhospitality tradition’ in antitrust - namely, that firms engaged in non-standard business practices like vertical integration, customer and territorial restrictions, tie-ins, franchising, and so on, must be seeking monopoly gains. In the ten years between the celebrated Schwinn (1967) and GTE-Sylvania (1977) cases, Williamson argues, economists began to incorporate transaction cost considerations into their understanding of vertical restrictions. This change in the intellectual climate was reflected in the Supreme Court’s reversal in GTE-Sylvania of its earlier position that vertical restraints are necessarily anticompetitive.

Joskow (1991, pp. 79-80) points out that this change may reflect sensitivity to claims that vertical integration and restraints need not reduce competition, rather than to claims that such arrangements provide contractual safeguards. While the NIE argued that nonstandard business practices may reduce transaction costs, Chicago-school writers like Posner, Peltzman and Bork were maintaining that such practices do not necessarily result in reduced competition. Of course, these arguments are largely complementary. Moreover, the Chicago position on vertical restraints relies largely (though not explicitly) on transaction-cost reasoning (Meese, 1997). In this sense, NIE has played an
important role in recent changes in antitrust enforcement, even if its contribution has not always been recognized.

NIE and transaction cost economics in particular, also has direct implications for many other contracting practices and regulations, though it does not yet appear to have influenced those areas. Barker and Chapman (1989) argue, for example, that closed-shop agreements in labor markets may serve to protect workers’ job-specific training rather than to exploit a monopoly position. They attack New Zealand’s ‘blanket coverage clause’, which effectively prohibits the closed shop, supporting their claims with arguments based on TCE. Studies of optimal contract design such as Crocker and Reynolds’s (1993) examination of Air Force procurement contracts are also relevant as a guide to public policy toward government purchases of goods and services. Other contracts between government agencies and private firms, such as franchise contracts for the provision of public utilities (like cable TV), can be evaluated using TCE reasoning. TCE also points out how the potential for opportunism by the state affects private incentives to make specific investments (Levy and Spiller, 1994). This is particularly important for economic and political reform in the former communist countries, where the need to provide incentives for private investment is paramount.

9. Summary

Speaking of Lionel Robbins’s influential *Nature and Significance of Economic Science* (1932), Coase (1992, p. 714) remarked that ‘in Robbins’s view, an economist does not interest himself in the internal arrangements within organizations but only in what happens on the market’. Even Coase himself believed, as late as 1988, that ‘[w]hy firms exist, what determines the number of firms, what determines what firms do ... are not questions of interest to most economists’ (Coase, 1988, p. 5).

Today, this is clearly no longer the case. The preceding survey provides a brief (and admittedly unbalanced) sketch of the new institutional economics. The literature in NIE is expanding rapidly and gaining increasing adherents and influence in economics, political science, law, strategy, sociology, growth and development, history and other disciplines. It is a highly diverse field and its many branches are rich in theoretical insight, relevant for policy and empirically useful.

Acknowledgements

I am grateful to Nicholas Argyres, John Drobak, Richard Langlois, Jackson Nickerson, David Robinson and Oliver Williamson for helpful suggestions and
clarifications. This paper draws on material in Shelanski and Klein (1995) and Klein and Shelanski (1996).

Bibliography on New Institutional Economics (0530)

Benson, Bruce L. (1990), The Enterprise of Law: Justice Without the State, San Francisco, Pacific Research Institute for Public Policy.


Other References


Knight, Frank H. (1921), Risk, Uncertainty, and Profit, New York, Hart, Schaffner & Marx.

Langlois, Richard N. (1989), ‘What was Wrong with the ‘Old’ Institutional Economics? (And What is Still Wrong with the ‘New’)?’, 1 Review of Political Economy, 272-300.


