
The Modern Corporation: Origins, Evolution, Attributes

Author(s): Oliver E. Williamson

Source: *Journal of Economic Literature*, Dec., 1981, Vol. 19, No. 4 (Dec., 1981), pp. 1537-1568

Published by: American Economic Association

Stable URL: <https://www.jstor.org/stable/2724566>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Economic Literature*

JSTOR

The Modern Corporation: Origins, Evolution, Attributes

By OLIVER E. WILLIAMSON

University of Pennsylvania

This paper has benefitted from the very helpful comments of Moses Abramovitz, Alfred Chandler, Sanford Grossman, Paul Joskow, Scott Masten, Richard Nelson, and Douglass North. Parts of it were given at Rice University as a 1981 Peterkin Lecture, and comments of the faculty and students in attendance were also helpful. For related recent assessments of the modern corporation which, however, emphasize somewhat different aspects, see Richard Caves (1980), Robin Marris and Dennis Mueller (1980), and Richard Cyert and Charles Hedrick (1972).

THERE IS VIRTUAL UNANIMITY with the proposition that the modern corporation is a complex and important economic institution. There is much less agreement on what its attributes are and on how and why it has successively evolved to take on its current configuration. While I recognize that there have been numerous contributing factors, I submit that the modern corporation is mainly to be understood as the product of a series of organizational innovations that have had the purpose and effect of economizing on transaction costs.

Note that I do not argue that the modern corporation is to be understood exclusively in these terms. Other important factors include the quest for monopoly gains and the imperatives of technology. These mainly have a bearing on market shares and on the absolute size of specific technological units; but decisions to make or buy,

which determine the distribution of economic activity, as between firms and markets, and the internal organization (including both the shape and the aggregate size) of the firm are not explained, except perhaps in trivial ways, in these terms. Inasmuch as these are core issues, a theory of the modern corporation that does not address them is, at best, seriously incomplete.

Specifically, the study of the modern corporation should actively concern itself with and provide consistent explanations for the following features of the organization of economic activity: What are the factors that determine the degree to which firms integrate—in backward, forward, and lateral respects? What economic purposes are served by the widespread adoption of divisionalization? What ramifications, if any, does internal organization have for the long-standing

dilemma posed by the separation of ownership from control? Can the "puzzle" of the conglomerate be unravelled? Do similar considerations apply in assessing multinational enterprise? Can an underlying rationale be provided for the reported association between innovation and direct foreign investment?

It is my contention that transaction cost economizing figures prominently in explaining these (as well as related) major features of the business environment. Since transaction-cost economizing is socially valued, it follows that the modern corporation serves affirmative economic purposes. But complex institutions often serve a variety of purposes—and the corporation can and sometimes is used to pursue antisocial objectives. I submit, however, that (1) objectionable purposes can normally be recognized and dealt with separately and (2) failure to understand the main purposes of the corporation has been the source of much confusion and ill-conceived public policy.¹ Specifically, antisocial purposes have often been attributed where none existed.

Inasmuch as a sensitivity to transactions and transaction-cost economizing can be traced to the 1930s (John Commons, 1934; Ronald Coase, 1937), it is somewhat surprising that the importance of the modern corporation as a means of reducing transaction costs has been so long neglected. The main reason is that the origins of transaction costs must often be sought in influences and motives that lie outside the normal domain of economics. Accordingly, a large gap separated an identification of transaction costs, as the main factor to which the study of the organization of economic activity must repair, and efforts to give operational content to that insight.

¹ This argument is elaborated in Williamson (1981). It is briefly discussed below in conjunction with what is referred to as the "inhospitality tradition" within antitrust. See Section 1.

This paper is organized in two parts. Sections 1 and 2 sketch the background and set forth the arguments that are subsequently employed to interpret a series of organizational innovations that have successively yielded the modern corporation. Sections 3 and 4 deal with these changes. My discussion of organizational innovation begins with the latter half of the nineteenth century. In this regard, I follow Alfred Chandler who traces the origins of complex hierarchical forms of business organization to this period (1977). To be sure, others have identified interesting organizational developments in both Japanese² and English³ business history that predate, if not prefigure, those in the U.S. But these earlier developments were not widely adopted by other firms—and in any event represent very primitive forms of divisionalization.⁴ As a consequence, these earlier developments were of isolated economic importance and are properly distinguished from the general transformation of American industry that began in the nineteenth century and has continued since.

Key legal features of the corporation—limited liability and the transferability of ownership—are taken as given. Failure to discuss these does not reflect a judgment that these are either irrelevant or uninteresting. The main focus of this essay, however, is on the internal organization of the corporation. Since any of a number of internal structures is consistent with these legal features, an explanation for the spe-

² Sadao Takatera and Nobaru Nishikawa, in an unpublished manuscript (undated), discuss the "Genesis of Divisional Management and Accounting Systems in the House of Mitsui, 1710–1730."

³ Gary Anderson, Robert E. McCormick, and Robert D. Tollison, in an unpublished manuscript (May 1981), describe the "East India Company as a Multidivisional Enterprise" early in the 18th century.

⁴ Primitive divisionalization is often confused with but needs to be distinguished from multidivisionalization. See Alfred Sloan (1965) and Chandler (1962) for a discussion of the origins of the M-form structure in the twentieth century.

cific organizational innovations that were actually adopted evidently resides elsewhere. Among the more significant of these innovations, and the ones addressed here, are: the development of line-and-staff organization by the railroads; selective forward integration by manufacturers into distribution; the development of the multidivisional corporate form; the evolution of the conglomerate; and the appearance of the multinational enterprise. The first three of these changes have been studied by business historians, the contributions of Chandler (1962; 1977) being the most ambitious and notable.

1. *Some Background*

1.1 *General*

Assessing the organization of economic activity in an advanced society requires that a bewildering variety of market, hierarchical, and mixed modes be evaluated. Economists, organization theorists, public policy specialists, and historians all have an interest and each have offered interpretations of successive organizational innovations. A coherent view, however, has not emerged.

Partly this is because the principal hierarchical structure to be assessed—the modern corporation—is formidably complex in its great size, diversity, and internal organization. The natural difficulties which thereby resulted would have been overcome sooner, however, had it not been for a number of conceptual barriers to an understanding of this institution. Chief among these are the following: (1) the neoclassical theory of the firm, which is the main referent to which economists appeal, is devoid of interesting hierarchical features; (2) organization theorists, who are specialists in the study of internal organization and unencumbered by an intellectual commitment to neoclassical economic models, have been preoccupied with hierarchy to the neglect of market

modes of organization and the healthy tension that exists between markets and hierarchies; (3) public policy analysts have maintained a deeply suspicious attitude toward nonstandard or unfamiliar forms of economic organization; and (4) organizational innovation has been relatively neglected by business and economic historians.

To be sure, this indictment sweeps too broadly. As discussed in 1.2 below, there have been important exceptions. The main features, however, are as I have described. Thus neoclassical theory treats the firm as a production function to which a profit maximization objective has been ascribed. Albeit useful for many purposes, such a construction is unhelpful in attempting to assess the purposes served by hierarchical modes of organization. The firm as production function needs to make way for the view of the *firm as governance structure* if the ramifications of internal organization are to be accurately assessed. Only recently has this latter orientation begun to make headway—and is still in a primitive state of development.

The preoccupation of organization theory specialists with internal organization is a potentially useful corrective. An understanding of the purposes served by internal organization has remained elusive, however, for at least two reasons. First, efficiency analysis plays a relatively minor role in the studies of most organization theory specialists—many of whom are more inclined to emphasize power. The economizing factors that are crucial to an understanding of the modern corporation are thus effectively suppressed. Second, and related, firms and markets are treated separately rather than in active juxtaposition with one another. The propositions that (1) firms and markets are properly regarded as alternative governance structures to which (2) transactions are to be assigned in discriminating (mainly transaction cost economizing) ways are unfa-

miliar to most organization theory specialists and alien to some.

Public policy analysts with an interest in the modern corporation might also have been expected to entertain a broader view. In fact, however, many of these likewise adopted a production function orientation—whereby markets were regarded as the “natural, hence efficient” way by which to mediate transactions between technologically separable entities. This was joined by a pervasive sense that the purposes of competition are invariably served by maintaining many autonomous traders. Even sensitive observers were trapped by this combined technological/atomistic logic. Thus Donald Turner, at a time when he headed the Antitrust Division, expressed skepticism over non-standard business practices by observing that “I approach territorial and customer restrictions not hospitably in the common law tradition, but inhospitably in the tradition of antitrust law.”⁵ The possibility that efficiency might be served by imposing restraints on autonomous market trading was evidently thought to be slight. This inhospitality tradition also explains ingrained public policy animosity towards vertical integration and conglomerate organization; more generally, industrial organization specialists were encouraged to discover what were often fanciful “distortions” at the expense of a more basic understanding of the modern corporation in economizing terms.

The neglect of organizational innovations by business and economic historians has been general but by no means complete and shows recent signs of being corrected.⁶ Mainly, however, interpreta-

⁵ The quotation is attributed to Turner by Stanley Robinson, 1968, N.Y. State Bar Association, Antitrust Symposium, p. 29.

⁶ For an interesting commentary and contribution, see Douglass North (1978). The earlier Lance Davis and North book, however, gave relatively little attention to institutional changes that occurred within firms (1971, p. 143).

tion has played a secondary role to description in most historical studies of organizational change—which, while understandable, contributes to the continuing confusion over the purposes served by the changing organizational features of the corporation.

This essay attempts to provide a coherent view of the modern corporation by (1) augmenting the model of the firm as production function to include the concept of the firm as governance structure, (2) studying firms and markets as alternative governance structures in a comparative institutional way, (3) supplanting the presumption that organizational innovations have anticompetitive purposes by the rebuttable presumption that organizational innovations are designed to economize on transaction costs, and (4) interpreting business history from a transaction cost perspective. Such an approach to the study of the modern corporation (and, more generally, to the study of organizational innovation) owes its origins to antecedent contributions of four kinds.

1.2 *Antecedents*

(a) Theory of Firms and Markets

The unsatisfactory state of the theory of the firm was recognized by Ronald Coase in his classic 1937 article on “The Nature of the Firm.” As he observed:

Outside the firm, price movements direct production, which is co-ordinated through a series of exchange transactions on the market. Within a firm, these market transactions are eliminated and in place of the complicated market structure with exchange transactions is substituted the entrepreneur-co-ordinator, who directs production. It is clear that these are *alternative means of co-ordinating production* [1952, p. 333; emphasis added].

Coase went on to observe that firms arose because there were costs of using the price system (1952, pp. 336–38). But internal organization was no cost panacea, since it experienced distinctive costs of its own

(1952, pp. 340–42). A balance is struck when the firm has expanded to the point where “the costs of organizing an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange in the open market or the costs of organizing in another firm” (1952, p. 341).

Related insight on the study of firms and markets was offered by Friedrich A. Hayek, who dismissed equilibrium economics with the observation that “the economic problem of society is mainly one of adaptation to changes in particular circumstances of time and place” (1945, p. 524), and who held that the “marvel” of the price system was that it could accomplish this without “conscious direction” (1945, p. 527). Setting aside the possibility that Hayek did not make adequate allowance for the limitations of the price system, three things are notable about these observations. First is his emphasis on change and the need to devise adaptive institutional forms. Second, his reference to particular circumstances, as distinguished from statistical aggregates, reflects a sense that economic institutions must be sensitive to dispersed knowledge of a microanalytic kind. And third was his insistence that attention to the details of social processes and economic institutions was made necessary by the “unavoidable imperfection of man’s knowledge” (1945, p. 530).

The organization of firms and markets has been a subject to which Kenneth Arrow has made repeated contributions. He has addressed himself not only to the economics of the internal organization (Arrow, 1964) but also to an assessment of the powers and limits of markets (Arrow, 1969). Like Coase, he expressly recognizes that firms and markets are alternative modes of organizing economic activity (Arrow, 1974). Moreover, whereas the limits of markets were glossed over by Hayek, Arrow specifically traces these to transac-

tion cost origins: “market failure is not absolute; it is better to consider a broader category, that of transaction costs, which in general impede and in particular cases block the formation of markets” (1969, p. 48)—where by transaction costs Arrow has reference to the “costs of running the economic system” (1969, p. 48).

(b) Organization Theory

Although organization theorists have not in general regarded efficiency as their central concern, there have been notable exceptions. The early works of Chester Barnard (1938) and Herbert Simon (1947) both qualify.

Barnard was a businessman rather than a social scientist and he addressed internal organizational issues that many would regard as outside the scope of economics. Economizing was nevertheless strongly featured in his approach to the study of organizations. Understanding the employment relation was among the issues that intrigued him. Matters that concerned him in this connection included: the need to align incentives, including noneconomic inducements, to achieve enterprise viability; the importance of assent to authority; a description of the authority relation within which hierarchical organizations operate; and the role of “informal organization” in supporting the working rules upon which formal organization relies. The rationality of internal organization, making due allowance for the attributes of human actors, was a matter of continuous concern to Barnard.

Simon expressly relies on Barnard and carries rationality analysis forward. A more precise vocabulary than Barnard’s is developed in the process. Simon traces the problem of organization to the joining of rational purposes with the cognitive limits of human actors: “it is precisely in the realm where human behavior is *intendedly* rational, but only *limitedly* so, that there is room for a genuine theory

of organization and administration" (1957, p. xxiv). Intended rationality supplies purpose, but meaningful economic and organizational choices arise only in a limited (or bounded) rationality context.

Simon makes repeated reference to the criterion of efficiency (1957, pp. 14, 39–41, 172–97), but he also cautions that organizational design should be informed by "a knowledge of those aspects of the social sciences which are relevant to the broader purposes of the organization" (1957, p. 246). A sensitivity to subgoal pursuit, wherein individuals identify with and pursue local goals at the possible expense of global goals⁷ (Simon, 1957, p. 13), and the "outguessing" or gaming aspects of human behavior (Simon, 1957, p. 252) are among these.

Although Simon examines the merits of centralized versus decentralized modes of organization (1947, pp. 234–40), it is not until his later writing that he expressly addresses the matter of factoring problems according to rational hierarchical principles (Simon, 1962). The issues here are developed more fully in Section 2.

(c) Nonstrategic Purposes

The "inhospitality tradition" referred to above maintained a presumption of illegality when nonstandard or unfamiliar business practices were brought under review. These same practices, when viewed "through the lens of price theory"⁸ by

⁷ The term "local goals" subsumes both the functional goals of a subunit of the enterprise and the individual goals of the functional managers. In a perfectly harmonized system, private goals are consonant with functional goals, the realization of which in turn promotes global goals. Frequently, however, managers become advocates for parochial interests that conflict with global goal attainment. If, for example, R&D claims a disproportionate share of resources—because of effective but distorted partisan representations from the management and staff of this group—profits (global goals) will suffer. Aggressive subgoal (or local goal) pursuit of this kind is a manifestation of opportunism (see 2.2, below).

⁸ The phrase is Richard Posner's (1979, p. 928).

Aaron Director (and his students and colleagues at Chicago), were regarded rather differently. Whereas Turner and others held that anticompetitive purposes were being served, Director and his associates reported instead that tie-ins, resale price maintenance, and the like were promoting more efficient resource allocation.

In fact, nonstandard business practices (such as tie-ins) are anomalies when regarded in the full information terms associated with static price theory. Implicitly, however, Chicago was also relying on the existence of transaction costs—which, after all, were the reason why comprehensive price discrimination could not be effected through simple contracts unsupported by restrictive practices from the outset.⁹ Be that as it may, Chicago's insistence that economic behavior be assessed with respect to its economizing properties was a healthy antidote and encouraged further scrutiny of these same matters—with the eventual result that an economizing orientation is now much more widely held. Indirectly, these views have spilled over and influenced thinking about the modern corporation as an economizing, rather than mainly a monopolizing, entity.¹⁰

(d) Business History

The study of organizational innovation has been relatively neglected by business and economic historians. Aside from the Research Center in Entrepreneurial History at Harvard, which was established in 1948 and closed its doors a decade later, there has not been a concerted effort to work through and establish the impor-

⁹ For a discussion of this point, see Williamson (1975, pp. 11–13, 109–10).

¹⁰ Although the nonstrategic tradition inspired by Aaron Director makes insufficient allowance for anticompetitive behavior, it was a useful counterweight to the inhospitality tradition to which it was paired. For a critique of the more extreme versions of this nonstrategic—or, as Posner (1979, p. 932) puts it, the "diehard Chicago"—tradition, see Williamson (1981).

tance of organizational innovation. Probably the most important reason for this neglect is that business history has not had “the support of an established system of theory” (Henrietta Larson, 1948, p. 135).

Despite this general neglect, notable contributions have nevertheless been made. The works of Lance Davis and Douglass North (1971) and of Alfred Chandler (1962; 1977) have been especially important. The first of these takes a sweeping view of institutional change and employs a market failure theory for assessing successive changes. It pays only limited attention, however, to the corporation as a unit whose attributes need to be assessed.¹¹

Chandler, by contrast, is expressly and deeply concerned with the organization form changes which, over the past 150 years, have brought us the modern corporation as we know it. The story is told in two parts, the first being the evolution of the large, multifunctional enterprise (Chandler, 1977), the second being the subsequent divisionalization of these firms (Chandler, 1962). Both of these transformations are described and interpreted in Sections 3 and 4 below. Suffice it to observe here that (1) Chandler’s is the first treatment of business history that de-

scribes organizational changes in sufficient detail to permit a transaction cost interpretation to be applied, (2) Chandler’s 1962 book was significant not only for its business history contributions but because it clearly established that organization form had an important impact on business performance—which neither economics nor organization theory had done (nor, for the most part, even attempted) previously, and (3) although Chandler is more concerned with the description than with the interpretation of organizational change, his careful descriptions are nevertheless suggestive of the economic factors that are responsible for the changes observed.

2. *Transaction Cost Economics*

Each of the antecedent literatures just described has a bearing on the transaction-cost approach to the study of economic institutions in general and the modern corporation in particular. Following Commons (1934), the transaction is made the basic unit of analysis. Specifically, attention is focused on the transaction costs of running the economic system (Coase, 1937; Arrow, 1969), with emphasis on adaptation to unforeseen, and often unforeseeable, circumstances (Hayek, 1945). The issues of special interest are connected with the changing structure of the corporation over the past 150 years (Chandler, 1962; 1977). Rather than regard these inhospitably, the new approach maintains the rebuttable presumption that the evolving corporate structure has the purpose and effect of economizing on transaction costs. These transaction-cost and business history literatures are linked by appeal to selective parts of the (mainly older) organization theory literature.

As Barnard (1938) emphasized, differences in internal organization often had significant performance consequences and could and should be assessed from a rationality viewpoint. Simon (1947) ex-

¹¹ Davis and North make repeated reference to the limited liability and unlimited life features of the corporate form and explicitly discuss the importance of organizational changes made by the railroads (1971, pp. 143–44). Their treatment of organization form changes in manufacturing, however, emphasizes economies of scale, monopolization (cartelization), protection against foreign competition, and resistance to regulation (1971, pp. 167–90). A sense that the corporation is progressively refining structures that economize on transaction costs—in labor, capital, and intermediate product markets—is nowhere suggested.

Although this is partly rectified in North’s recent survey paper, where he observes that recent organizational changes have had transaction cost origins, he defines transaction costs narrowly in terms of the “measurement of the separable dimensions of a good or services” (1978, p. 971). As developed below, measurement is only one aspect—and not, in my judgment, the most important one—for understanding the modern corporation.

tended and refined the argument that internal organization mattered and that the study of internal organization needed to make appropriate allowance for the attributes of human actors—for what Frank Knight has felicitously referred to as “human nature as we know it” (1965, p. 270). Then, and only then, does the comparative institutional assessment of alternative organizational forms take on its full economic significance.

2.1 *Comparative Institutional Analysis*

The costs of running the economic system to which Arrow refers can be usefully thought of in contractual terms. Each feasible mode of conducting relations between technologically separable entities can be examined with respect to the ex ante costs of negotiating and writing, as well as the ex post costs of executing, policing, and, when disputes arise, remedying the (explicit or implicit) contract that joins them.

A transaction may thus be said to occur when a good or service is transferred across a technologically separable interface. One stage of processing or assembly activity terminates and another begins. A mechanical analogy, while imperfect, may nevertheless be useful. A well-working interface, like a well-working machine, can be thought of as one where these transfers occur smoothly.

In neither case, however, is smoothness desired for its own sake: the benefits must be judged in relation to the cost. Both investment and operating features require attention. Thus extensive prior investment in finely tuned equipment and repeated lubrication and adjustment during operation are both ways of attenuating friction, slippage, or other loss of mechanical energy. Similarly, extensive pre-contract negotiation that covers all relevant contingencies may avoid the need for periodic intervention to realign the interface during execution so that a contract may

be brought successfully to completion. Simultaneous attention to both investment (pre-contract costs) and operating expenses (harmonizing costs) is needed if mechanical (contractual) systems are to be designed effectively. The usual study of economizing in a production function framework is thus extended to include an examination of the *comparative costs of planning, adapting, and monitoring task completion under alternative governance structures*—where by governance structure I have reference to the explicit or implicit contractual framework within which a transaction is located (markets, firms, and mixed modes—e.g., franchising—included).

The study of transaction-cost economizing is thus a comparative institutional undertaking which recognizes that there are a variety of distinguishably different transactions on the one hand and a variety of alternative governance structures on the other. The object is to match governance structures to the attributes of transactions in a discriminating way. Microanalytic attention to differences among governance structures and microanalytic definition of transactions are both needed in order for this to be accomplished.

Although more descriptive detail than is associated with neoclassical analysis is needed for this purpose, a relatively crude assessment will often suffice. As Simon has observed, comparative institutional analysis commonly involves an examination of discrete structural alternatives for which marginal analysis is not required: “In general, much cruder and simpler arguments will suffice to demonstrate an inequality between two quantities than are required to show the conditions under which these quantities are equated at the margin” (1978, p. 6).

2.2 *Behavioral Assumptions*

Human nature as we know it is marvelously rich and needs to be reduced to

manageable proportions. The two behavioral assumptions on which transaction-cost analysis relies—and without which the study of economic organization is pointless—are bounded rationality and opportunism. As a consequence of these two assumptions, the human agents that populate the firms and markets with which I am concerned differ from economic man (or at least the common caricature thereof) in that they are less competent in calculation and less trustworthy and reliable in action. A condition of bounded rationality is responsible for the computational limits of organization man. A proclivity for (at least some) economic agents to behave opportunistically is responsible for their unreliability.

The term bounded rationality was coined by Simon to reflect the fact that economic actors, who may be presumed to be “intendedly rational,” are not hyper-rational. Rather, they experience limits in formulating and solving complex problems and in processing (receiving, storing, retrieving, transmitting) information (Simon, 1957, p. 198). Opportunism is related to but is a somewhat more general term than the condition of “moral hazard” to which Knight referred in his classic statement of economic organization (1965, pp. 251–56).¹² Opportunism effectively extends the usual assumption of self-interest seeking to make allowance for self-interest seeking with guile.

But for the *simultaneous* existence of both bounded rationality and opportu-

ism,¹³ all economic contracting problems are trivial and the study of economic institutions is unimportant. Thus, but for bounded rationality, all economic exchange could be effectively organized by contract. Indeed, the economic theory of comprehensive contracting has been fully worked out.¹⁴ Given bounded rationality, however, it is impossible to deal with complexity in all contractually relevant respects (Radner, 1968). As a consequence, incomplete contracting is the best that can be achieved.

Ubiquitous, albeit incomplete, contracting would nevertheless be feasible if economic agents were completely trustworthy. Principals would simply extract promises from agents that they will behave in a stewardship fashion, while agents would reciprocally ask principals to behave in good faith. Such devices will not work, however, if some economic actors (either principals or agents) are dishonest (or, more generally, disguise attributes or preferences, distort data, obfuscate issues, and otherwise confuse transactions) and it is very costly to distinguish opportunistic from nonopportunistic types *ex ante*.

Although the dual assumptions of bounded rationality and opportunism complicate the study of economic behavior and may be inessential for some purposes, the study of alternative modes of organization does not qualify as an excep-

¹² Moral hazard is a technical term with a well defined meaning in the insurance literature. It refers to an *ex post* insurance condition and is clearly distinguished from adverse selection, which is responsible for a troublesome *ex ante* insurance screening problem. Opportunism is a less technical but more general term that applies to a wide set of economic behavior—of which adverse selection and moral hazard are specific kinds. Unless, therefore, moral hazard is given a broader meaning, the substitution of moral hazard for opportunism focuses attention on a subset of the full range of human and economic conditions of concern.

¹³ The co-existence of cunning and bounded rationality is troublesome to some. How can economic agents simultaneously be more clever and less competent than the hyper-rational man that populates neoclassical models? Is he a maximizer or is he not? This is not a useful dichotomy. Maximizing is an analytical convenience the use of which is often justified by the fact that human agents are “*intendedly rational*” (Simon, 1957b, p. xxiv). As discussed in the text, however, comprehensive contracting, which is an ambitious form of maximizing, is infeasible. Opportunism has important economic ramifications for this reason.

¹⁴ I have reference, of course, to the Arrow–Debreu contracting model.

tion. To the contrary, failure to recognize and make allowance for both is virtually to invite mistaken assessments of alternative modes.¹⁵ Taking these two behavioral assumptions into account, the following compact statement of the problem of economic organization is suggested: assess alternative governance structures in terms of their capacities to economize on bounded rationality while simultaneously safeguarding transactions against opportunism. This is not inconsistent with the imperative "maximize profits!," but it focuses attention somewhat differently.

2.3 *Dimensionalizing*

As Coase observed in 1972, his 1937 paper was "much cited but little used" (1972, p. 63). The reasons for this are many, including a preoccupation by economists with other matters during the intervening 35 years. The main reason, however, is that transaction costs had not been operationalized and it was not obvious how this could be accomplished.

The postwar market failure literature, especially Arrow's insight (1969) that market failures had transaction costs origins, served to focus attention on the troublesome issues. A recognition that market (and internal) failures of all kinds could be ultimately traced to the human factors described above was a second step. The remaining step was to identify the critical dimensions with respect to which transactions differ.

The attributes of transactions that are of special interest to the economics of organization are: (1) the frequency with

which transactions recur, (2) the uncertainty to which transactions are subject,¹⁶ and (3) the degree to which transactions are supported by durable, transaction-specific investments (Williamson, 1979). A considerable amount of explanatory power turns on the last.¹⁷

Asset specificity can arise in any of three ways: site specificity, as when successive stations are located in cheek-by-jowl relation to each other so as to economize on inventory and transportation expenses; physical asset specificity, as where specialized dies are required to produce a component; and human asset specificity that arises in a learning-by-doing fashion. The reason why asset specificity is critical is that, once the investment has been made, buyer and seller are effectively operating in a bilateral (or at least quasi-bilateral) exchange relation for a considerable period thereafter. Inasmuch as the value of highly specific capital in other uses is, by definition, much smaller than the specialized use for which it has been intended, the supplier is effectively "locked into" the transaction to a significant degree. This is symmetrical, moreover, in that the buyer cannot turn to alternative sources of supply and obtain the item on favorable terms, since the cost of supply from unspecialized capital is presumably great. The buyer is thus committed to the transaction as well. Accordingly, where asset specificity is great, buyer and seller will make special efforts to design an exchange relation that has good continuity properties. Autonomous contracting gives way to

¹⁵ The argument that effective ex ante competition for the right to supply service (franchise bidding) vitiates the need to regulate decreasing cost industries sometimes goes through but not always. Incomplete contracting (bounded rationality) coupled with the hazards of ex post opportunism place great strain on the franchise bidding mode if assets are durable and specific. For a critique of what I believe was a mistaken assessment of the feasibility of using franchise bidding for CATV, see Williamson (1976).

¹⁶ As Knight observes: "With uncertainty entirely absent, every individual being in possession of perfect knowledge of the situation, there would be no occasion for anything of the nature of responsible management or control of productive activity" (1965, p. 267).

¹⁷ Williamson (1979). Also see Benjamin Klein, Robert Crawford, and Armen Alchian (1978) for an illuminating discussion of transaction specific investments in the context of what they refer to as "appropriate quasi-rents."

more complex forms of market contracting and sometimes to internal organization for this reason.

2.4 *Three Principles of Organizational Design*

The criterion for organizing commercial transactions is assumed to be the strictly instrumental one of cost economizing. Essentially this takes two parts: economizing on production expense and economizing of transaction costs. In fact, these are not independent and need to be addressed simultaneously. The study of the latter, however, is much less well developed and is emphasized here.

The three principles of organizational design employed here are neither exhaustive nor refined. They nevertheless offer considerable explanatory power in dealing with the main changes in corporate organization reported by Chandler and addressed here. Transaction cost reasoning supports all three, although only the first, the asset-specificity principle, is tightly linked to dimensionalizing. Bounded rationality and opportunism, however, operate with respect to all three.

The asset-specificity principle turns on the above described transformation of an exchange relation from a large-numbers to a small-numbers condition during the course of contract execution. The second, the externality principle, is often discussed under the heading of "free rider" effects. The more general phenomenon, however, is that of subgoal pursuit, that is, in the course of executing contracts, agents also pursue private goals which may be in some degree inconsistent with the contract's intended purpose. These two principles influence the choice of contracting form (mainly firm or market). In fact, however, the efficacy of internal organization depends on whether sound principles of internal organizational design are respected, which is to say that the details of internal organization matter.

The hierarchical decomposition principle deals with this last.

It will be convenient to assume that transactions will be organized by markets unless market exchange gives rise to serious transaction costs. In the beginning, so to speak, there were markets. Both bureaucratic and production cost considerations favor this presumption. The bureaucratic argument is simply this: market exchange serves to attenuate the bureaucratic distortions to which internal exchange is subject. (Although the reasons for this have been set out elsewhere—James Thompson, 1967, pp. 152–54; Williamson, 1975, Chapter 7—the study of firm and market organization is greatly in need of a more adequate theory of bureaucracy.) The production cost advantages of market procurement are three: static scale economies can be more fully exhausted by buying rather than making if the firm's needs are small in relation to the market; markets can aggregate uncorrelated demands, to thereby realize risk pooling benefits; and markets may enjoy economies of scope¹⁸ in supplying a related set of activities of which the firm's requirements are only one. Accordingly, transactions will be organized in markets *unless* transaction cost disabilities appear.¹⁹

¹⁸ Whereas scale economies refer to declining average costs associated with increasing output of a single line of commerce, scope economies are realized "where it is less costly to combine two or more product lines in one firm rather than to produce them separately" (John Panzar and Robert Willig, 1981, p. 268). Retail outlets that carry many products and brands (drug stores, department stores) presumably enjoy significant economies of scope in the retailing function.

¹⁹ Bureaucratic disabilities aside, any given firm could realize all of these production benefits for itself by an appropriate increase in the scale and scope of its activities. Pursuit of this logic, however, leads to the following anomaly: all firms, of which there will be few, will be comprehensively integrated and diversified in sufficient degree to obviate the need for market exchange. The fact that we do not observe comprehensive integration—as Coase puts it, "Why is not all production carried on by one big firm?"

(a) Asset Specificity Principle (All Transactions)

Recall that transactions are described in terms of three attributes: frequency, uncertainty, and asset specificity. Although interesting organizational issues are posed when transactions are of only an occasional kind (Williamson, 1979, pp. 245–54), this paper deals entirely with the governance of recurring transactions. Also, it will facilitate the analysis to hold uncertainty constant in intermediate degree—which is to say that we are dealing neither with steady state nor highly uncertain events. Accordingly, asset specificity is the transactional dimension of special interest. The first principle of efficient organizational design is this: *the normal presumption that recurring transactions for technologically separable goods and services will be efficiently mediated by autonomous market contracting is progressively weakened as asset specificity increases.*

The production cost advantages of markets decrease and the (comparative) governance costs of markets increase as assets become progressively more specific. Thus as assets become more fully specialized to a single use or user, hence are less transferable to other uses and users, economies of scale can be as fully realized when a firm operates the asset under its own internal direction as when its services are obtained externally by contract. And the market's advantage in pooling risks likewise shrinks. Simultaneously, the transactions in question take on a stronger bilateral character, and the governance costs of markets increase relatively.

(1952, p. 340)—suggests that the bureaucratic disabilities of internal organization are very serious. But since we do observe that some transactions are organized within firms, this poses the question of which and why. The answer resides in the transaction cost disabilities of markets that arise when asset specificity and demand externalities appear.

The distinction between ex ante and ex post competition is essential to an understanding of this condition. What may have been (and commonly is) an effective large-numbers-bidding situation at the outset is sometimes *transformed* into a bilateral trading relation thereafter. This obtains if, despite the fact that large numbers of qualified bidders were prepared to enter competitive bids for the initial contract, the winning bidder realizes advantages over nonwinners at contract renewal intervals because nontrivial investments in durable specific assets are put in place (or otherwise accrue, say in a learning-by-doing fashion) during contract execution. As set out elsewhere (Williamson, 1979), the efficient governance of recurring transactions will vary as follows: classical market contracting will be efficacious wherever assets are nonspecific to the trading parties; bilateral or obligational market contracting will appear as assets become semi-specific; and internal organization will displace markets as assets take on a highly specific character.²⁰

Internal organization enjoys advantages over market contracting for transactions that are supported by highly specific assets that are supported by highly specific assets at both contract-writing and contract-execution stages. Since highly specific assets cannot be redeployed without sacrificing productivity, both suppliers and purchasers will insist upon contractual safeguards before undertaking such projects. Writing and negotiating such contracts is costly. Additionally, implementation problems

²⁰ Note that the nature of the asset specificity matters. If the assets in question are mobile and the specificity is due to physical but not human asset features, market procurement may still be feasible. This can be accomplished by having the buyer own the specific assets (e.g., dies). He puts the business up for bid and awards it to the low bidder, to whom he ships the dies. Should contractual difficulties arise, however, he is not locked into a bilateral exchange. He reclaims the dies and reopens the bidding. This option is not available if the specific assets are of a human asset kind or if they are nonmobile. See David Teece (1980) for a related discussion.

need to be faced. The internal direction of firms confers execution advantages over bilateral trading in three respects. First, common ownership reduces the incentives of the trading units to pursue local goals. Second, and related, internal organization is able to invoke fiat to resolve differences whereas costly adjudication is needed when an impasse develops between autonomous traders. Third, internal organization has easier and more complete access to the relevant information when disputes must be settled. The incentive to shift bilateral transactions from markets to firms also increases as uncertainty increases, since the costs of harmonizing a relation among parties vary directly with the need to adjust to changing circumstances.

(b) Externality Principle (Forward Integration)

Whereas the asset specificity principle refers to transactions that are transformed from large- to small-numbers bidding situations—as buyers, who initially obtained assets or their services in a competitive market, subsequently face suppliers with some degree of monopoly power—the externality principle involves no such market transformation. Also, the asset-specificity principle applies to backward, forward, and lateral integration; by contrast, the externality principle mainly applies to distribution stages.

The externalities of concern are those that arise in conjunction with the unintended debasement of quality for a branded good or service. As discussed below, such debasement is explained by costly metering. The externality is thus a manifestation of the measurement problems to which North refers in his discussion of transaction costs (1978, p. 972). It appears mainly at the interface between production and distribution. The differential ease of inspecting, and thereby controlling, the quality of components and

materials that are purchased from earlier-stage and lateral suppliers as compared with the cost of exercising quality controls over distributors is responsible for this condition.²¹

End-games and fly-by-night distributors aside, the unintended debasement of quality by distributors poses a problem only where the activities of individual distributors affect one another, as when one retailer's poor service in installation or repair injures a product's reputation for performance and limits the sales of other retailers. More generally, if the quality enhancement (debasement) efforts of distributors give rise to positive (negative) externalities, the benefits (costs) of which can be incompletely appropriated by (assigned to) the originators, failure to extend quality controls over distribution will result in suboptimization. Autonomous contracting thus gives way to obligational market contracting (e.g., franchising) if not forward integration into distribution²² as demand interaction effects become more important. More generally, the second principle of efficient organizational design is this: *the normal presumption that exchange between producers of differentiated goods and distribution stages will be efficiently mediated by autonomous contracting is progressively weakened as demand externalities increase.*

Product differentiation is a necessary but not a sufficient condition for troublesome demand externalities to appear. Manufacturers can sometimes insulate a product against deterioration by special packaging (say by selling the item in hermetic containers with an inert atmosphere and providing replacement guar-

²¹ Manufacturers may, of course, decide to integrate into components if work-in-process inspections are much cheaper than final inspections.

²² Franchising will be more prevalent if aggregation economies are present at the distribution stage. It will be inefficient in these circumstances for a single product firm to integrate forward into distribution.

antees). If, however, such safeguards are very costly, and if follow-on checks and penalties to discourage distributors from debasing the quality image of a product are likewise expensive, autonomous trading will give way to forms of distribution that have superior quality control properties.

(c) Hierarchical Decomposition Principle (Internal Organization)²³

Merely to transfer a transaction out of the market into the firm does not, by itself, assure that the activity will be effectively organized thereafter. Not only are bounded rationality and opportunism ubiquitous, but the problems presented by both vary with changes in internal organization. Accordingly, a complete theory of value will recognize that firm structure as well as market structure matters.

Simon makes provision for bounded rationality effects in arguing that the organizational division of decision making labor is quite as important as the neoclassical division of production labor, where, from "the information processing point of view, division of labor means factoring the total system of decisions that need to be made into relatively independent subsystems, each one of which can be designed with only minimal concern for its interactions with the others" (Simon, 1973, p. 270). This applies to both vertical and horizontal aspects of the organization. In both respects the object is to recognize and give effect to conditions of near decomposability. The vertical slice entails grouping the operating parts into separable entities, the interactions within which are strong and between which are weak. The horizontal

slice has temporal ramifications of a strategic versus operating kind. Problems are thus factored in such a way that the higher frequency (or short run dynamics) are associated with the operating parts while the lower frequency (or long run dynamics) are associated with the strategic system (Simon, 1962, p. 477). These operating and strategic distinctions correspond with the lower and higher levels in the organizational hierarchy, respectively. Internal incentives and information flows need, of course, to be aligned, lest distortions be deliberately or inadvertently introduced into the internal information summary and transmittal processes.

The hierarchical decomposition principle can thus be stated as follows: *internal organization should be designed in such a way as to effect quasi-independence between the parts, the high frequency dynamics (operating activities) and low frequency dynamics (strategic planning) should be clearly distinguished, and incentives should be aligned within and between components* so as to promote both local and global effectiveness.

Each of these three principles of organizational design is responsive to considerations of both bounded rationality and opportunism. Thus asset specificity would pose no problems if comprehensive contracting were feasible (which is tantamount to unbounded rationality) or if winning bidders could be relied upon to behave in an utterly reliable and trustworthy fashion (absence of opportunism). The externality principle is mainly a reflection of opportunism (autonomous distributors permit their suppliers' reputations to be degraded because they bear only part of the costs), but, of course, quality control checks would be unneeded if all relevant information could be costlessly displayed and assessed. The hierarchical decomposition principle recognizes the need to divide problems into manageable units and at the same time prevent agents from en-

²³ The hierarchical decomposition principle is due to Simon (1962; 1973). As he observes, the anatomy of an organization can be viewed either in terms of the groupings of human beings or the flows and transformations of symbols (1973, p. 270). He emphasizes the latter, which is in the spirit of transaction cost analysis.

gaging in dysfunctional pursuit of local goals, which reflect bounded rationality and opportunism concerns, respectively.

A more comprehensive analysis would embed these principles of organization within a larger optimizing framework where demand as well as cost consequences are recognized and where production versus transaction costs tradeoffs are made explicit.²⁴ For the purposes at hand, however, which take product design as given and focus on distinguishably different rather than close cases, such refinements do not appear to be necessary.

3. *The Nineteenth Century Corporation*

The 1840s mark the beginning of a great wave of organizational change that has evolved into the modern corporation (Chandler, 1977). According to Stuart Bruchey, the fifteenth century merchant of Venice would have understood the form of organization and methods of managing men, records, and investment used by Baltimore merchants in 1790 (1956, pp.

²⁴ Thus, whereas I argue that the object is to minimize the sum of production and transaction costs, taking output and design as given, the more general problem is to maximize profits, treating output and design as decision variables. A rudimentary statement of the optimizing problem, for a given organization form (f), is to choose output (Q) and design (D) so as to maximize:

$$\pi(Q,D;f) = P(Q,D) \cdot Q - C_r(Q,D;S) - G_r(Q,D),$$

where π denotes profit, $P(Q,D)$ is the demand curve, S denotes combinatorial economies of scope, and C_r and G_r are the production costs and governance (transaction) costs of mode f . Transaction costs become relatively more important to this calculus as the assets needed to support specialized designs become progressively more specific—which they normally will as designs become more idiosyncratic.

Plainly the tradeoffs that run through this optimizing relation are more extensive than my earlier discussion discloses, but a detailed assessment of these is not needed for the types of purposes to which the asset specificity principle is herein applied. Both the externality and hierarchical decomposition principles should likewise be qualified to recognize tradeoffs. Again, however, second order refinements are not needed for the comparative institutional purposes to which these are applied below.

370–71). These practices evidently remained quite serviceable until after the 1840s. The two most significant developments were the appearance of the railroads and, in response to this, forward integration by manufacturers into distribution.

3.1 *The Railroads*

Although a number of technological developments—including the telegraph (Chandler, 1977, p. 189), the development of continuous process machinery (Chandler, pp. 252–53), the refinement of interchangeable parts manufacture (Chandler, 1977, pp. 75–77), and related mass manufacturing techniques (Chandler, Chap. 8)—contributed to organizational changes in the second half of the nineteenth century, none was more important than the railroads (Glenn Porter and Harold Livesay, 1971, p. 55). Not only did the railroads pose distinctive organizational problems of their own, but the incentive to integrate forward from manufacturing into distribution would have been much less without the low cost, reliable, all-weather transportation afforded by the railroads. Forward integration is discussed in 3.2 below; the railroads are treated here.

The appearance and purported importance of the railroads have been matters of great interest to economic historians. But with very few exceptions, the organizational—as opposed to the technological—significance of the railroads has been neglected. Thus Robert Fogel (1964) and Albert Fishlow (1965) “investigated the railroad as a construction activity and as a means of transport, but not as an organizational form. As with most economists, the internal workings of the railroad organizations were ignored. This appears to be the result of an implicit assumption that the organization form used to accomplish an objective does not matter” (Peter Temin, 1980, p. 3).

The economic success of the railroads

entailed more, however, than the substitution of one technology (rails) for another (canals). Rather, organizational aspects also required attention. As Chandler puts it:

[the] safe, regular, reliable movement of goods and passengers, as well as the continuing maintenance and repair of locomotives, rolling stock, and track, roadbed, stations, round-houses, and other equipment, required the creation of a sizeable administrative organization. It meant the employment of a set of managers to supervise these functional activities over an extensive geographical area; and the appointment of an administrative command of middle and top executives to monitor, evaluate, and coordinate the work of managers responsible for the day-to-day operations. It meant, too, the formulation of brand new types of internal administrative procedures and accounting and statistical controls. Hence, the operational requirements of the railroads demanded the creation of the first administrative hierarchies in American business [1977, p. 87].

The “natural” railroad units, as these first evolved, were lines of about fifty miles in length. These roads employed about fifty workers and were administered by a superintendent and several managers of functional activities (Chandler, 1977, p. 96). This was adequate as long as traffic flows were uncomplicated and short hauls prevailed. The full promise of the railroads could be realized, however, only if traffic densities were increased and longer hauls introduced. How was this to be effected?

In principle, successive end-to-end systems could be joined by contract. The resulting contracts would be tightly bilateral in negotiation, interpretation and execution, however, since investments in site-specific assets by each party were considerable. Severe contractual difficulties would, therefore, predictably arise.²⁵ Un-

²⁵ Problems of two kinds would need to be faced. Not only would the railroads need to reach agreement on how to deal with a series of complex operating matters—equipment utilization, costing, and maintenance; adapting cooperatively to unantic-

less supporting governance structure were simultaneously created,²⁶ the potential of the railroads for long-haul and high-density traffic would evidently go unrealized. One possibility was for heavily traveled end-to-end links to be joined under common ownership.

But while the consolidation of ownership reduced the restraints on long-haul operations, it did not guarantee that the end-to-end systems would work smoothly. Indeed, early operation of the Western and Albany road, which was just over 150 miles in length and was built in three sections each operated as a separate division with its own set of functional managers, quickly proved otherwise (Chandler, 1977, pp. 96–97). As a consequence, a new organizational structure was fashioned whereby the first “formal administrative structure manned by full-time salaried managers” in the U.S. appeared (Chandler, 1977, pp. 97–98).

This structure was progressively perfected, and the organizational innovation that the railroads eventually evolved is characterized by Chandler as the “decentralized line-and-staff concept of organization.” This provided that “the managers on the line of authority were responsible for ordering men involved with the basic function of the enterprise, and other functional managers (the staff executives) were responsible for setting standards” (Chandler, 1977, p. 106). Geographic divisions were defined and the superintendents in charge were held responsible for the “day-to-day movement of trains and traffic by

pated disturbances; assigning responsibility for customer complaints, breakdown, etc.—but problems of customers contracting with a set of autonomous end-to-end suppliers would need to be worked out. Plainly, complex contracting issues proliferate.

²⁶ Railroad regulation can be interpreted, in part, as an effort to deal with these contractual difficulties by inventing specialized governance structures. Pursuit of these matters is beyond the scope of this paper. Aspects of the general problem are dealt with in Williamson (1976) and Victor Goldberg (1976).

an express delegation of authority" (Chandler, 1977, p. 102). These division superintendents were on the "direct line of authority from the president through the general superintendent" (Chandler, 1977, p. 106), and the functional managers within the geographic divisions—who dealt with transportation, motive power, maintenance of way, passenger, freight, and accounting—reported to them rather than to their functional superiors at the central office (Chandler, 1977, pp. 106–07).

Confronted, as they were, by the contractual dilemmas that arise when highly specific assets are in place and by complexities that exceeded, perhaps by several orders of magnitude, those that had been faced by earlier business enterprise, the managements of the railroads supplanted markets by hierarchies of a carefully crafted kind. Although military organizations had earlier devised similar structures, the railroad innovators brought engineering rather than military backgrounds to the task (Chandler, 1977, Chapter 3). The hierarchical structure that they evolved was consistent, at least broadly, with the hierarchical principles stated by Simon. Thus support activities (lower frequency dynamics) were split off from operations (higher frequency dynamics), and the linkages within each of these classes of activity were stronger than the linkages between. This organizational innovation, in Chandler's judgment, paved the way for modern business enterprise. As with most significant organizational developments, it evolved in a piecemeal rather than a full-blown way (Richard Nelson and Sidney Winter, 1981). Failure to recognize the opportunities for decomposition of functions and to perfect the hierarchical governance structures by which these could be realized would have arrested the development of the modern corporation at a very primitive stage.

3.2 *Forward Integration*

Forward integration by manufacturers into distribution was one of the significant consequences of the appearance of the railroads. Low cost transportation combined with telegraph and telephone communication permitted manufacturers efficiently to service a larger market and, as a consequence, realize greater economies of scale in production. The points of connection between manufacturing, wholesaling, and retailing, however, also required attention. Forward integration was a common but by no means uniform response. To the contrary, it was highly selective rather than comprehensive, and it is this selectivity that is the matter of special interest to this paper.

At least four degrees of forward integration can be recognized. From least to most, these are:

- A: none—in which event traditional wholesale and retail distribution was continued (many grocery, drug, hardware, jewelry, liquor, and dry goods items were of this kind) [Porter and Livesay, 1971, p. 214].
- B: minor—efforts to presell product and to monitor wholesale inventories, but not to include the ownership and operation of wholesale plants, are examples. Certain branded nondurables (soups, soaps), especially those for which staling was a problem (cigarettes, cereals), are included.
- C: wholesale—this was undertaken for perishable, branded items that required special handling;²⁷ often specialized investments in refrigeration were involved (meat and beer are examples). [Chandler, 1977, p. 299].

²⁷ The Whitman candy case involved the use of two different merchandising methods. Wholesalers were bypassed in the sale of high-grade, packaged candies. Small, inexpensive, bar and packaged candies, by contrast, were sold through the usual jobber and wholesale grocer network. Control of the wholesaling function for the former was arguably more important for quality control purposes. These high-grade items were "sold directly to retailers so that the company could regulate the flow of the perishable items and avoid alienating customers," (Porter and Livesay, 1971, p. 220)—who were presumably prepared to pay a premium to avoid stale candy.

D: retail-integration into retail was rare and was reserved for "new, complex, high priced machines that required specialized marketing services—demonstration, installation, consumer credit, after-sales service and repair" (Chandler, 1977, p. 288). Certain consumer durables (sewing machines, automobiles) and producer durables (some electrical machinery and office machines) were of this kind.

Actually, there is a variant of this fourth category that I will designate "mistaken" retail integration. Such integration involved none of the transaction specific investments in sales and service referred to above but had the purpose of foreclosing rivals. The ill-fated efforts of American Tobacco to integrate forward into the wholesaling and retailing of cigars (Porter and Livesay, 1971, p. 210) and of American Sugar Refining to "drive its competitor John Arbuckle out of business by buying into wholesale and retail houses" (Porter and Livesay, 1971, p. 211, p. 52) are examples.²⁸

The question is how to interpret these developments. Although the data that would be needed for a quantitative analysis have yet to be worked up, a systematic qualitative interpretation along the lines of the discussion in Sections 2.2 and 2.3 above is nevertheless feasible. The attributes of the five integration classes are set out in Table 1, where ++ denotes considerable, + denotes some, ~ is uncertain, and 0 is negligible.

Markets remain the main mode for effecting distribution for classes A and B. Markets enjoy substantial economies of scope for these products while asset specificity is negligible and externalities are

²⁸ This is not to say that foreclosure will never be successful unless accompanied by transaction specific investments. But it should not entail sacrifice of scale economies. Forward integration by the motion picture producers into theatres may have been a viable means of foreclosing entry into the production stage because theatre ownership by major producers entailed little or no sacrifice of scale economies.

dealt with by monitoring inventory. Integration into wholesale occurs for products that involve some asset specificity (refrigeration) and the reputation of branded products needs protection. Integration into retail does not occur, however, until asset specificity at the retail level is great (and these are products for which separate sales and service entails negligible loss of scope economies).²⁹ Finally, mistaken retail integration involves the sacrifice of scope economies without offsetting governance cost benefits (externalities and asset specificity are negligible). This pattern of integration is broadly consistent with transaction cost reasoning and explains why forward integration occurred selectively rather than comprehensively in response to the transportation and communication infrastructure.³⁰

TABLE 1

Integration Class	Economies of Scope	Externalities	Asset Specificity
A: none	++	0	0
B: minor	+	+	0
C: wholesale	~	+	+
D ₁ : retail/viable	0	+	++
D ₂ : retail/mistaken	+	0	0

4. *The 20th Century Corporation*

Three developments are particularly noteworthy in the evolution of the modern corporation in the 20th century. The first of these was the appearance of the

²⁹ Concessions in department stores are devices for effecting retail sales for products that are efficiently marketed in conjunction with others but which nevertheless require transaction specific investments. Chandler does not discuss such products, but a more comprehensive microanalytic analysis would, I conjecture, disclose the existence of some where mixed modes arise because aggregation economies and asset specificity are simultaneously present.

³⁰ For a more complete assessment, on which the above is based, see Williamson (1980).

multidivisional (or M-form) organization. Later developments are the conglomerate and the multinational corporation.

4.1 *The Multidivisional Structure*

The most significant organizational innovation of the 20th century was the development in the 1920s of the multidivisional structure. Surprisingly, this development was little noted or widely appreciated as late as 1960. Leading management texts extolled the virtues of “basic departmentation” and “line and staff authority relationships,” but the special importance of multidivisionalization went unremarked.³¹

Chandler’s pathbreaking study of business history, *Strategy and Structure*, simply bypassed this management literature. He advanced the thesis that “changing developments in business organization presented a challenging area for comparative analysis” and observed that “the study of [organizational] innovation seemed to furnish the proper focus for such an investigation” (1966, p. 2). Having identified the multidivisional structure as one of the more important of such innovations, he proceeded to trace its origins, identify the factors that gave rise to its appearance, and describe the subsequent diffusion of this organizational form. It was uninformed and untenable to argue that organization form was of no account after the appearance of Chandler’s book.

The leading figures in the creation of the multidivisional (or M-form) structure were Pierre S. du Pont and Alfred P. Sloan; the period was the early 1920s; the firms were Du Pont and General Motors; and the organizational strain of trying to cope with economic adversity under the old structure was the occasion to innovate in both. The structures of the two companies, however, were different.

³¹ The treatment of these matters by Harold Koontz and Cyril O’Donnell (1955) is representative.

Du Pont was operating under the centralized, functionally departmentalized or unitary (U-form) structure. General Motors, by contrast, had been operated more like a holding company by William Durant—whose genius in perceiving market opportunities in the automobile industry (Livesay, 1979, pp. 232–34) evidently did not extend to organization. Chandler summarizes the defects of the large U-form enterprise in the following way:

The inherent weakness in the centralized, functionally departmentalized operating company . . . became critical only when the administrative load on the senior executives increased to such an extent that they were unable to handle their entrepreneurial responsibilities efficiently. This situation arose when the operations of the enterprise became too complex and the problems of coordination, appraisal, and policy formulation too intricate for a small number of top officers to handle both long-run, entrepreneurial, and short-run, operational administrative activities [1966, pp. 382–83].

The ability of the management to handle the volume and complexity of the demands placed upon it became strained and even collapsed. Unable meaningfully to identify with or contribute to the realization of global goals, managers in each of the functional parts attended to what they perceived to be operational subgoals instead (Chandler, 1966, p. 156). In the language of transaction cost economics, bounds on rationality were reached as the U-form structure labored under a communication overload while the pursuit of subgoals by the functional parts (sales, engineering, production) was partly a manifestation of opportunism.

The M-form structure fashioned by du Pont and Sloan involved the creation of semi-autonomous operating divisions (mainly profit centers) organized along product, brand, or geographic lines. The operating affairs of each were managed separately. More than a change in decomposition rules were needed, however, for the M-form to be fully effective. Du Pont

and Sloan also created a general office “consisting of a number of powerful general executives and large advisory and financial staffs” (Chandler, 1977, p. 460) to monitor divisional performance, allocate resources among divisions, and engage in strategic planning. The reasons for the success of the M-form innovation are summarized by Chandler as follows:

The basic reason for its success was simply that it clearly removed the executives responsible for the destiny of the entire enterprise from the more routine operational activities, and so gave them the time, information, and even psychological commitment for long-term planning and appraisal. . . .

[The] new structure left the broad strategic decisions as to the allocation of existing resources and the acquisition of new ones in the hands of a top team of generalists. Relieved of operating duties and tactical decisions, a general executive was less likely to reflect the position of just one part of the whole [1966, pp. 382–83].

In contrast with the holding company—which is also a divisionalized form but has little general office capability and hence is little more than a corporate shell—the M-form organization adds (1) a strategic planning and resource allocation capability and (2) monitoring and control apparatus. As a consequence, cash flows are reallocated among divisions to favor high yield uses, and internal incentive and control instruments are exercised in a discriminating way. In short, the M-form corporation takes on many of the properties of (and is usefully regarded as) a miniature capital market,³² which is a much more ambitious concept of the corporation than the term holding company contemplates.

Although the structure was imitated very slowly at first, adoption by U.S. firms proceeded rapidly during the period 1945 to 1960. Acceptance of this structure by European firms came later. Lawrence

³² Others who reported that the modern corporation was assuming capital market resource allocation and control functions include Richard Heflebower (1960) and Armen Alchian (1969).

Franko (1972) reports that most large European companies administered their domestic operations through U-form or holding company structures until late in the 1960s, but that rapid reorganization along M-form lines has occurred since. The advent of zero tariffs within the European Economic Community and the post-war penetration of European markets by American multinationals were, in his judgment, important contributing factors.

As W. Ross Ashby has observed, it is not sufficient to determine the behavior of a whole machine to know the behavior of its parts: “only when the details of coupling are added does the whole’s behavior become determinate” (1956, p. 53). The M-form structure represented a different solution to the coupling problem than the earlier unitary form structure. It effected decomposability along product or brand lines to which profit center standing could be assigned and it more clearly separated operating from strategic decision making. It carried Simon’s hierarchical decomposition principles to a higher degree of refinement.³³

As compared with the U-form organization of the same activities, the M-form organization of the large, complex corporation served both to economize on bounded rationality and attenuate opportunism. Specifically:

Operating decisions were no longer forced to the top but were resolved at the divisional level, which relieved the communication load. Strategic decisions were reserved for the general office, which reduced partisan political input into the resource allocation process. And the internal auditing and control techniques which the general office had access to served to overcome information impactedness conditions and permit fine timing controls to be exercised over the operating parts [Williamson, 1975, pp. 137–38].

³³ Moreover, whereas the line-and-staff structure that the railroads adopted in the 1850s could be said to have been prefigured by the military, there is no such military precedent for the M-form. Rather, the reorganization of the military after World War II has certain M-form attributes.

4.2 *The Conglomerate*

Chandler's studies of organizational innovation do not include the conglomerate and multinational form of corporate enterprise. These are more recent developments, the appearance of which would not have been feasible but for the prior development of the M-form structure. Inasmuch as transaction cost economizing is socially valued and has been relatively neglected by prior treatments, my discussion of both of these emphasizes affirmative aspects. But this is intended to redress an imbalance and should not be construed to suggest either that a transaction cost interpretation is fully adequate or that conglomerates and multinationals pose no troublesome public policy issues.³⁴ Unrelieved hostility to these two forms of organization, however, is clearly inappropriate. Specifically, conglomerates that have the capacity to allocate resources to high valued uses and multinationals that use the M-form to facilitate technology transfer warrant more sympathetic assessments.

Although diversification as a corporate strategy certainly predates the 1960s, when general awareness of the conglomerate began to appear, the conglomerate is essentially a post World War II phenomenon. To be sure, General Electric's profit centers number in the hundreds and GE has been referred to as the world's most diversified firm. Until recently, however, General Electric's emphasis has been the manufacture and distribution of electrical appliances and machinery. Similarly, although General Motors was more than an automobile company, it took care to limit its portfolio. Thus Sloan remarked that "tetraethyl lead was clearly a misfit for GM. It was a chemical product, rather

than a mechanical one. And it had to go to market as part of the gasoline and thus required a gasoline distribution system" (Burton and Kuhn, 1979, p. 6). Accordingly, although GM retained an investment position, the Ethyl Corporation became a free-standing entity rather than an operating division (Sloan, 1965, p. 224). Similarly, although Durant had acquired Frigidaire, and Frigidaire's market share of refrigerators exceeded 50 percent in the 1920s, the position was allowed to deteriorate as rivals developed market positions in other major appliances (radios, ranges, washers, etc.) while Frigidaire concentrated on refrigerators. The suggestion that GM get into air conditioners "did not register on us, and the proposal was not . . . adopted" (Sloan, 1965, p. 361). As Richard Burton and Arthur Kuhn conclude, GM's "deep and myopic involvement in the automobile sector of the economy, [prevented] product diversification opportunities in other market areas—even in product lines where GM had already achieved substantial penetration—[from being] recognized" (1979, pp. 10–11).

The conglomerate form of organization, whereby the corporation consciously took on a diversified character and nurtured its various parts, evidently required a conceptual break in the mind-set of Sloan and other prewar business leaders. This occurred gradually, more by evolution than by grand design (Robert Sobel, 1974, p. 377); and it involved a new group of organizational innovators—of which Royal Little was one (Sobel, 1974). The natural growth of conglomerates, which would occur as the techniques for managing diverse assets were refined, was accelerated as antitrust enforcement against horizontal and vertical mergers became progressively more severe. Conglomerate acquisitions—in terms of numbers, assets acquired, and as a proportion of total acquisitions—grew rapidly with the result

³⁴ For a discussion of the public policy issues posed by conglomerates, see Williamson (1975, pp. 163–71).

that "pure" conglomerate mergers, which in the period 1948–1953 constituted only 3 percent of the assets acquired by merger, had grown to 49 percent by 1973–1977 (Frederic Scherer, 1980, p. 124).

Morris Adelman's (1961) explanation for the conglomerate is that this form of organization has attractive portfolio diversification properties. But why should the conglomerate appear in the 1960s rather than much earlier? After all, holding companies, which long predated the conglomerate, can accomplish portfolio diversification. And individual stockholders, through mutual funds and otherwise, are able to diversify their own portfolios. At best the portfolio diversification thesis is a very incomplete explanation for the postwar wave of conglomerate mergers.³⁵

The Federal Trade Commission also ventured an early assessment of the conglomerate in which organization form features were ignored. The conglomerate was a natural target for the inhospitality tradition. Thus the FTC Staff held that the conglomerate had the following properties:

With the economic power which it secures through its operations in many diverse fields, the giant conglomerate corporation may attain an almost impregnable economic position. Threatened with competition in any one of its various activities, it may sell below cost in that field, offsetting its losses through profits made in its other lines—a practice which is frequently explained as one of meeting competition. The conglomerate corporation is thus in a position to strike out with great force against smaller business in a variety of different industries [1948, p. 59].

³⁵ The diversification of personal portfolios is not a perfect substitute for conglomerate diversification because bankruptcy has real costs that the firm, but not individuals, can reduce by portfolio diversification. Bankruptcy costs have not sharply increased in the past 30 years, however, hence these differences do not explain the appearance of the conglomerate during this interval.

I submit that some phenomena, of which changing internal organization is one, need to be addressed on their own terms. Adopting this view, the conglomerate is best understood as a logical outgrowth of the M-form mode for organizing complex economic affairs. Thus once the merits of the M-form structure for managing separable, albeit related, lines of business (e.g., a series of automobile or a series of chemical divisions) were recognized and digested, its extension to manage less closely related activities was natural. This is not to say that the management of product variety is without problems of its own. But the basic M-form logic, whereby strategic and operating decisions are distinguished and responsibilities are separated, carried over. The conglomerates in which M-form principles of organization are respected are usefully thought of as internal capital markets whereby cash flows from diverse sources are concentrated and directed to high yield uses.

The conglomerate is noteworthy, however, not merely because it permitted the M-form structure to take this diversification step. Equally interesting are the unanticipated systems consequences which developed as a byproduct. Thus once it was clear that the corporation could manage diverse assets in an effective way, the possibility of takeover by tender offer suggested itself. In principle, incumbent managements could always be displaced by waging a proxy contest. In fact, this is a very expensive and relatively ineffective way to achieve management change (Williamson, 1970, Chapter 6). Moreover, even if the dissident shareholders should succeed, there was still a problem of finding a successor management.

Viewed in contractual terms, the M-form conglomerate can be thought of as substituting an administrative interface between an operating division and the stockholders where a market interface had existed previously. Subject to the con-

dition that the conglomerate does not diversify to excess, in the sense that it cannot competently evaluate and allocate funds among the diverse activities in which it is engaged, the substitution of internal organization can have beneficial effects in goal pursuit, monitoring, staffing, and resource allocation respects. The goal-pursuit advantage is that which accrues to M-form organizations in general: since the general management of an M-form conglomerate is disengaged from operating matters, a presumption that the general office favors profits over functional goals is warranted. Relatedly, the general office can be regarded as an agent of the stockholders whose purpose is to monitor the operations of the constituent parts. Monitoring benefits are realized in the degree to which internal monitors enjoy advantages over external monitors in access to information—which they arguably do (Williamson, 1975, pp. 145–48). The differential ease with which the general office can change managers and reassign duties where performance failures or distortions are detected is responsible for the staffing advantage. Resource-allocation benefits are realized because cash flows no longer return automatically to their origins but instead revert to the center, thereafter to be allocated among competing uses in accordance with prospective yields.³⁶

This has a bearing on the problem of separation of ownership from control, noted by Adolph Berle and Gardiner C.

³⁶ To be sure, this substitution of internal organization for the capital market is subject to tradeoffs and diminishing returns. Breadth—that is, access to the widest range of alternatives—is traded off for depth—that is, more intimate knowledge of a narrower range of possible investment outlets—(Alchian and Harold Demsetz, 1972, p. 29), where the general office may be presumed to have the advantage in the latter respect. The diminishing returns feature suggests that the net benefits of increased diversity eventually become negative. Were further diversification thereafter to be attempted, effective control would pass back into the hands of the operating divisions with problematic performance consequences.

Means in 1932. Thus they inquired, “have we any justification for assuming that those in control of a modern corporation will also choose to operate it in the interests of the stockholders” (1932, p. 121). The answer, then as now, is almost certainly no. Indeed, the evident disparity of interest between managers and stockholders gave rise in the 1960s to what has become known as the managerial discretion literature (William Baumol, 1959; Robin Marris, 1964; Williamson, 1964).

There are important differences, however, between the U-form structure, which was the prevailing organization form at the time Berle and Means were writing, and the M-form structure, which in the U.S. was substantially in place by the 1960s. For one thing, as argued above, U-form managers identified more strongly with functional interests and hence were more given to subgoal pursuit. Secondly, and related, there was a confusion between strategic and operating goals in the U-form structure which the M-form served to rectify—with the result that the general office was more fully concerned with enterprise goals, of which profits is the leading element. Third, the market for corporate control, which remained ineffectual so long as the proxy contest was the only way to challenge incumbent managements, was activated as conglomerates recognized that tender offers could be used to effect corporate takeovers. As a consequence, managements that were otherwise secure and would have permitted managerial preferences to prevail were brought under scrutiny and induced to self-correct against egregious managerial distortions.

To be sure, managerial preferences (for salary and perquisites) and stockholder preferences for profits do not become perfectly consonant as a result of conglomerate organization and the associated activation of the capital market. The continuing tension between management and stock-

holder interests is evident in the numerous efforts that incumbent managements have taken to protect target firms against takeover (William Cary, 1969; Williamson, 1979; George Benston, 1980). Changes in internal organization have nevertheless relieved these concerns. A study of capitalist enterprises which makes no allowance for organization form changes and their capital market ramifications will naturally overlook the possibility that the corporate control dilemma posed by Berle and Means has since been alleviated more by *internal* than it has by regulatory or external organizational reforms.

Not all conglomerates respected M-form principles when they were first organized. The above argument applies only to those where rational decomposition principles were observed and leads to the following testable proposition: conglomerates that were organized along holding company rather than M-form lines (as many were initially) would be less able to cope when adversity appeared, at which time they would be reorganized as M-form firms. Voluntary divestiture is also an interesting conglomerate phenomenon. Such a rationalization of assets is commonly accompanied by internal organizational reforms. Growth maximization theories are mainly at a loss to explain such behavior.

4.3 *Multinational Enterprise*

The discussion of the multinational enterprise (MNE) that follows deals mainly with recent developments and, among these, emphasizes organizational aspects—particularly those associated with technology transfer in manufacturing industries. As Mira Wilkins has reported, direct foreign investment by American firms has a long history: the book value of cumulative U.S. direct foreign investment, expressed as a percentage of GNP, was in the range of 7 to 8 percent in 1914, 1929, and 1970 (Wilkins, 1974, p. 437).

Both the character of this investment and, relatedly, the organization structure within which this investment takes place have been changing, however. It is not accidental that the term MNE was coined neither in 1914 or 1929 but is of much more recent origin.

Thus whereas the ratio of the book value of U.S. foreign investments in manufacturing as compared with all other (petroleum; trade; mining; public utilities) was 0.47 in 1950, this had increased to 0.71 in 1970 (Wilkins, 1974, p. 329). Also, "what impressed Europeans about American plants in Europe and the United States [in 1929] was mass production, standardization, and scientific management; in the 1960s, Europeans were remarking that America's superiority was based on technological and managerial advantage . . . [and] that this expertise was being exported via direct investment" (Wilkins, 1974, p. 436).

The spread of the multinational corporation in the post World War II period has given rise to considerable scrutiny, some puzzlement, and even some alarm (Yoshihiro Tsurumi, 1977, p. 74). One of the reasons for this unsettled state of affairs is that transaction-cost economizing and organization form issues have been relatively neglected in efforts to assess MNE activity. An important exception is the work of Peter Buckley and Mark Casson (1976).

Organization form is relevant in two related respects. First is the matter of U.S.-based as compared with foreign-based investment rates. Tsurumi reports in this connection that the rate of foreign direct investments by U.S. firms increased rapidly after 1953, peaked in the mid-1960s, and has leveled off and declined since (Tsurumi, 1977, p. 97). The pattern of foreign direct investments by foreign firms, by contrast, has lagged that of the U.S. by about a decade (Tsurumi, 1977, pp. 91-92).

Recall that the conglomerate uses the

M-form structure to extend asset management from specialized to diversified lines of commerce. The MNE counterpart is the use of the M-form structure to extend asset management from a domestic base to include foreign operations. Thus the domestic M-form strategy for decomposing complex business structures into semi-autonomous operating units was subsequently applied to the management of foreign subsidiaries. As noted in 4.1 above, the transformation of the corporation along M-form lines came earlier in the U.S. than in Europe and elsewhere. U.S. corporations were for this reason better qualified to engage in foreign direct investments at an earlier date than were foreign-based firms. Only as the latter took on the M-form structure did this multinational management capability appear. The pattern of foreign direct investments recorded by Tsurumi and reported above is consistent with the temporal differences of U.S. and foreign firms in adopting the M-form structure.

That U.S. corporations possessed an M-form capability earlier than their foreign counterparts does not, however, establish that they used it to organize foreign investment. John Stopford and Louis Wells have studied this issue. They report that while initial foreign investments were usually organized as autonomous subsidiaries, divisional status within an M-form structure invariably appeared as the size and complexity of foreign operations increased (Stopford and Wells, 1972, p. 21). This transformation usually followed the organization of domestic operations along M-form lines (Stopford and Wells, 1972, p. 24). The adoption of a "global" strategy or "worldwide perspective"—whereby "strategic planning and major policy decisions" are made in the central office of the enterprise—could only be accomplished within a multidivisional framework (Stopford and Wells, 1972, p. 25).

Even more interesting than these orga-

nization form issues is the fact that foreign direct investments by U.S. firms have been concentrated in a few industries. Manufacturing industries that have made substantial foreign direct investments include chemicals, drugs, automobiles, food processing, electronics, electrical and non-electrical machinery, nonferrous metals, and rubber. Tobacco, textiles and apparel, furniture, printing, glass, steel, and aircraft have, by comparison, done little foreign direct investment (Tsurumi, 1977, p. 87).

Stephen Hymer's "dual" explanation for the multinational enterprise is of interest in this connection. Thus Hymer observes that direct foreign investment "allows business firms to transfer capital, technology, and organizational skill from one country to another. It is also an instrument for restraining competition between firms of different nations" (1970, p. 443).

Hymer is surely correct that the MNE can service both of these purposes and examples of both kinds can doubtlessly be found. It is nevertheless useful to ask whether the overall character of MNE investment, in terms of its distribution among industries, is more consistent with the efficiency purposes to which Hymer refers (transfer of capital, technology, and organizational skill) or with the oligopolistic restraint hypothesis. Adopting a transaction cost orientation discloses that the observed pattern of investment is more consistent with the efficiency part of Hymer's dual explanation.

For one thing, oligopolistic purposes can presumably be realized by portfolio investment coupled with a limited degree of management involvement to segregate markets. Put differently, direct foreign investment and the organization of foreign subsidiaries within an M-form structure are not needed to effect competitive restraints. Furthermore, if competitive restraints were mainly responsible for these investments, then presumably all concen-

trated industries—which would include tobacco, glass, and steel—rather than those associated with rapid technical progress would be active in MNE creation. Finally, although many of the leading U.S. firms that engaged in foreign direct investment enjoyed “market power,” this was by no means true for all.

By contrast, the pattern of foreign direct investments reported by Tsurumi appears to be consistent with a transaction cost economizing interpretation. Raymond Vernon’s 1970 study of the *Fortune* 500 corporations disclosed that 187 of these firms had a substantial multinational presence. R&D expenditures as a percentage of sales were higher among these 187 than among the remaining firms in the *Fortune* 500 group. Furthermore, according to Vernon, firms that went multinational tended to be technological innovators at the time of making their initial foreign direct investments.

This raises the question of the attributes of firms and markets for accomplishing technology transfer. The difficulties with transferring technology across market interface are of three kinds: recognition, disclosure, and team organization (Arrow, 1962; Williamson, 1975, pp. 31–33, 203–07; Teece, 1977).³⁷ Of these three, recognition is probably the least severe. To be sure, foreign firms may sometimes fail to perceive the opportunities to apply technological developments originated elsewhere. But enterprising domestic firms that have made the advance can be expected to identify at least some of the potential applications abroad.

Suppose, therefore, that recognition problems are set aside and consider disclosure. Technology transfer by contract can break down if convincing disclosure to buyers effectively destroys the basis for

exchange. A very severe information asymmetry problem exists, on which account the less informed party (in this instance the buyer) must be wary of opportunistic representations by the seller.³⁸ Although sometimes this asymmetry can be overcome by sufficient ex ante disclosure (and veracity checks thereon), this may shift rather than solve the difficulty. The “fundamental paradox” of information is that “its value for the purchaser is not known until he has the information, but then he has in effect acquired it without costs” (Arrow, 1971, p. 152).

Suppose, *arguendo*, that buyers concede value and are prepared to pay for information in the seller’s possession. The incentive to trade is then clear and for some items this will suffice. The formula for a chemical compound or the blueprints for a special device may be all that is needed to effect the transfer. Frequently, however, and probably often, new knowledge is diffusely distributed and is poorly defined (Nelson, 1981). Where the requisite information is distributed among a number of individuals all of whom understand their speciality in only a tacit, intuitive way, a simple contract to transfer the technology cannot be devised. See Michael Polanyi (1962).

Transfer need not cease, however, because simple contracts are not feasible. If the benefits of technology transfer are sufficiently great, exchange may be accomplished either by devising a complex trade or through direct foreign investment.

³⁸ Markets for information are apt to be especially costly and/or hazardous when transmission across a national boundary is attempted. Language differences naturally complicate the communication problem, and differences in the technological base compound these difficulties. If, moreover, as is commonly the case, cultural differences foster suspicion, the trust that is needed to support informational exchange may be lacking. Not only will contract negotiations be more complex and costly on this account, but execution will be subject to more formal and costly procedures than would obtain under a regime of greater trust.

³⁷ The material that follows is based on Williamson and Teece (1979). Our argument is similar to that advanced by Buckley and Casson (1976).

Which will be employed depends on the circumstances. If only a one-time (or very occasional) transfer of technology is contemplated, direct foreign investment is a somewhat extreme response.³⁹ The complex contractual alternative is to negotiate a tie-in sale whereby the technology and associated know-how are transferred as a package. Since the know-how is concentrated in the human assets who are already familiar with the technology, this entails the creation of a “consulting team” by the seller to accompany the physical technology transfer—the object being to overcome start up difficulties and to familiarize the employees of the foreign firm, through teaching and demonstration, with the idiosyncracies of the operation.⁴⁰

Inasmuch as many of the contingencies that arise in the execution of such contracts will be unforeseen and as it will be too costly to work out appropriate *ex ante* responses for others, such consulting contracts are subject to considerable strain. Where a succession of transfers is contemplated, which is to say, when the frequency shifts from occasional to recurring, complex contracting is apt to give way to direct foreign investment. A more harmonious and efficient exchange relation—better disclosure, easier reconciliation of differences, more complete cross-cultural adaptation, more effective team organization and reconfiguration—all predictably result from the substitution of an internal governance relation for bilateral trading under these recurrent trading circumstances for assets, of which complex technology transfer is an example, that have a highly specific character.⁴¹

³⁹ This is an implication of transaction cost reasoning in which the frequency dimension has explanatory power (Williamson, 1979, pp. 245–54).

⁴⁰ On the importance of on-site observation and of teaching-by-doing, see Polanyi (1962), Peter Doeringer and Michael Piore (1971, pp. 15–16), and Williamson, Michael Wachter, and Jeffrey Harris (1975).

⁴¹ The argument can be extended to deal with such observations as those of Edwin Mansfield, Anthony

The upshot is that while puzzlement with and concerns over MNEs will surely continue, a transaction cost interpretation of this phenomenon sheds insight on the following conspicuous features of multinational investment: (1) the reported concentration of foreign direct investment in manufacturing industries where technology transfer is of special importance; (2) the organization of these investments within M-form structures; and (3) the differential timing of foreign direct investment between U.S. and foreign manufacturing enterprises (which difference also has organization form origins). I furthermore conjecture that the application of transaction cost reasoning will lead to a deeper understanding of other specific features of MNE activity as these are discovered and/or become subject to public policy scrutiny.

5. *Concluding Remarks*

There is widespread agreement, among economists and noneconomists alike, with the proposition that the modern corporation is an important and complex economic institution. Such agreement is mainly explained by the obtrusive size of the largest firms—running to tens of billions of dollars of assets and sales, with employment numbering in the hundreds of thousands. The economic factors that lie behind the size, shape, and performance of the modern corporation, however, are poorly understood.

This puzzlement is not of recent origin. Edward Mason complained over twenty years ago that “the functioning of the corporate system has not to date been ade-

Romeo and Samuel Wagner (1979), who report that firms use subsidiaries to transfer their newest technology overseas but rely on licensing or joint ventures for older technology. The transaction cost argument is that the latter are more well defined, hence are more easily reduced to contract, and require less firm specific know-how to effect successful transfer.

quately explained. . . . The man of action may be content with a system that works. But one who reflects on the properties or characteristics of this system cannot help asking why it works and whether it will continue to work" (1960, p. 4). The predicament to which Mason refers is, I submit, largely the product of two different (but not unrelated) intellectual traditions. The first of these holds that the structural features of the corporation are irrelevant. This is the neoclassical theory of the firm that populates intermediate theory textbooks. Structural differences are suppressed as the firm is described as a production function to which a profit maximization objective has been assigned. The second has public policy roots; this is the inhospitality tradition that I referred to earlier. In this tradition, distinctive structural features of the corporation are believed to be the result of unwanted (anti-competitive) intrusions into market processes.

The transaction-cost approach differs from both. Unlike neoclassical analysis, internal organization is specifically held to be important. Unlike the inhospitality tradition, structural differences are assumed to arise primarily in order to promote economy in transaction costs. The assignment of transactions between firms and markets and the economic ramifications of internal structure both come under scrutiny in these terms. The application of these ideas to the study of transactions in general and of the modern corporation in particular requires that (1) the transaction be made the principal unit of analysis, (2) an elementary appreciation for "human nature as we know it" supplant the fiction of economic man, (3) transactions be dimensionalized, (4) rudimentary principles of market and hierarchical organization be recognized, and (5) a guiding principle of comparative institutional study be the hypothesis that transactions are assigned to and organized within gov-

ernance structures in a discriminating (transaction-cost economizing) way.

The view that the corporation is first and foremost an efficiency instrument does not deny that firms also seek to monopolize markets, sometimes by engaging in strategic behavior, or that managers sometimes pursue their own goals to the detriment of system goals. But specific structural preconditions need to be satisfied if strategic behavior is to be feasible⁴²—and most firms do not qualify, which is to say that strategic behavior is the exception rather than the rule. Furthermore, most firms will be penalized if efficiency norms are seriously violated for extended periods of time—which serves to curb managerial discretion. The strongest argument favoring transaction cost economizing, however, is that this is the only hypothesis that is able to provide a discriminating rationale for the succession of organizational innovations that have occurred over the past 150 years and out of which the modern corporation has emerged.

To recapitulate, although railroad mergers between parallel roads can have monopolizing purposes, the joining of end-to-end systems under common management is explained by transaction cost economics. The hierarchical structures evolved by the railroads were the outcome of internal efforts to effect coordination across interfaces to which common operating responsibilities had been assigned. Older and simpler structures were unable to manage such complex networks, while coordination by end-to-end contracts between successive stations was prohibitively costly.

Forward integration out of manufacturing into distribution was widespread at the turn of the century. More interesting,

⁴² For a discussion of these preconditions—mainly high concentration coupled with high barriers to entry—see Joskow and Klevorick (1979), and Williamson (1981).

however, than this general movement is the fact that forward integration was selective—being extensive in some industries (e.g., sewing machines), negligible in others (e.g., dry goods), and mistaken in still others (e.g., sugar). This selective pattern is predicted by and consistent with transaction-cost reasoning—whereas no other hypothesis makes comparably detailed predictions.

The efficiency incentive to shift from the earlier U-form to the M-form structure is partly explained in managerial discretion terms: the older structure was more subject to distortions of a managerial discretion kind—which is to say that opportunism had become a serious problem in the large U-form firm. Equally and probably more important, however, is that the managerial hierarchy is the U-form enterprise was simply overburdened as the firm became large and complex. The M-form structure represented a more rational decomposition of the affairs of the firm and thereby served to economize on bounded rationality.⁴³ The subsequent diffusion of this structure was hastened by a combination of product market (pressure on rivals) and capital market (takeover) competition.

The M-form structure, which was originally adopted by firms in relatively specialized lines of commerce was subsequently extended to manage diversified assets (the conglomerate) and foreign direct investments (MNE). A breadth-for-depth tradeoff is involved in the former case, as the firm selectively internalizes functions ordinarily associated with the capital market. MNE activity has also been selective—being concentrated in the more technologically progressive industries where higher rates of R&D are reported and technology transfer arguably

⁴³ Had “normal” managerial preferences prevailed, the U-form, which favored the exercise of those preferences, would presumably have been retained.

poses greater difficulties than is true of technologically less progressive industries. This pattern of foreign direct investment cannot be explained as the pursuit of monopoly but is consistent with transaction-cost reasoning.

The upshot is that a transaction-cost approach to the study of the modern corporation permits a wide variety of significant organizational events to be interpreted in a coherent way.⁴⁴ It links up comfortably with the type of business history studies that have been pioneered by Chandler. It has ramifications for the study of regulation (Williamson, 1976; Goldberg, 1976) and for antitrust enforcement. Applications to aspects of labor economics and comparative systems have been made, and others would appear to be fruitful. More generally, while there is room for and need for refinement, a comparative approach to the study of economic institutions in which the economy of transaction costs is the focus of analysis, appears to have considerable promise.

REFERENCES

- ADELMAN, M. A. “The Antimerger Act, 1950–1960,” *Amer. Econ. Rev.*, May 1961, 51, pp. 236–44.
ALCHIAN, A. A. “Corporate Management and Prop-

⁴⁴ Recent contributions to the theory of the firm that are held to have a bearing on the study of the modern corporation are Alchian and Demsetz (1972) and Michael Jensen and William Meckling (1976). Both, however, deal with a microcosm much smaller than the modern corporation. Thus Alchian and Demsetz focus on the reasons why technological nonseparabilities give rise to team organization. Although small groups may be explained in this way (manual freight loading, whereby two men are required to lift coordinately, is the standard example), the existence of complex hierarchies cannot be explained in terms of the imperatives of such nonseparabilities. (The largest work group which, to my knowledge, qualifies is the symphony orchestra.)

Similarly, while the Jensen and Meckling paper is an important contribution to the principal-agent literature, it does not generalize to the modern corporation—as they expressly acknowledge (1976, p. 356). Although they conjecture that their analysis can be applied to the large, diffusely owned corporation whose managers own little or no equity (1976, p. 356), I have serious doubts.

- erty Rights," in *Economic policy and regulation of corporate securities*. Edited by H. G. MANNE. Washington: American Enterprise Institute for Public Policy Research, 1969, pp. 337-60.
- _____. AND DEMSETZ, H. "Production, Information Costs, and Economic Organization," *Amer. Econ. Rev.*, Dec. 1972, 62(5), pp. 777-95.
- ARROW, KENNETH J. "Economic Welfare and the Allocation of Resources of Invention," in *The rate and direction of inventive activity: Economic and social factors*. Edited by NATIONAL BUREAU OF ECONOMIC RESEARCH. Princeton: Princeton University Press, 1962, pp. 609-25.
- _____. "Control in Large Organizations," *Manage. Science*, April 1964, 10(3), pp. 397-408.
- _____. "The Organization of Economic Activity: Issues Pertinent to the Choice of Market Versus Nonmarket Allocation," in *The analysis and evaluation of public expenditure: The PPB system*, Vol. 1. U.S. Joint Economic Committee, 91st Congress, 1st Session, U.S. Government Printing Office, 1969, pp. 59-73.
- _____. *Essay in the theory of risk-bearing*. Chicago: Markham Pub. Co., 1971.
- _____. *The limits of organization*. First edition. New York: W. W. Norton & Co., 1974.
- ASHBY, W. R. *An introduction to cybernetics*. New York: John Wiley and Sons, 1956.
- BARNARD, C. I. *The functions of the executive*. Cambridge, Mass.: Harvard University Press, 1938.
- BAUMOL, W. J. *Business behavior, value and growth*. New York: Macmillan, 1959; Harcourt, Brace and World, 1967.
- BENSTON, GEORGE J. *Conglomerate mergers: Causes, consequences and remedies*. Washington, D.C.: American Enterprise Institute for Public Policy Research, 1980.
- BERLE, A. A. AND MEANS, G. C. *The modern corporation and private property*. New York: Macmillan, 1932.
- BRUCHEY, STUART W. *Robert Oliver, merchant of Baltimore, 1783-1819*. Baltimore: Johns Hopkins University Press, 1956.
- BUCKLEY, P. J. AND CASSON, M. *The future of multinational enterprise*. New York: Holmes and Meier, 1976.
- BURTON, R. H. AND KUHN, A. J. "Strategy Follows Structure: The Missing Link of Their Intertwined Relation," Working Paper No. 260, Fuqua School of Business, Duke University, May 1979.
- CARY, W. "Corporate Devices Used to Insulate Management from Attack," *Antitrust Law Journal*, 1969-70, 39(1), pp. 318-24.
- CAVES, R. E. "Corporate Strategy and Structure," *J. Econ. Lit.*, March 1980, 18(1), pp. 64-92.
- CHANDLER, A. D., JR. *Strategy and structure: Chapters in the history of the industrial enterprise*. Cambridge, Mass.: MIT Press, 1962; Garden City, N.J.: Doubleday & Co., 1966.
- _____. *The visible hand: The managerial revolution in American business*. Cambridge, Mass.: Belknap Press, 1977.
- COASE, R. H. "The Nature of the Firm," *Economica N.S.*, 1937, 4, pp. 386-405; and in *Readings in price theory*. Edited by G. J. STIGLER AND K. E. BOULDING. Chicago: Richard D. Irwin for the American Economic Association, 1952.
- _____. "Industrial Organization: A Proposal for Research," in *Policy issues and research opportunities in industrial organization: Economic research: Retrospect and prospect*. Edited by VICTOR R. FUCHS. New York: NBER; distributed by Columbia University Press, New York and London, 1972, pp. 59-73.
- COMMONS, JOHN R. *Institutional economics; its place in political economy*. New York: Macmillan, [1934] 1951.
- CYERT, RICHARD M. AND HEDRICK, CHARLES L., "Theory of the Firm: Past, Present, and Future; An Interpretation," *J. Econ. Lit.*, June 1972, 10(2), pp. 398-412.
- DAVIS, LANCE E. AND NORTH, DOUGLASS C. *Institutional change and American economic growth*. Cambridge, England: Cambridge University Press, 1971.
- DOERINGER, P. AND PIORI, M. *Internal labor markets and manpower analysis*. Boston: D.C. Heath and Co., 1971.
- DRUCKER, P. *Management: Tasks, responsibilities, practices*. New York: Harper & Row, 1974.
- FISHLOW, ALBERT. *American railroads and the transformation of the antebellum economy*. Cambridge, Mass.: Harvard University Press, 1965.
- FOGEL, WILLIAM R. *Railroads and American economic growth: Essays in econometric history*. Baltimore: Johns Hopkins University Press, 1964.
- FRANKO, LAWRENCE G. "The Growth, Organizational Efficiency of European Multinational Firms: Some Emerging Hypotheses," *Colloques international aux C.N.R.S.*, 1972, pp. 335-66.
- GOLDBERG, V. P. "Regulation and Administered Contracts," *Bell J. Econ.*, Autumn 1976, 7(2), pp. 426-52.
- HAYEK, F. "The Use of Knowledge in Society," *Amer. Econ. Rev.*, Sept. 1945, 35, pp. 519-30.
- HEFLEBOWER, R. B. "Observation on Decentralization in Large Enterprises," *J. Ind. Econ.*, Nov. 1960, 9, pp. 7-22.
- HYMER, S. "The Efficiency (Contradictions) of Multinational Corporations," *Amer. Econ. Rev.*, May 1970, 60(2), pp. 441-48.
- JENSEN, M. C. AND MECKLING, W. H. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *J. Finan. Econ.*, Oct. 1976, 3(4), pp. 305-60.
- JOSKOW, PAUL L. AND KLEVORICK, ALVIN K. "A Framework for Analyzing Predatory Pricing Policy," *Yale Law J.*, Dec. 1979, 89, pp. 213-70.
- KLEIN, B.; CRAWFORD, R. A. AND ALCHIAN, A. A. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *J. Law Econ.*, Oct. 1978, 21(2), pp. 297-326.
- KNIGHT, FRANK H. *Risk, uncertainty and profit*. New York: Harper & Row, [1921] 1965.
- KOONTZ, H. AND O'DONNELL, C. *Principles of management; an analysis of managerial functions*. New York: McGraw-Hill, 1955.
- LARSON, HENRIETTA M. *Guide to business history;*

- materials for the study of American business history and suggestions for their use. Cambridge, Mass.: Harvard University Press, 1948.
- LIVESAY, H. C. *American made: Men who shaped the American economy*. First edition. Boston: Little, Brown, 1979.
- MANSFIELD, E.; ROMEO, A. AND WAGNER, S. "Foreign Trade and U.S. Research and Development," *Rev. Econ. Statist.*, Feb. 1979, 61(1), pp. 49-57.
- MARRIS, R. *The economic theory of managerial capitalism*. New York: Free Press, 1964.
- ____ AND MUELLER, D. C. "The Corporation, Competition, and the Invisible Hand," *J. Econ. Lit.*, March 1980, 18(1), pp. 32-63.
- MARSCHAK, J. AND RADNER, R. *Economic theory of teams*. New Haven: Yale University Press, 1972.
- MASON, E. S. "Introduction," in *The corporation in modern society*. Edited by E. S. MASON. Cambridge, Mass.: Harvard University Press, 1960, pp. 1-24.
- NELSON, R. R. "Assessing Private Enterprise: An Exegesis of Tangled Doctrine," *Bell J. Econ.*, Spring 1981, 12(1), pp. 93-111.
- ____ AND WINTER, S. G. *An evolutionary theory of economic behavior and capabilities*. Cambridge, Mass.: Harvard University Press, 1981.
- NORTH, D. C. "Structure and Performance: The Task of Economic History," *J. Econ. Lit.*, Sept. 1978, 16(3), pp. 963-78.
- PANZAR, JOHN C. AND WILLIG, ROBERT D. "Economies of Scope," *Amer. Econ. Rev.*, *Papers and Proceedings*, May 1981, 71(2), pp. 268-72.
- POLANYI, M. *Personal knowledge: Towards a post-critical philosophy*. New York: Harper & Row, 1962.
- PORTER, G. AND LIVESAY, H. C. *Merchants and manufacturers: Studies in the changing structure of nineteenth century marketing*. Baltimore: Johns Hopkins University Press, 1971.
- POSNER, R. A. "The Chicago School of Antitrust Analysis," *Univ. Pennsylvania Law Rev.*, April 1979, 127(4), pp. 925-48.
- RADNER, ROY. "Competitive Equilibrium Under Uncertainty," *Econometrica*, Jan. 1968, 36(1), pp. 31-58.
- SCHERER, F. M. *Industrial market structure and economic performance*. Second edition. Chicago: Rand McNally College Pub. Co., 1980.
- SIMON, H. A. *Models of man: Social and rational mathematical essays on rational human behavior in a social setting*. New York: John Wiley and Sons, 1957a.
- ____. *Administrative behavior; a study of decision-making processes in administrative organization*. Second edition. New York: Macmillan, [1947] 1957b.
- ____. "The Architecture of Complexity," *Proceedings of the American Philosophical Society*, Dec. 1962, 106(6), pp. 467-82.
- ____. "Applying Information Technology to Organization Design," *Pub. Admin. Rev.*, May-June 1973, 33(3), pp. 268-78.
- ____. "Rationality as Process and as Product of Thought," *Amer. Econ. Rev.*, May 1978, 68(2), pp. 1-16.
- SLOAN, A. P., JR. *My years with General Motors*. New York: MacFadden-Bartell, [1963] 1965.
- SOBEL, R. *The entrepreneurs: Explorations within the American business tradition*. New York: Weybright and Talley, 1974.
- STOPFORD, JOHN M. AND WELLS, LOUIS T., JR. *Managing the multinational enterprise; organization of the firm and ownership of the subsidiaries*. New York: Basic Books, 1972.
- TEECE, D. J. "Technology Transfer by Multinational Firms," *Econ. J.*, June 1977, 87, pp. 242-61.
- ____. "Economies of Scope and the Scope of the Enterprise," *J. Econ. Behavior Org.*, Sept. 1980, 1(3), pp. 223-45.
- TEMIN, P. "The Future of the New Economic History," *J. of Interdisciplinary Hist.*, Autumn 1981, 12(2), pp. 179-97.
- THOMPSON, JAMES D. *Organizations in action; social science bases of administrative theory*. New York: McGraw-Hill, 1967.
- TSURUMI, Y. *Multinational management: Business strategy and government policy*. Cambridge, Mass.: Ballinger, 1977.
- U.S. FEDERAL TRADE COMMISSION. *Report of the Federal Trade Commission on the merger movement: A summary report*, Washington, D.C.: U.S. Government Printing Office, 1948.
- VERNON, R. *Sovereignty at bay: The multinational spread of U.S. enterprises*. New York: Basic Books, 1971.
- WILKINS, MIRA. *The maturing of multinational enterprise: American business abroad from 1914 to 1970*. Cambridge, Mass.: Harvard University Press, 1974.
- WILLIAMSON, O. E. *The economics of discretionary behavior: Managerial objectives in a theory of the firm*. Englewood Cliffs, N.J.: Prentice-Hall, 1964.
- ____. *Corporate control and business behavior*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- ____. *Markets and hierarchies: Analysis and anti-trust implications: A study in the economics of internal organization*. New York: Free Press, 1975.
- ____. "Franchise Bidding for Natural Monopolies—in General and with Respect to CATV," *Bell J. Econ.*, Spring 1976, 7(1), pp. 73-104.
- ____. "Transaction-Cost Economics: The Governance of Contractual Relations," *J. Law Econ.*, Oct. 1979, 22(2), pp. 233-61.
- ____. "On the Governance of the Modern Corporation," *Hofstra Law Rev.*, Fall 1979, 8(1), pp. 63-78.
- ____. "Organizational Innovation: The Transaction-Cost Approach." Discussion Paper No. 82, Center for the Study of Organizational Innovation, University of Pennsylvania, Sept. 1980.
- ____. "Antitrust Enforcement: Where It's Been; Where It's Going." Discussion Paper No. 102, Center for the Study of Organizational Innovation, University of Pennsylvania, May 1981.

——— AND TEECE, D. J. "European Economic and Political Integration: The Markets and Hierarchies Approach," in *New approaches to European integration*. Edited by P. SALMON. Forthcoming.

———; WACHTER, MICHAEL L. AND HARRIS, JEFFREY E. "Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange," *Bell J. Econ.*, Spring 1975, 6(1), pp. 50–280.