

# 4

## Comparative Economic Organization: The Analysis of Discrete Structural Alternatives

Although microeconomic organization is formidably complex and has long resisted systematic analysis, that has been changing as new modes of analysis have become available, as recognition of the importance of institutions to economic performance has grown, and as the limits of earlier modes of analysis have become evident. Information economics, game theory, agency theory, and population ecology have all made significant advances.

This chapter approaches the study of economic organization from a comparative institutional point of view in which transaction-cost economizing is featured. Comparative economic organization never examines organization forms separately but always in relation to alternatives. Transaction-cost economics places the principal burden of analysis on comparisons of transaction costs—which, broadly, are the “costs of running the economic system” (Arrow, 1969, p. 48).

My purpose in this chapter is to extend and refine the apparatus out of which transaction-cost economics works, thereby to respond to some of the leading criticisms. Four objections to prior work in this area are especially pertinent. One objection is that the two stages of the new institutional economics research agenda—the institutional environment and the institutions of governance—have developed in disjunct ways. The first of these paints on a very large historical canvas and emphasizes the institutional rules of the game: customs, laws, politics (North, 1986). The latter is much more microanalytic and focuses on the comparative efficacy with which alternative generic forms of governance—markets, hybrids, hierarchies—economize on transaction costs. Can this disjunction problem be overcome? Second, transaction-cost economics has been criticized because it deals with polar forms—markets and hierarchies—to the neglect of intermediate or hybrid forms. Although that objection has begun to be addressed by recent treatments of long-term contracting in which bilateral dependency conditions are supported by a variety of specialized

governance features (hostages, arbitration, take-or-pay procurement clauses, tied sales, reciprocity, regulation, etc.), the abstract attributes that characterize alternative modes of governance have remained obscure. What are the key attributes and how do they vary among forms? This is responsive to the third objection, namely, that efforts to operationalize transaction-cost economics have given disproportionate attention to the abstract description of transactions as compared with the abstract description of governance. The dimensionalization of both is needed. Finally, there is the embeddedness problem: Transaction-cost economics purports to have general application but has been developed almost entirely with reference to Western capitalist economies (Hamilton and Biggart, 1988). Is a unified treatment of Western and non-Western, capitalist and noncapitalist economies really feasible? This paper attempts to address these objections by posing the problem of organization as one of discrete structural analysis.

## 1. Discrete Structural Analysis

The term discrete structural analysis was introduced into the study of comparative economic organization by Simon, who observed that

As economics expands beyond its central core of price theory, and its central concern with quantities of commodities and money, we observe in it . . . [a] shift from a highly quantitative analysis, in which equilibration at the margin plays a central role, to a much more qualitative institutional analysis, in which discrete structural alternatives are compared. . . .

[S]uch analyses can often be carried out without elaborate mathematical apparatus or marginal calculation. 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, pp. 6–7).

But what exactly is discrete structural analysis? Is it employed only because “there is at present no [satisfactory] way of characterizing organizations in terms of continuous variation over a spectrum” (Ward, 1967, p. 38)? Or is there a deeper rationale?

Of the variety of factors that support discrete structural analysis, I focus here on the following: (1) firms are not merely extensions of markets but employ different means, (2) discrete contract law differences provide crucial support for and serve to define each generic form of governance, and (3) marginal analysis is typically concerned with second-order refinements to the neglect of first-order economizing.

### 1.1. Different Means

Although the study of economic organization deals principally with markets and market mechanisms, it is haunted by a troublesome fact: a great deal of economic activity takes place within firms (Barnard, 1938; Chandler, 1962,

1977). Conceivably, however, no novel economizing issues are posed within firms, because technology is largely determinative—the firm is mainly defined by economies of scale and scope and is merely an instrument for transforming inputs into outputs according to the laws of technology—and because market mechanisms carry over into firms. I have taken exception with the technology view elsewhere (Williamson, 1975). Consider, therefore, the latter.

In parallel with von Clausewitz's (1980) views on war, I maintain that hierarchy is not merely a contractual act but is also a contractual instrument, a continuation of market relations by other means. The challenge to comparative contractual analysis is to discern and explicate the different means. As developed below, each viable form of governance—market, hybrid, and hierarchy—is defined by a syndrome of attributes that bear a supporting relation to one another. Many hypothetical forms of organization never arise, or quickly die out, because they combine inconsistent features.

## 1.2. Contract Law

The mapping of contract law onto economic organization has been examined elsewhere (Williamson, 1979b, 1985a). Although some of that is repeated here, there are two significant differences. First, I advance the hypothesis that each generic form of governance—market, hybrid, and hierarchy—needs to be supported by a different form of contract law. Second, the form of contract law that supports hierarchy is that of forbearance.

### 1.2.1. *Classical contract law*

Classical contract law applies to the ideal transaction in law and economics in which the identity of the parties is irrelevant. "Thick" markets are ones in which individual buyers and sellers bear no dependency relation to each other. Instead, each party can go its own way at negligible cost to another. If contracts are renewed period by period, that is only because current suppliers are continuously meeting bids in the spot market. Such transactions are monetized in extreme degree; contract law is interpreted in a very legalistic way: more formal terms supersede less formal should disputes arise between formal and less formal features (e.g., written agreements versus oral amendments), and hard bargaining, to which the rules of contract law are strictly applied, characterizes these transactions. Classical contract law is congruent with and supports the autonomous market form of organization (Macneil, 1974, 1978).

### 1.2.2. *Neoclassical contract law and excuse doctrine*

Neoclassical contract law and excuse doctrine, which relieves parties from strict enforcement, apply to contracts in which the parties to the transaction maintain autonomy but are bilaterally dependent to a nontrivial degree. Identity plainly matters if premature termination or persistent maladaptation would place burdens on one or both parties. Perceptive parties reject classical contract

law and move into a neoclassical contracting regime because this better facilitates continuity and promotes efficient adaptation.

As developed below, hybrid modes of contracting are supported by neoclassical contract law. The parties to such contracts maintain autonomy, but the contract is mediated by an elastic contracting mechanism. Public utility regulation, in which the relations between public utility firms and their customers are mediated by a regulatory agency, is one example (Goldberg, 1976a; Williamson, 1976). Exchange agreements or reciprocal trading in which the parties experience (and respond similarly to) similar disturbances is another illustration (Williamson, 1983). Franchising is another way of preserving semi-autonomy, but added supports are needed (Klein, 1980; Hadfield, 1990). More generally, long-term, incomplete contracts require special adaptive mechanisms to effect realignment and restore efficiency when beset by unanticipated disturbances.

Disturbances are of three kinds: inconsequential, consequential, and highly consequential. Inconsequential disturbances are ones for which the deviation from efficiency is too small to recover the costs of adjustment. The net gains from realignment are negative for minor disturbances because (as discussed below) requests for adjustments need to be justified and are subject to review, the costs of which exceed the prospective gains.

Middle-range or consequential disturbances are ones to which neoclassical contract law applies. These are transactions for which Karl Llewellyn's concept of "contract as framework" is pertinent (1931, p. 737). The thirty-two-year coal supply agreement between the Nevada Power Company and the Northwest Trading Company illustrates the elastic mechanisms employed by a neoclassical contract. That contract reads in part as follows:

In the event an inequitable condition occurs which adversely affects one Party, it shall then be the joint and equal responsibility of both Parties to act promptly and in good faith to determine the action required to cure or adjust for the inequity and effectively to implement such action. Upon written claim of inequity served by one Party upon the other, the Parties shall act jointly to reach an agreement concerning the claimed inequity within sixty (60) days of the date of such written claim. An adjusted base coal price that differs from market price by more than ten percent (10%) shall constitute a hardship. The Party claiming inequity shall include in its claim such information and data as may be reasonably necessary to substantiate the claim and shall freely and without delay furnish such other information and data as the other Party reasonably may deem relevant and necessary. If the Parties cannot reach agreement within sixty (60) days the matter shall be submitted to arbitration.

By contrast with a classical contract, this contract (1) contemplates unanticipated disturbances for which adaptation is needed, (2) provides a tolerance zone (of  $\pm 10\%$ ) within which misalignments will be absorbed, (3) requires information disclosure and substantiation if adaptation is proposed, and (4) provides for arbitration in the event voluntary agreement fails.

The forum to which this neoclassical contract refers disputes is (initially at least) that of arbitration rather than the courts. Fuller described the procedural differences between arbitration and litigation:

[T]here are open to the arbitrator . . . quick methods of education not open to the courts. An arbitrator will frequently interrupt the examination of witnesses with a request that the parties educate him to the point where he can understand the testimony being received. This education can proceed informally, with frequent interruptions by the arbitrator, and by informed persons on either side, when a point needs clarification. Sometimes there will be arguments across the table, occasionally even within each of the separate camps. The end result will usually be a clarification that will enable everyone to proceed more intelligently with the case. (1963, pp. 11–12)

Such adaptability notwithstanding, neoclassical contracts are not indefinitely elastic. As disturbances become highly consequential, neoclassical contracts experience real strain, because the autonomous ownership status of the parties continuously poses an incentive to defect. The general proposition here is that when the “lawful” gains to be had by insistence upon literal enforcement exceed the discounted value of continuing the exchange relationship, defection from the spirit of the contract can be anticipated.

When, in effect, arbitration gives way to litigation, accommodation can no longer be presumed. Instead, the contract reverts to a much more legalistic regime—although, even here, neoclassical contract law averts truly punitive consequences by permitting appeal to exceptions that qualify under some form of excuse doctrine. The legal system’s commitment to the keeping of promises under neoclassical contract law is modest (Macneil, 1974, p. 731).

From an economic point of view, the tradeoff that needs to be faced in excusing contract performance is between stronger incentives and reduced opportunism. If the state realization in question was unforeseen and unforeseeable (different in degree and/or especially in kind from the range of normal business experience), if strict enforcement would have truly punitive consequences, and especially if the resulting “injustice” is supported by (lawful) opportunism, then excuse can be seen mainly as a way of mitigating opportunism, ideally without adverse impact on incentives. If, however, excuse is granted routinely whenever adversity occurs, then incentives to think through contracts, choose technologies judiciously, share risks efficiently, and avert adversity will be impaired. Excuse doctrine should therefore be used sparingly—which it evidently is (Farnsworth, 1968, p. 885; Buxbaum, 1985).

The relief afforded by excuse doctrine notwithstanding, neoclassical contracts deal with consequential disturbances only at great cost: arbitration is costly to administer and its adaptive range is limited. As consequential disturbances and, especially, as highly consequential disturbances become more frequent, the hybrid mode supported by arbitration and excuse doctrine incurs added costs and comes under added strain. Even more elastic and adaptive arrangements warrant consideration.

### 1.2.3. *Forbearance*

Internal organization, hierarchy, qualifies as a still more elastic and adaptive mode of organization. What type of contract law applies to internal organization? How does this have a bearing on contract performance?

Describing the firm as a “nexus of contracts” (Alchian and Demsetz, 1972; Jensen and Meckling, 1976; Fama, 1980) suggests that the firm is no different from the market in contractual respects. Alchian and Demsetz originally took the position that the relation between a shopper and his grocer and that between an employer and employee was identical in contractual respects:

The single consumer can assign his grocer to the task of obtaining whatever the customer can induce the grocer to provide at a price acceptable to both parties. That is precisely all that an employer can do to an employee. To speak of managing, directing, or assigning workers to various tasks is a deceptive way of noting that the employer continually is involved in renegotiation of contracts on terms that must be acceptable to both parties. . . . Long-term contracts between employer and employee are not the essence of the organization we call a firm. (1972, p. 777)

That it has been instructive to view the firm as a nexus of contracts is evident from the numerous insights that this literature has generated. But to regard the corporation only as a nexus of contracts misses much of what is truly distinctive about this mode of governance. As developed below, bilateral adaptation effected through fiat is a distinguishing feature of internal organization. But wherein do the fiat differences between market and hierarchy arise? If, moreover, hierarchy enjoys an “advantage” with respect to fiat, why can’t the market replicate this?

One explanation is that fiat has its origins in the employment contract (Coase, 1937; Barnard, 1938; Simon, 1951; Masten, 1988). Although there is a good deal to be said for that explanation, I propose a separate and complementary explanation: The implicit contract law of internal organization is that of forbearance. Thus, whereas courts routinely grant standing to firms should there be disputes over prices, the damages to be ascribed to delays, failures of quality, and the like, courts will refuse to hear disputes between one internal division and another over identical technical issues. Access to the courts being denied, the parties must resolve their differences internally. Accordingly, hierarchy is its own court of ultimate appeal.

What is known as the “business judgment rule” holds that “absent bad faith or some other corrupt motive, directors are normally not liable to the corporation for mistakes of judgment, whether those mistakes are classified as mistakes of fact or mistakes of law” (Gilson, 1986, p. 741). Not only does that rule serve as “a quasi-jurisdictional barrier to prevent courts from exercising regulatory powers over the activities of corporate managers” (Manne, 1967, p. 271), but “The courts’ abdication of regulatory authority through the business judgment rule may well be the most significant common law contribution to corporate governance” (Gilson, 1986, p. 741). The business judgment rule, which applies to the relation between shareholders and directors, can be interpreted as a particular manifestation of forbearance doctrine, which applies to the management of the firm more generally. To review alleged mistakes of judgment or to adjudicate internal disputes would sorely test the competence of courts and would undermine the efficacy of hierarchy.

Accordingly, the reason why the market is unable to replicate the firm with respect to fiat is that market transactions are defined by contract law of an altogether different kind. There is a logic to classical market contracting and there is a logic for forbearance law, and the choice of one regime precludes the other. Whether a transaction is organized as make or buy—internal procurement or market procurement, respectively—thus matters greatly in dispute-resolution respects: the courts will hear disputes of the one kind and will refuse to be drawn into the resolution of disputes of the other. Internal disputes between one division and another regarding the appropriate transfer prices, the damages to be ascribed to delays, failures of quality, and the like, are thus denied a court hearing.

To be sure, not all disputes within firms are technical. Personnel disputes are more complicated. Issues of worker safety, dignity, the limits of the “zone of acceptance,” and the like sometimes pose societal spillover costs that are undervalued in the firm’s private net benefit calculus. Underprovision of human and worker rights could ensue if the courts refused to consider issues of these kinds. Also, executive compensation agreements can sometimes be written in ways that make it difficult to draw a sharp line between personnel and technical issues. Even with personnel disputes, however, there is a presumption that such differences will be resolved internally. For example, unions may refuse to bring individual grievances to arbitration:

[G]iving the union control over all claims arising under the collective agreement comports so much better with the functional nature of a collective bargaining agreement. . . . Allowing an individual to carry a claim to arbitration whenever he is dissatisfied with the adjustment worked out by the company and the union . . . discourages the kind of day-to-day cooperation between company and union which is normally the mark of sound industrial relations—a relationship in which grievances are treated as problems to be solved and contracts are only guideposts in a dynamic human relationship. When . . . the individual’s claim endangers group interests, the union’s function is to resolve the competition by reaching an accommodation or striking a balance. (Cox, 1958, p. 24)

As compared with markets, internal incentives in hierarchies are flat or low-powered, which is to say that changes in effort expended have little or no immediate effect on compensation. This is mainly because the high-powered incentives of markets are unavoidably compromised by internal organization (Williamson, 1985b, chap. 6; 1988d). Also, however, hierarchy uses flat incentives because these elicit greater cooperation and because unwanted side effects are checked by added internal controls (see Williamson, 1988d; Holmstrom, 1989). Not only, therefore, will workers and managers be more willing to accommodate, because their compensation is the same whether they “do this” or “do that,” but an unwillingness to accommodate is interpreted not as an excess of zeal but as a predilection to behave in a noncooperative way. Long-term promotion prospects are damaged as a consequence. Defection from the spirit of the agreement in favor of litigiousness is quite perverse if neither immediate nor long-term gains are thereby realized. The combination

of fiat with low-powered incentives is a manifestation of the syndrome condition of economic organization to which I referred earlier (and develop more fully below).

The underlying rationale for forbearance law is twofold: (1) parties to an internal dispute have deep knowledge—both about the circumstances surrounding a dispute as well as the efficiency properties of alternative solutions—that can be communicated to the court only at great cost, and (2) permitting the internal disputes to be appealed to the court would undermine the efficacy and integrity of hierarchy. If fiat were merely advisory, in that internal disputes over net receipts could be pursued in the courts, the firm would be little more than an “inside contracting” system (Williamson, 1985b, pp. 218–22). The application of forbearance doctrine to internal organization means that parties to an internal exchange can work out their differences themselves or appeal unresolved disputes to the hierarchy for a decision. But this exhausts their alternatives. When push comes to shove, “legalistic” arguments fail. Greater reliance on instrumental reasoning and mutual accommodation result. This argument contradicts Alchian and Demsetz’s claim that the firm “has no power of fiat, no authority, no disciplinary action any different in the slightest degree from ordinary market contracting” (1972, p. 777). That is exactly wrong: firms can and do exercise fiat that markets cannot. Prior neglect of contract law differences and their ramifications explain the disparity.

### 1.3. First-Order Economizing

Although the need to get priorities straight is unarguable, first-order economizing—effective adaptation and the elimination of waste—has been neglected. Adaptation is especially crucial. As developed below, it is the central economic problem. But as Frank Knight insisted, the elimination of waste is also important (1941, p. 252).

Relatedly, but independently, Oskar Lange held that “the real danger of socialism is that of the bureaucratization of economic life, and not the impossibility of coping with the problem of allocation of resources” (1938, p. 109). Inasmuch, however, as Lange believed that this argument belonged “in the field of sociology” he concluded that it “must be dispensed with here” (1938, p. 109). Subsequent informed observers of socialism followed this lead, whereupon the problems of bureaucracy were, until recently, given scant attention. Instead, the study of socialism was preoccupied with technical features—marginal cost pricing, activity analysis, and the like—with respect to which a broadly sanguine consensus took shape (Bergson, 1948; Montias, 1976; Koopmans, 1977).

The natural interpretation of the organizational concerns expressed by Knight and Lange—or, at least, the interpretation that I propose here—is that economics was too preoccupied with issues of allocative efficiency, in which marginal analysis was featured, to the neglect of organizational efficiency, in which discrete structural alternatives were brought under scrutiny. Partly that is because the mathematics for dealing with clusters of attributes is only



now beginning to be developed (Topkis, 1978; Milgrom and Roberts, 1990b; Holmstrom and Milgrom, 1991). Even more basic, however, is the propensity to focus exclusively on market mechanisms to the neglect of discrete structural alternatives. The argument, for example, that all systems of honest trade are variants on the reputation-effect mechanisms of markets (Milgrom, North, and Weingast, 1990, p. 16) ignores the possibility that some ways of infusing contractual integrity (e.g., hierarchy) employ altogether different means. Market-favoring predispositions need to be disputed, lest the study of economic organization in all of its forms be needlessly and harmfully truncated.

## 2. Dimensionalizing Governance

What are the key attributes with respect to which governance structures differ? The discriminating alignment hypothesis to which transaction-cost economics owes much of its predictive content holds that transactions, which differ in their attributes, are aligned with governance structures, which differ in their costs and competencies, in a discriminating (mainly, transaction-cost-economizing) way. But whereas the dimensionalization of transactions received early and explicit attention, the dimensionalization of governance structures has been relatively slighted. What are the factors that are responsible for the aforementioned differential costs and competencies?

One of those key differences has been already indicated: market, hybrid, and hierarchy differ in contract law respects. Indeed, were it the case that the very same type of contract law were to be uniformly applied to all forms of governance, important distinctions between these three generic forms would be vitiated. But there is more to governance than contract law. Crucial differences in adaptability and in the use of incentive and control instruments are also germane.

### 2.1. Adaptation as the Central Economic Problem

Hayek insistently argued that “economic problems arise always and only in consequence of change” and that this truth was obscured by those who held that “technological knowledge” is of foremost importance (1945, p. 523). He disputed the latter and urged that “the economic problem of society is mainly one of rapid adaptation in the particular circumstances of time and place” (1945, p. 524). Of special importance to Hayek was the proposition that the price system, as compared with central planning, is an extraordinarily efficient mechanism for communicating information and inducing change (1945, pp. 524–27).

Interestingly, Barnard (1938) also held that the main concern of organization was that of adaptation to changing circumstances, but this concern was with adaptation within internal organization. Confronted with a continuously fluctuating environment, the “survival of an organization depends upon the maintenance of an equilibrium of complex character. . . . [This] calls for read-

justment of processes internal to the organization . . . , [whence] the center of our interest is the processes by which [adaptation] is accomplished” (Barnard, 1938, p. 6).

That is very curious. Both Hayek and Barnard hold that the central problem of economic organization is adaptation. But whereas Hayek locates this adaptive capacity in the market, it was the adaptive capacity of internal organization on which Barnard focused attention. If the “marvel of the market” (Hayek) is matched by the “marvel of internal organization” (Barnard), then wherein does one outperform the other?

The marvel to which Hayek referred had spontaneous origins: “The price system is . . . one of those formations which man has learned to use . . . after he stumbled on it without understanding it” (1945, p. 528). The importance of such spontaneous cooperation notwithstanding, it was Barnard’s experience that intended cooperation was important and undervalued. The latter was defined as “that kind of cooperation among men that is conscious, deliberate, purposeful” (Barnard, 1938, p. 4) and was realized through formal organization, especially hierarchy.

I submit that adaptability is the central problem of economic organization and that both Hayek and Barnard are correct, because they are referring to adaptations of different kinds, both of which are needed in a high-performance system. The adaptations to which Hayek refers are those for which prices serve as sufficient statistics. Changes in the demand or supply of a commodity are reflected in price changes, in response to which “individual participants . . . [are] able to take the right action” (Hayek, 1945, p. 527). I will refer to adaptations of this kind as adaptation (A), where (A) denotes autonomy. This is the neoclassical ideal in which consumers and producers respond independently to parametric price changes so as to maximize their utility and profits, respectively.

That would entirely suffice if all disturbances were of this kind. Some disturbances, however, require coordinated responses, lest the individual parts operate at cross-purposes or otherwise suboptimize. Failures of coordination may arise because autonomous parties read and react to signals differently, even though their purpose is to achieve a timely and compatible combined response. The “nonconvergent expectations” to which Malmgren (1961) referred is an illustration. Although, in principle, convergent expectations could be realized by asking one party to read and interpret the signals for all, the lead party may behave strategically—by distorting information or disclosing it in an incomplete and selective fashion.

More generally, parties that bear a long-term bilateral dependency relation to one another must recognize that incomplete contracts require gapfilling and sometimes get out of alignment. Although it is always in the collective interest of autonomous parties to fill gaps, correct errors, and effect efficient realignments, it is also the case that the distribution of the resulting gains is indeterminate. Self-interested bargaining predictably obtains. Such bargaining is itself costly. The main costs, however, are that transactions are maladapted to the environment during the bargaining interval. Also, the prospect of ex

post bargaining invites *ex ante* prepositioning of an inefficient kind (Grossman and Hart, 1986).

Recourse to a different mechanism is suggested as the needs for coordinated investments and for uncontested (or less contested) coordinated realignments increase in frequency and consequentiality. Adaptations of these coordinated kinds will be referred to as adaptation (C), where (C) denotes cooperation. The conscious, deliberate, and purposeful efforts to craft adaptive internal coordinating mechanisms were those on which Barnard focused. Independent adaptations here would at best realize imperfect realignments and could operate at cross-purposes. Lest the aforementioned costs and delays associated with strategic bargaining be incurred, the relation is reconfigured by supplanting autonomy by hierarchy. The authority relation (*fiat*) has adaptive advantages over autonomy for transactions of a bilaterally (or multilaterally) dependent kind.

## 2.2. Instruments

Vertical and lateral integration are usefully thought of as organization forms of last resort, to be employed when all else fails. That is because markets are a “marvel” in adaptation (A) respects. Given a disturbance for which prices serve as sufficient statistics, individual buyers and suppliers can reposition autonomously. Appropriating, as they do, individual streams of net receipts, each party has a strong incentive to reduce costs and adapt efficiently. What I have referred to as high-powered incentives result when consequences are tightly linked to actions in this way (Williamson, 1988a). Other autonomous traders have neither legitimate claims against the gains nor can they be held accountable for the losses. Accounting systems cannot be manipulated to share gains or subsidize losses.

Matters get more complicated when bilateral dependency intrudes. As discussed above, bilateral dependency introduces an opportunity to realize gains through hierarchy. As compared with the market, the use of formal organization to orchestrate coordinated adaptation to unanticipated disturbances enjoys adaptive advantages as the condition of bilateral dependency progressively builds up. But these adaptation (C) gains come at a cost. Not only can related divisions within the firm make plausible claims that they are causally responsible for the gains (in indeterminate degree), but divisions that report losses can make plausible claims that others are culpable. There are many ways, moreover, in which the headquarters can use the accounting system to effect strategic redistributions (through transfer pricing changes, overhead assignments, inventory conventions, etc.), whatever the preferences of the parties. The upshot is that internal organization degrades incentive intensity, and added bureaucratic costs result (Williamson, 1985b, chap. 6; 1988d).

These three features—adaptability of type A, adaptability of type C, and differential incentive intensity—do not exhaust the important differences between market and hierarchy. Also important are the differential reliance

on administrative controls and, as developed above, the different contract law regimes to which each is subject. Suffice it to observe here that (1) hierarchy is buttressed by the differential efficacy of administrative controls within firms, as compared with between firms, and (2) incentive intensity within firms is sometimes deliberately suppressed. Incentive intensity is not an objective but is merely an instrument. If added incentive intensity gets in the way of bilateral adaptability, then weaker incentive intensity supported by added administrative controls (monitoring and career rewards and penalties) can be optimal.

Markets and hierarchies are polar modes. As indicated at the outset, however, a major purpose of this chapter is to locate hybrid modes—various forms of long-term contracting, reciprocal trading, regulation, franchising, and the like—in relation to these polar modes. Plainly, the neoclassical contract law of hybrid governance differs from both the classical contract law of markets and the forbearance contract law of hierarchies, being more elastic than the former but more legalistic than the latter. The added question is How do hybrids compare with respect to adaptability (types A and C), incentive intensity, and administrative control?

The hybrid mode displays intermediate values in all four features. It preserves ownership autonomy, which elicits strong incentives and encourages adaptation to type A disturbances (those to which one party can respond efficiently without consulting the other). Because there is bilateral dependency, however, long-term contracts are supported by added contractual safeguards and administrative apparatus (information disclosure, dispute-settlement machinery). These facilitate adaptations of type C but come at the cost of incentive attenuation. Concerns for “equity” intrude. Thus the Nevada Power Company–Northwest Trading Company coal contract, whose adaptation mechanics were set out above, begins with the following: “It is the intent of the Parties hereto that this agreement, as a whole and in all of its parts, shall be equitable to both Parties throughout its term.” Such efforts unavoidably dampen incentive-intensity features.

One advantage of hierarchy over the hybrid with respect to bilateral adaptation is that internal contracts can be more incomplete. More importantly, adaptations to consequential disturbances are less costly within firms because (1) proposals to adapt require less documentation, (2) resolving internal disputes by fiat rather than arbitration saves resources and facilitates timely adaptation, (3) information that is deeply impacted can more easily be accessed and more accurately assessed, (4) internal dispute resolution enjoys the support of informal organization (Barnard, 1938; Scott, 1987), and (5) internal organization has access to additional incentive instruments—including especially career reward and joint profit sharing—that promote a team orientation. Furthermore, highly consequential disturbances that would occasion breakdown or costly litigation under the hybrid mode can be accommodated more easily. The advantages of hierarchy over hybrid in adaptation C respects are not, however, realized without cost. Weaker incentive intensity (greater bureaucratic costs) attend the move from hybrid to hierarchy, *ceteris paribus*.

*Table 4.1. Distinguishing Attributes of Market, Hybrid, and Hierarchy Governance Structures<sup>a</sup>*

Attributes	Governance Structure		
	Market	Hybrid	Hierarchy
<b>Instruments</b>			
Incentive intensity	++	+	0
Administrative controls	0	+	++
<b>Performance attributes</b>			
Adaptation (A)	++	+	0
Adaptation (C)	0	+	++
Contract law	++	+	0

<sup>a</sup>++ = strong; + = semi-strong; 0 = weak.

Summarizing, the hybrid mode is characterized by semistrong incentives, an intermediate degree of administrative apparatus, displays semi-strong adaptations of both kinds, and works out of a semi-legalistic contract law regime. As compared with market and hierarchy, which are polar opposites, the hybrid mode is located between the two of these in all five attribute respects. Based on the foregoing, and denoting strong, semi-strong, and weak by ++, +, and 0, respectively, the instruments, adaptive attributes, and contract law features that distinguish markets, hybrids, and hierarchies are shown in Table 4.1.

### 3. Discriminating Alignment

Transaction-cost economics subscribes to Commons' view (1924, 1934) that the transaction is the basic unit of analysis. That important insight takes on operational significance upon identifying the critical dimensions with respect to which transactions differ. As heretofore indicated, these include the frequency with which transactions recur, the uncertainty to which transactions are subject, and the type and degree of asset specificity involved in supplying the good or service in question (Williamson, 1979b). Although all are important, transaction-cost economics attaches special significance to this last (Williamson, 1975, 1979b; Klein, Crawford, and Alchian, 1978; Grossman and Hart, 1986).

Asset specificity has reference to the degree to which an asset can be redeployed to alternative uses and by alternative users without sacrifice of productive value. Asset-specificity distinctions of six kinds have been made: (1) site specificity, as where successive stations are located in a cheek-by-jowl relation to each other so as to economize on inventory and transportation expenses; (2) physical asset specificity, such as specialized dies that are required to produce a component; (3) human-asset specificity that arises in learning by doing; (4) brand name capital; (5) dedicated assets, which are discrete investments in general purpose plant that are made at the behest of a particular

customer; to which (6) temporal specificity, which is akin to technological nonseparability and can be thought of as a type of site specificity in which timely responsiveness by on-site human assets is vital has been added (Masten, Meehan, and Snyder, 1991). Asset specificity, especially in its first five forms, creates bilateral dependency and poses added contracting hazards. It has played a central role in the conceptual and empirical work in transaction-cost economics.

The analysis here focuses entirely on transaction costs: neither the revenue consequences nor the production-cost savings that result from asset specialization are included. Although that simplifies the analysis, note that asset specificity increases the transaction costs of all forms of governance. Such added specificity is warranted only if these added governance costs are more than offset by production-cost savings and/or increased revenues. A full analysis will necessarily make allowance for effects of all three kinds (Riordan and Williamson, 1985). Only a truncated analysis appears here.

### 3.1. Reduced-Form Analysis

The governance-cost expressions set out herein are akin to reduced forms, in that governance costs are expressed as a function of asset specificity and a set of exogenous variables. The structural equations from which these reduced forms are derived are not set out. The key features that are responsible for cost differences among governance structures are nonetheless evident in the matrix version of the models set out below.<sup>1</sup>

Although asset specificity can take a variety of forms, the common consequence is this: a condition of bilateral dependency builds up as asset specificity deepens. The ideal transaction in law and economics—whereby the identities of buyers and sellers is irrelevant—obtains when asset specificity is zero. Identity matters as investments in transaction-specific assets increase, since such specialized assets lose productive value when redeployed to best alternative uses and by best alternative users.

Assume, for simplicity, that asset specificity differences are entirely due to physical or site specificity features. I begin with the situation in which classical market contracting works well: autonomous actors adapt effectively to exogenous disturbances. Internal organization is at a disadvantage for transactions of this kind, since hierarchy incurs added bureaucratic costs to which no added benefits can be ascribed. That, however, changes as bilateral dependency sets in. Disturbances for which coordinated responses are required become more numerous and consequential as investments in asset specificity deepen. The high-powered incentives of markets here impede adaptability, since each party to an autonomous exchange that has gotten out of alignment and for which mutual consent is needed to effect an adjustment will want to

1. Developing the deeper structure that supports the reduced forms—by explicating contractual incompleteness and its consequences in a more microanalytic way and by developing the bureaucratic cost consequences of internal organization more explicitly—is an ambitious but important undertaking.

appropriate as much as possible (ideally, all but epsilon) of the adaptive gains to be realized. When bilaterally dependent parties are unable to respond quickly and easily, because of disagreements and self-interested bargaining, maladaptation costs are incurred. Although the transfer of such transactions from market to hierarchy creates added bureaucratic costs, those costs may be more than offset by the bilateral adaptive gains that result.

Let  $M = M(k; \theta)$  and  $H = H(k; \theta)$  be reduced-form expressions that denotes market and hierarchy governance costs as a function of asset specificity ( $k$ ) and a vector of shift parameters ( $\theta$ ). Assuming that each mode is constrained to choose the same level of asset specificity, the following comparative-cost relations obtain:  $M(0) < H(0)$  and  $M' > H' > 0$ .<sup>2</sup> The first of these two inequalities reflect the fact that the bureaucratic costs of internal organization exceed those of the market because the latter is superior in adaptation (A) respects—which is the only kind that matters if asset specificity is negligible. The intercept for market governance is thus lower than is the intercept for hierarchy. The second inequality reflects the marginal disability of markets as compared with hierarchies in adaptation (C) respects as asset specificity, hence bilateral dependency, becomes more consequential.

As described above, the hybrid mode is located between market and hierarchy with respect to incentives, adaptability, and bureaucratic costs. As compared with the market, the hybrid sacrifices incentives in favor of superior coordination among the parts. As compared with the hierarchy, the hybrid sacrifices cooperativeness in favor of greater incentive intensity. The distribution of branded product from retail outlets by market, hierarchy, and hybrid, where franchising is an example of this last, illustrates the argument.

Forward integration out of manufacturing into distribution would be implied by hierarchy. That would sacrifice incentive intensity but would (better) assure that the parts do not operate at cross-purposes with one another. The market solution would be to sell the good or service outright. Incentive intensity is thereby harnessed, but suboptimization (free riding on promotional efforts, dissipation of the brand name, etc.) may also result. Franchising awards greater autonomy than hierarchy but places franchisees under added rules and surveillance as compared with markets. Costs control and local adaptations are stronger under franchising than hierarchy, and suboptimization is reduced under franchising as compared with the market. The added autonomy (as compared with hierarchy) and the added restraints (as compared with the market) under which franchisees operate nevertheless come at a cost. If, for example, quality assurance is realized by constraining the franchisee to use materials supplied by the franchisor, and if exceptions to that practice are not permitted because of the potential for abuse that would result, then local opportunities to make “apparently” cost-effective procurements will be prohibited. Similarly, the added local autonomy enjoyed by franchisees may get in the way of some global adjustments.

2. A more general optimizing treatment in which the level of asset specificity varies with organization form is set out in Riordan and Williamson, 1985; a shorter version of which appears in Chapter 3.

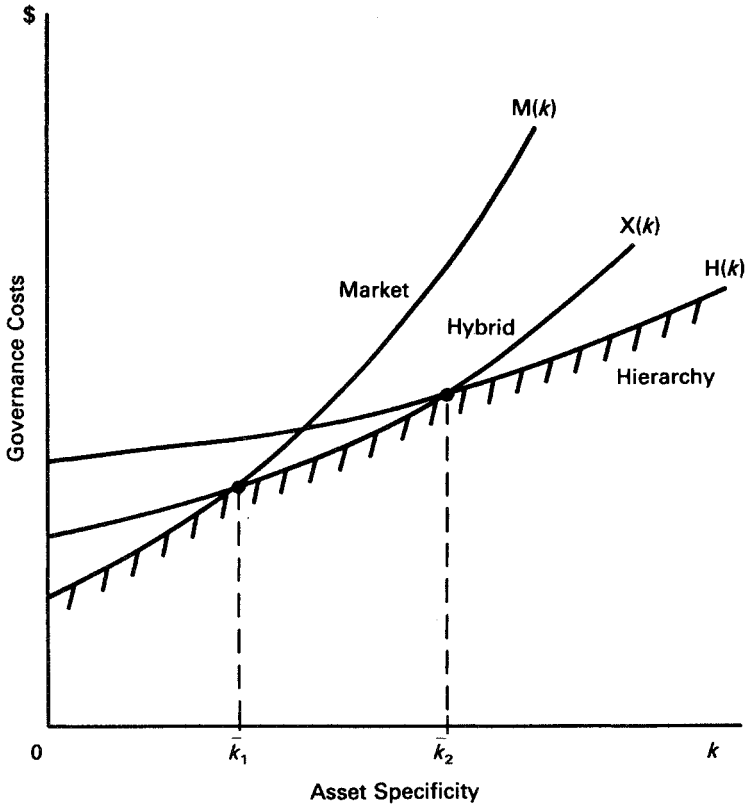


Figure 4.1. Governance costs as a function of asset specificity.

Transactions for which the requisite adaptations to disturbances are neither predominantly autonomous nor bilateral, but require a mixture of each, are candidates to be organized under the hybrid mode. Over some intermediate range of  $k$ , the mixed adaptation (A/C) that hybrids afford could well be superior to the A-favoring or C-favoring adaptations supported by markets and hierarchies, respectively.

Letting  $X = X(k; \theta)$  denote the governance costs of the hybrid mode as a function of asset specificity, the argument is that  $M(0) < X(0) < H(0)$  and that  $M' > X' > H' > 0$ .<sup>3</sup> The relations shown in Figure 4.1 then obtain. Efficient supply implies operating on the envelope, whence, if  $k^*$  is the optimal value of  $k$ , the rule for efficient supply is as follows: I, use markets for  $k^* < \bar{k}_1$ ; II, use hybrids for  $\bar{k}_1 < k^* < \bar{k}_2$ ; and III, use hierarchy for  $k^* > \bar{k}_2$ .

In a very heuristic way, moreover, one can think of moving along one of these generic curves as moving toward more intrusive controls. Thus, consider two forms of franchising, one of which involves less control than the other.

3. This assumes that  $X(0)$  is less than  $H(0)$  to a nontrivial degree, since otherwise the hybrid mode could be dominated throughout by the least-cost choice of either market or hierarchy, which may occur for certain classes of transactions, as discussed below.



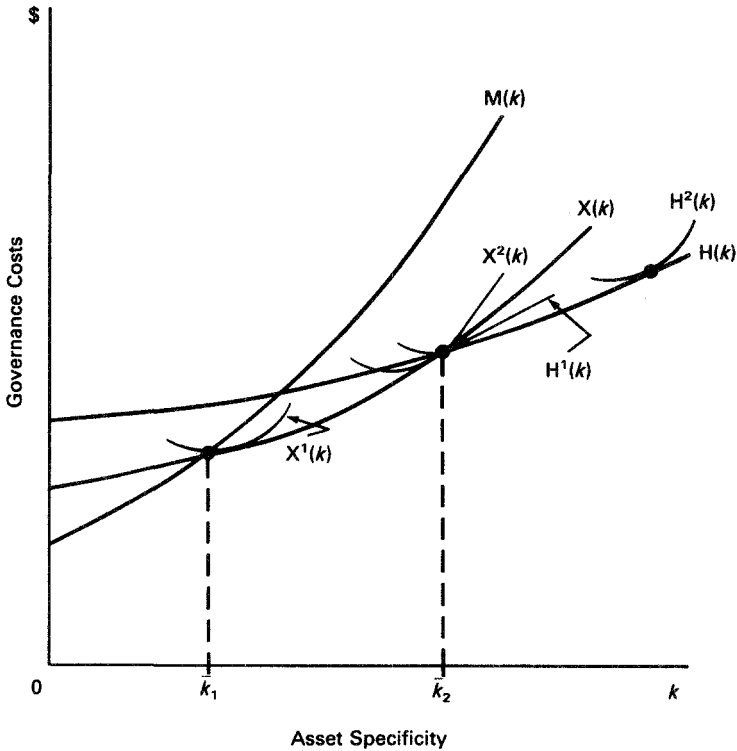


Figure 4.2. Governance differences within discrete structural forms.

If  $X^1(k)$  and  $X^2(k)$  refer to franchising with little and much control, respectively, then  $X^2(k)$  will be located to the right of  $X^1(k)$  in Figure 4.2. Or consider the M-form (multidivisional) and U-form (unitary or functionally organized) corporation. Because the former provides more market-like divisionalization than does the latter, the M-form is given by  $H^1(k)$  and is located closer to  $\bar{k}_2$  in Figure 4.2.

### 3.2. A Stochastic Representation

Suppose that disturbances are distinguished in terms of the type of response—autonomous or bilateral—that is needed to effect an adaptation. Suppose further that the type of adaptation depends on the degree of asset specificity. Let asset specificity be denoted by  $k_j$  and suppose that it can take on any of three values:  $k_1 = 0$  (generic investment),  $k_2 > 0$  (semi-specific investment), or  $k_3 \geq 0$  (highly specific investment). Assume that adjustments to disturbances can be any of four kinds: I, strictly autonomous; II, mainly autonomous; III, mainly coordinated; or IV, strictly coordinated. Let  $p_{ij}$  be the probability that an adaptation of type  $i = \text{I, II, } \dots, \text{IV}$  will be required if asset-specificity condition  $k_j$  ( $j = 1, 2, 3$ ) obtains and let the matrix  $[p_{ij}]$  be given by

		$k_1$	$k_2$	$k_3$
[ $p_{ij}$ ]:	I	1.00	.25	.10
	II	.00	.25	.10
	III	.00	.25	.40
	IV	.00	.25	.40

Note that, the  $k_1$  column excepted, positive probability is associated with every element in the matrix. What added asset specificity does is shift the distribution of required responses in favor of greater cooperativeness.

Assume that each adaptation, if costlessly and successfully implemented, would yield identical expected cost savings. For the reasons given above, however, the efficacy with which different modes adapt to disturbances of different kinds varies. Let  $e_{im}$  be the efficacy with which mode  $m$  ( $m = M, X, H$ ) is able to implement adaptations of type  $i$  ( $i = I, II, \dots, IV$ ) and assume that the matrix  $e_{im}$  is given by

		M	X	H
[ $e_{im}$ ]:	I	1.0	0.9	0.7
	II	0.7	0.9	0.4
	III	0.2	0.5	0.5
	IV	-0.2	0.0	0.5

where 1.0 is the ideal degree of adaptiveness and 0.0 is equivalent (in terms of efficacy) to no adaptation.

The efficacy assumptions embedded in this last matrix warrant remark: (1) Only the entry  $e_{IM}$  has a value of 1.0. This condition—market adaptations to a disturbance for which strictly autonomous adaptation is appropriate—corresponds to the ideal transaction in law and economics (classical market contracting); (2) The efficacy of the market falls off as bilateral dependency builds up, becoming negative (worse than no adaptation at all) for the strictly cooperative case (IV). This last reflects the conflictual nature of market exchange for transactions of the bilaterally dependent kind; (3) The hybrid mode is almost as good as the market for strictly autonomous adaptations, is better than the market in all other adaptation categories, and is as good or better than hierarchy in all categories save that for which strict coordination is indicated; (4) Hierarchy is burdened by bureaucracy and never scores high in efficacy for any category of adaptation.<sup>4</sup> What matters, however, is comparative efficacy. The hierarchy comes into its own (comparatively) where adaptations of a strictly cooperative kind are needed; and (5) The efficacy of hierarchy is lowest for disturbances requiring a mainly autonomous adaptation. As compared with strictly autonomous disturbances, where bureaucratic costs are

4. Hierarchy is able to deal with type I (strictly autonomous) disturbances reasonably well by instructing the operating parts to respond to local disturbances on their own motion and by using the market as an alternate source of supply and/or standard.

held in check by an objective market standard, ready recourse to the market is compromised by the need for some coordination. Because, however, the gains from coordination are not great, efforts to coordinate are problematic. If efforts to adapt autonomously are protested (my costs are greater because you moved without consulting me) while failures to adapt quickly are costly, the hierarchy is caught between the proverbial rock and a hard place.

Let  $C_{jm}$  be the expected maladaptation costs of using mode  $m$  to effect adaptations if asset specificity is of type  $k_j$ . Since inefficacy is given by  $1 - e_{im}$ , the expected maladaptation costs are  $C_{jm} = \sum_i p_{ij} (1 - e_{im})$ . That matrix is given by

		M	X	H
[ $C_{jm}$ ]:	$k_1$	<b>.000</b>	.100	.300
	$k_2$	.575	<b>.425</b>	.475
	$k_3$	.830	.620	<b>.490</b>

The lowest values in each row are realized by matching market, hybrid, and hierarchy with asset specificity conditions  $k_1$ ,  $k_2$ ,  $k_3$ , respectively. These costs are consonant with the reduced-form relations shown in Figure 4.1. Thus if  $\beta \geq 0$  is the irreducible setup costs of economic participation, then the bureaucratic cost intercepts associated with zero asset specificity ( $k_1$ ) for market, hybrid, and hierarchy will be given by  $\beta$  plus .000, .100, and .300, respectively. Also, the relation between the implied slopes associated with each mode in the matrix (expressed as a function of asset specificity) is that  $M' > X' > H'$ , which corresponds exactly to the relations shown in Figure 4.1.

#### 4. Comparative Statics

Transaction-cost economics maintains that (1) transaction-cost economizing is the "main case," which is not to be confused with the only case (Williamson, 1985b, pp. 22–23; 1989c, pp. 137–38), and (2) transaction costs vary with governance structures in the manner described above. Assuming that the institutional environment is unchanging, transactions should be clustered under governance structures as indicated. Variance will be observed, but the main case should be as described.

The purpose of this section is to consider how equilibrium distributions of transactions will change in response to disturbances in the institutional environment. That is a comparative static exercise. Both parts of the new institutional economics—the institutional environment and the institutions of governance—are implicated. The crucial distinctions are these:

The *institutional environment* is the set of fundamental political, social and legal ground rules that establishes the basis for production, exchange and

distribution. Rules governing elections, property rights, and the right of contract are examples. . . .

An *institutional arrangement* is an arrangement between economic units that governs the ways in which these units can cooperate and/or compete. It . . . [can] provide a structure within which its members can cooperate . . . or [it can] provide a mechanism that can effect a change in laws or property rights. (Davis and North, 1971, pp. 6–7)

The way that I propose to join these two is to treat the institutional environment as a set of parameters, changes in which elicit shifts in the comparative costs of governance. An advantage of a three-way setup—market, hybrid, and hierarchy (as compared with just market and hierarchy)—is that much larger parameter changes are required to induce a shift from market to hierarchy (or the reverse) than are required to induce a shift from market to hybrid or from hybrid to hierarchy. Indeed, as developed below, much of the comparative static action turns on differential shifts in the intercept and/or slope of the hybrid mode. The critical predictive action is that which is located in the neighborhood of  $\bar{k}_1$  (M to X) and  $\bar{k}_2$  (X to H) in Figure 4.1. Parameter changes of four kinds are examined: property rights, contract law, reputation effects, and uncertainty.

Among the limitations of the discrete structural approach is that parameter changes need to be introduced in a special way. Rather than investigate the effects of increases (or decreases) in a parameter (a wage rate, a tax, a shift in demand), as is customary with the usual maximizing setup, the comparative governance cost setup needs to characterize parameter changes as improvements (or not). It is furthermore limited by the need for those improvements to be concentrated disproportionately on one generic mode of governance. Those limitations notwithstanding, it is informative to examine comparative static effects.

#### 4.1. Property Rights

What has come to be known as the economics of property rights holds that economic performance is largely determined by the way in which property rights are defined. Ownership of assets is especially pertinent to the definition of property rights, where this “consists of three elements: (a) the right to use the asset [and delimitations that apply thereto] . . . , (b) the right to appropriate returns from the asset . . . , (c) the right to change the asset’s form and/or substance” (Furubotn and Pejovich, 1974, p. 4).

Most discussions of property rights focus on definitional issues. As is generally conceded, property rights can be costly to define and enforce and hence arise only when the expected benefits exceed the expected costs (Demsetz, 1967). That is not my concern here. Rather, I focus on the degree to which property rights, once assigned, have good security features. Security hazards of two types are pertinent: expropriation by the government and expropriation by commerce (rivals, suppliers, customers).

#### 4.1.1. *Governing expropriation*

Issues of “credible commitments” (see Chapter 5) and “security of expectations” (Michelman, 1967) are pertinent to expropriation by the government. If property rights could be efficiently assigned once and for all, so that assignments, once made, would not subsequently be undone—especially strategically undone—governmental expropriation concerns would not arise. Firms and individuals would confidently invest in productive assets without concern that they would thereafter be deprived of their just deserts.

If, however, property rights are subject to occasional reassignment, and if compensation is not paid on each occasion (possibly because it is prohibitively costly), then strategic considerations enter the investment calculus. Wealth will be reallocated (disguised, deflected, consumed) rather than invested in potentially expropriable assets if expropriation is perceived to be a serious hazard. More generally, individuals or groups who either experience or observe expropriation and can reasonably anticipate that they will be similarly disadvantaged in the future have incentives to adapt.

Michelman (1967) focused on cost-effective compensation. He argued that if compensation is costly and if the “demoralization costs” experienced by disadvantaged individuals and interested observers are slight, then compensation is not needed. If, however, demoralization costs can be expected to be great and losses can be easily ascertained, compensation is warranted. Michelman proposed a series of criteria by which to judge how this calculus works out. Suppose that the government is advised of these concerns and “promises” to respect the proposed criteria. Will such promises be believed? This brings us to the problem of credible commitments.

Promises are easy to make, but credible promises are another thing. Kornai’s observation that craftsmen and small shopkeepers fear expropriation in Hungary despite “repeated official declarations that their activity is regarded as a permanent feature of Hungarian socialism” (1986, pp. 1705–6) is pertinent. That “many of them are myopic profit maximizers, not much interested in building up lasting goodwill . . . or by investing in long-lived fixed assets” (1986, p. 1706) is partly explained by the fact that “These individuals or their parents lived through the era of confiscations in the forties” (Kornai, 1986, p. 1705).

But there is more to it than that. Not only is there a history of expropriation, but, as of 1986, the structure of the government had not changed in such a way as to assuredly forestall subsequent expropriations. Official declarations will be more credible only with long experience or if accompanied by a credible (not easily reversible) reorganization of politics. As one Polish entrepreneur recently remarked, “I don’t want expensive machines. If the situation changes, ‘I’ll get stuck with them’” (Newman, 1989, p. A10). Note, in this connection, that the objectivity of law is placed in jeopardy if the law and its enforcement are under the control of a one-party state (Berman, 1983, p. 37). Credibility will be enhanced if a monarch who has made the law “may not make it arbitrarily, and until he has remade it—lawfully—he is bound by it” (Berman,

1983, p. 9). Self-denying ordinances and, even more, inertia that has been crafted into the political process have commitment benefits (North and Weingast, 1989).

That this has not fully registered on Eastern Europe and the Soviet Union is suggested by the following remarks of Mikhail Gorbachev (advising U.S. firms to invest quickly in the Soviet Union rather than wait): “Those [companies] who are with us now have good prospects of participating in our great country . . . [whereas those who wait] will remain observers for years to come—we will see to it” (*International Herald Tribune*, 1990, p. 5, italics added). That the leadership of the Soviet Union “will see to it” that early and late movers will be rewarded and punished, respectively, reflects conventional carrot-and-stick incentive reasoning. What it misses is that ready access to administrative discretion is the source of contractual hazard. The paradox is that fewer degrees of freedom (rules) can have advantages over more (discretion) because added credible commitments can obtain in this way. Effective economic reform thus requires that reneging options be foreclosed if investor confidence is to be realized.

Lack of credible commitment on the part of the government poses hazards for durable, immobile investments of all kinds—specialized and unspecialized alike—in the private sector. If durability and immobility are uncorrelated with asset specificity, then the transaction costs of all forms of private-sector governance increase together as expropriation hazards increase. In that event, the values of  $\bar{k}_1$  and  $\bar{k}_2$  might then change little or not at all. What can be said with assurance is that the government sector will have to bear a larger durable investment burden in a regime in which expropriation risks are perceived to be great. Also, private-sector durable investments will favor assets that can be smuggled or are otherwise mobile—such as general-purpose human assets (skilled machinists, physicians) that can be used productively if emigration is permitted to other countries.

#### 4.1.2. *Leakage*

Not only may property rights be devalued by governments, but the value of specialized knowledge and information may be appropriated and/or dissipated by suppliers, buyers, and rivals. The issues here have recently been addressed by Teece (1986) in conjunction with “weak regimes of appropriability” and are related to earlier discussions by Arrow (1962) regarding property rights in information. If investments in knowledge cannot lawfully be protected or if nominal protection (e.g., a patent) is ineffective, then (1) the ex ante incentives to make such investments are impaired and (2) the incentives to embed such investments in protective ex post governance structures are increased. As Teece (1986) discussed, vertical or lateral integration into related stages of production where the hazards of leakage are greatest is sometimes undertaken for precisely these protective purposes. Trade secret protection is an example.

Interpreted in terms of the comparative governance cost apparatus employed here, weaker appropriability (increased risk of leakage) increases the

cost of hybrid contracting as compared with hierarchy. The market and hybrid curves in Figure 4.1 are both shifted up by increased leakage, so that  $\bar{k}_1$  remains approximately unchanged and the main effects are concentrated at  $\bar{k}_2$ . The value of  $\bar{k}_2$  thus shifts to the left as leakage hazards increase, so that the distribution of transactions favors greater reliance on hierarchy.

## 4.2. Contract Law

Improvements or not in a contract law regime can be judged by how the relevant governance-cost curve shifts. An improvement in excuse doctrine, for example, would shift the cost of hybrid governance down. The idea here is that excuse doctrine can be either too lax or too strict. If too strict, then parties will be reluctant to make specialized investments in support of one another because of the added risk of truly punitive outcomes should unanticipated events materialize and the opposite party insist that the letter of the contract be observed. If too lax, then incentives to think through contracts, choose technologies judiciously, share risks efficiently, and avert adversity will be impaired.

Whether a change in excuse doctrine is an improvement or not depends on the initial conditions and on how these trade-offs play out. Assuming that an improvement is introduced, the effect will be to lower the cost of hybrid contracting—especially at higher values of asset specificity, where a defection from the spirit of the contract is more consequential. The effect of such improvements would be to increase the use of hybrid contracting, especially as compared with hierarchy.

Hadfield has recently examined franchise law and has interpreted the prevailing tendency by the courts to fill in the gaps of an incomplete contract “by according the franchisor unfettered discretion, much as it would enjoy if it [the franchisor] were a vertically integrated corporation” as a mistaken application of forbearance reasoning from hierarchy (where the logic holds) to neoclassical contracting (where the logic fails) (1990, pp. 981–82). Such a failure of franchise law would increase the cost of franchising in relation to forward integration into distribution (Hadfield, 1990, p. 954). This would imply a shift in the value of  $\bar{k}_2$  in Figure 4.1 to the left.

A change in forbearance doctrine would be reflected in the governance cost of hierarchy. Thus, mistaken forbearance doctrine—for example, a willingness by the courts to litigate intrafirm technical disputes—would have the effect of shifting the costs of hierarchical governance up. This would disadvantage hierarchy in relation to hybrid modes of contracting ( $\bar{k}_2$  would shift to the right).

## 4.3. Reputation Effects

One way of interpreting a network is as a nonhierarchical contracting relation in which reputation effects are quickly and accurately communicated. Parties to a transaction to which reputation effects apply can consult not only their own experience but can benefit from the experience of others. To be sure,

the efficacy of reputation effects is easily overstated (Williamson, 1991b), but comparative efficacy is all that concerns us here and changes in comparative efficacy can often be established.

Thus, assume that it is possible to identify a community of traders in which reputation effects work better (or worse). Improved reputation effects attenuate incentives to behave opportunistically in interfirm trade—since the immediate gains from opportunism in a regime where reputation counts must be traded off against future costs. The hazards of opportunism in interfirm trading are greatest for hybrid transactions—especially those in the neighborhood of  $\bar{k}_2$ . Since an improvement in interfirm reputation effects will reduce the cost of hybrid contracting, the value of  $\bar{k}_2$  will shift to the right. Hybrid contracting will therefore increase, in relation to hierarchy, in regimes where interfirm reputation effects are more highly perfected, *ceteris paribus*. Reputation effects are pertinent within firms as well. If internal reputation effects improve, then managerial opportunism will be reduced and the costs of hierarchical governance will fail.

Ethnic communities that display solidarity often enjoy advantages of a hybrid contracting kind. Reputations spread quickly within such communities and added sanctions are available to the membership (Light, 1972). Such ethnic communities will predictably displace nonethnic communities for activities for which interfirm reputation effects are important. Nonethnic communities, to be viable, will resort to market or hierarchy (in a lower or higher  $k$  niche, respectively).

#### 4.4. Uncertainty

Greater uncertainty could take either of two forms. One is that the probability distribution of disturbances remains unchanged but that more numerous disturbances occur. A second is that disturbances become more consequential (due, for example, to an increase in the variance).

One way of interpreting changes of either kind is through the efficacy matrix, above. I conjecture that the effects of more frequent disturbances are especially pertinent for those disturbances for which mainly coordinated or strictly coordinated responses are required. Although the efficacy of all forms of governance may deteriorate in the face of more frequent disturbances, the hybrid mode is arguably the most susceptible. That is because hybrid adaptations cannot be made unilaterally (as with market governance) or by fiat (as with hierarchy) but require mutual consent. Consent, however, takes time. If a hybrid mode is negotiating an adjustment to one disturbance only to be hit by another, failures of adaptation predictably obtain (Ashby, 1960). An increase in market and hierarchy and a decrease in hybrid will thus be associated with an (above threshold) increase in the frequency of disturbances. As shown in Figure 4.3, the hybrid mode could well become nonviable when the frequency of disturbances reaches high levels.<sup>5</sup>

5. The range of asset specificity is from zero (purely generic) to complete (purely firm-specific). The range of frequency is from “low” (a positive lower bound in a nearly unchanging environment) to “very high.”



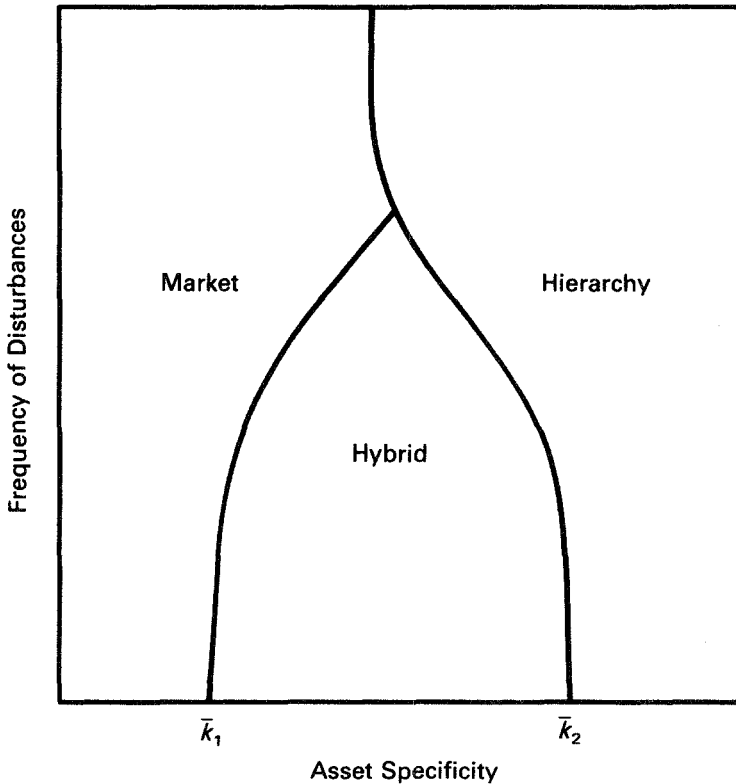


Figure 4.3. Organization form responses to changes in frequency.

If an increase in the variance of the disturbances uniformly increases the benefits to be associated with each successful adaptation, then the effect of increasing the consequentiality of disturbances can again be assessed through the effects on efficacy. Since outliers induce greater defection on the spirit of the agreement for hybrid modes, the efficacy of the hybrid is adversely affected by added variance. Unless similar disabilities can be ascribed to market or hierarchy, the hybrid is disfavored by greater variance, *ceteris paribus*.

## 5. Discussion

The foregoing is concerned with the organization of transactions for mature goods and services and introduces parameter shifts one at a time. Added complications arise when innovation is introduced and when a series of parameter shifts occur together.

### 5.1. Innovation

Some of the added problems posed by innovation take the form of weak property rights. These are discussed above in conjunction with leakage. A

second class of problems that confront innovation is that of timeliness. Non-standard forms of organization, such as parallel R&D (Nelson, 1961) and joint ventures, are sometimes employed because these facilitate timely entry.

Timing can be crucial if a party expects to be a “player” when events are fast-moving or if learning-by-doing is essential. Although transaction-cost economics can relate to some of the pertinent issues, such as those posed by tacit knowledge (Polanyi, 1962) and the limits of imitation (Williamson, 1975, pp. 31–32, 203–7), added apparatus is needed to deal with the full set of issues that arise when responsiveness in real time, rather than equilibrium contracting, is the central concern. Awaiting such developments, the apparatus developed here should not be applied uncritically. For example, joint ventures are sometimes described as hybrids. If, however, joint ventures are temporary forms of organization that support quick responsiveness, and if that is their primary purpose, then both successful and unsuccessful joint ventures will commonly be terminated when contracts expire. Successful joint ventures will be terminated because success will often mean that each of the parties, who chose not to merge but, instead, decided to combine their respective strengths in a selective and timely way, will have learned enough to go it alone. Unsuccessful joint ventures will be terminated because the opportunity to participate will have passed them by. Joint ventures that are designed to give a respite should be distinguished from the types of hybrid modes analyzed here, which are of an equilibrium kind.

The need to distinguish continuing from temporary supply does not, however, mean that transaction-cost economizing principles do not apply to each. To the contrary, although the particulars differ, I would urge that the same general transaction-cost economizing framework has application (Williamson, 1985b). The quasi-firms described by Eccles (1981), for example, can be interpreted as the efficient solution to a particular type of recurrent contracting problem. But the details do matter.

## 5.2. Simultaneous Parameter Shifts

The comparative static analysis set out above treats each generic form of organization as a syndrome of attributes and introduces parameter shifts one at a time. Suppose, instead, that a series of shifts were to occur together. Could these be processed as a sequence of independent changes? If such changes were in fact independent, that is precisely what I would propose. If, however, a related set of changes is made simultaneously, it will not do to treat these independently. If strong interaction effects exist, these must be treated as a cluster.

Relying extensively on the recent work of Aoki (1988, 1990), I interpret the Japanese corporation as follows: (1) three key factors—employment, subcontracting, and banking—are fundamentally responsible for the success of the Japanese firm; (2) the efficacy of each of these rests on distinctive institutional supports; and (3) the three factors bear a complementary relation to each other (see Chapter 12).

The search for key factors and their institutional supports is wholly consistent with the spirit of this chapter. Because employment, subcontracting, and banking changes are linked, however, the American corporation cannot expect to replicate the Japanese corporation by making changes in only one of these practices and not in the others. That is not to say that American firms cannot learn by observing subcontracting practices in Japanese firms. Exact replication of individual practices will be suboptimal, however, if linkages are important.

Similar considerations apply to economic reforms in China and Eastern Europe. If, for example, the efficacy of privatization turns crucially on the manner in which banking is organized and on the security of property rights, then piecemeal proposals that ignore the support institutions are fraught with hazard. The study of viable clusters of organization is a combined law, economics, and organizations undertaking. Although the apparatus in this paper is pertinent, applications to economic reform need to make express provision for contextual differences between alternative forms of capitalism (Hamilton and Biggart, 1988).

## 6. Conclusion

This chapter advances the transaction-cost economics research agenda in the following five respects: (1) the economic problem of society is described as that of adaptation, of which autonomous and coordinated kinds are distinguished; (2) each generic form of governance is shown to rest on a distinctive form of contract law, of which the contract law of forbearance, which applies to internal organization and supports fiat, is especially noteworthy; (3) the hybrid form of organization is not a loose amalgam of market and hierarchy but possesses its own disciplined rationale; (4) more generally, the logic of each generic form of governance—market, hybrid, and hierarchy—is revealed by the dimensionalization and explication of governance herein developed; and (5) the obviously related but hitherto disjunct stages of institutional economics—the institutional environment and the institutions of governance—are joined by interpreting the institutional environment as a locus of shift parameters, changes in which parameters induce shifts in the comparative costs of governance. A large number of refutable implications are derived from the equilibrium and comparative static analyses of governance that result. The growing empirical literature, moreover, is broadly corroborative (for summaries, see Williamson, 1985b, chap. 5; Joskow, 1988; Klein and Shelanski, 1995).

Further developments of conceptual, theoretical, and empirical kinds are needed. Taken together with related developments in information economics, agency theory, and population ecology, there is reason to be optimistic that a “new science of organization” will take shape by the turn of the century (see Chapter 2). Whether that materializes or not, organization theory is being renewed in law, economics, and organizational respects. These are exciting times for interdisciplinary social theory.