Review: The Firm Is Dead; Long Live The Firm a Review of Oliver E. Williamson's The Economic Institutions of Capitalism
Author(s): Armen A. Alchian and Susan Woodward
Review by: Armen A. Alchian and Susan Woodward
Published by: American Economic Association
Stable URL: http://www.jstor.org/stable/2726609

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp
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.
The Firm Is Dead; Long Live the Firm
A Review of Oliver E. Williamson’s
The Economic Institutions of Capitalism*

By ARMEN A. ALCHIAN
University of California, Los Angeles

and

SUSAN WOODWARD
Council of Economic Advisers

The authors thank unnamed referees for improvement in content and exposition. The usual caveats apply.

ONCE UPON A TIME, the organization of cooperative economic activity we call the firm was a black box. Into this box went labor and capital, and out came products. The mechanism was driven by wealth maximization and governed by the laws of returns. Some venture-some economists have wondered what the black box contained, seeing as how market prices could guide all gains from specialization. What more could a firm do?

No one has been more venturesome than Oliver Williamson in his eminently readable and imperially titled book, The Economic Institutions of Capitalism, essentially a compendium, elaboration, and revision of certain of his publications in this decade. It summarizes his many significant insights that gave substance to Coase’s suggestion that firms reduce transactions costs, and presages a research agenda for himself and others. For those seeking a more accessible presentation of the transactions approach, the new book is both a more general

and at the same time more precise exposition of Williamson’s contributions.

Because the book is a summary of Williamson’s contributions to the analysis of the firm, this review will of necessity, and deservedly, become a sort of analytic review of both (a) the prevailing transaction cost interpretation of agreements among owners of cooperatively used resources and (b) some important implications that are being derived from that model.

The emphasis in Williamson’s book is not on capitalism’s “free markets,” but on the constraints that are voluntarily arrived at when transactors are free to impose restrictions upon themselves. The restrictions that interest Williamson are not “anticompetitive” (though some used to be presumed so by some economists) but rather those crafted in competitive markets and that minimize transactions costs both across markets and within firms.

I. The Puzzle

The traditional price-theoretic paradigm in which gains from trade depend on a comparative advantage in production (i.e., by marginal

cost equalization across producers) is distinguished by two features: First, no one relies on someone else for directions about what to do; market prices alone direct production and exchange. Second, production results from cooperative teamwork or cooperative production, leaving no role for contracts or any other constraints (such as rigid prices) on the options of cooperating parties. Yet in a wide array of economic activities people rely on and follow the administrative directions of other people, and both explicit and implicit agreements restrict options. In other words, "firms" or organized and managed "coalitions" exist. Why?

An answer—transactions costs—has been developed in two complementary, but different, directions. One emphasizes the administering, directing, negotiating, and monitoring of the joint productive teamwork in a firm. The other emphasizes assuring the quality or performance of contractual agreements. Both activities prevail across markets and within firms, but to different degrees and with different opportunities for containment. Where these costs are high, market transactions tend to be replaced with internal production and direction and common ownership of more of the jointly used resources so that quality is controlled (managed) and monitored through and during the production process.

A. Bounded Rationality and Opportunism

The first three chapters concern how Williamson views bounded rationality and opportunism as a source of many transactions costs, both across markets and within firms. By "bounded rationality," Williamson means that people have limited information and limited ability to process it. This implies incomplete information about market opportunities, limited ability to predict the future and derive implications from predictions, and limited ability to prespecify responses to future events. People don't know everything and so they make mistakes; moreover each person may know different things.

Opportunism follows from bounded rationality plus self-interest. When a conflict arises between what people want and what they have agreed to do for others, they will act in their own interest insofar as it is costly for others to know their behavior (others face costly information). Opportunism, not merely self-interest, is the original and deadly sin recognized by Williamson.

Opportunism covers more than the propensity for mutually reliant parties to mislead, distort, disguise, obfuscate, or otherwise confuse (p. 47) in order to expropriate wealth from one another. It includes honest disagreements. Even when both parties recognize the genuine goodwill of the other, different but honest perceptions can lead to disputes that are costly to resolve. The point is important because many business arrangements interpreted as responses to potential "dishonest" opportunism are equally appropriate for avoiding costly disputes between honest, ethical people who disagree about what event transpired and what adjustment would have been agreed to initially had the event been anticipated.

B. Transactions, Exchanges, and Contracts

The notion of a "transaction" includes both exchanges and contracts. An exchange is a transfer of property rights to resources that involves no promises or latent future responsibility. In contrast, a contract promises future performance, typically because one party makes an investment, the profitability of which depends on the other party's future behavior. The transactions that are the focus of Williamson's approach are contractual, not just spot exchanges or even a long-lasting series of spot exchanges. In a contract a promise of future performance is exchanged, and investments are made, the value of which becomes dependent on the fulfillment of the other party's promises.

For Williamson, transaction cost is more than the cost of finding other people, inspecting goods, seeking agreeable terms, and writing exchange agreements; it includes, almost to the extent of ignoring the former, the costs incurred in making contracts enforceable by law or by self-enforcement, and extends to the precautions against potential expropriation of the value of investments relying on contractual performance as well as costs of informing and administering terms of contractual relations. These costs are associated with the containment of opportunism.

Williamson defines ex ante costs as the costs of actions and tasks involved in establishing a contract. Ex post costs are those incurred in
subsequently administering, informing, monitoring, and enforcing the contractually promised performance—features that dominate the transaction cost approach. Williamson labels the transition from precontract to postcontract the Fundamental Transformation. Options available in the former stage are lost in the latter; as a result, the value of some resources becomes dependent on particular unique other parties because of loss of significant substitutability by equivalent resources. Other terms used for this dependence are asset specificity, interspecificity, reliance, and Williamson’s previously used term idiosyncratic.

The contract terms that make possible the Fundamental Transformation include both promises of performance and agreements on price. If price (or a price formula) were to be changed or renegotiated whenever either party proposed, no effective commitment of performance would be present. Precommitments to a price by both parties constrain the options and restrict future renegotiability, which is precisely the point of a contract—to protect dependent, reliant resources from malperformance, and to guarantee them a prespecified compensation.

Although this is called a “transaction cost” approach, it involves identifying factors that make spot market transactions less viable, and transactions governed by contracts more viable. It also involves identifying the factors that shape contractural restraints. But we know what the behavioral factors are—bounded rationality and opportunism. Evidently, these can occur in various ways and circumstances, which calls for different institutions for their management. Indeed, we submit that by drawing some distinctions among kinds of opportunism, Williamson’s analysis can be made more powerful.

II. Moral Hazard and Holdup

Paradoxically Williamson’s exposition and, in part the analysis, does not distinguish analytically between the two kinds of opportunism: moral hazard and holdup. Fortunately, it only rarely leads Williamson to errors. But the distinction will help the reader get to many of Williamson’s conclusions. In this section we set out the distinction.

A. Holdup and Specificity

Uncertainty about price or compensation arises when the value of a collection of resources dependent on continued association for their maximum product exceeds their summed market values. Resources dependent upon one another in this way are referred to by Williamson as “specific” or “interspecific,” and those making investments specific to other assets will seek protection against reneging or “holdup” by the other parties. This view gives us insight into contracts, pricing policies, and agreements that constrain markets, and also gives us a theory of the types of assets the firm will own.

Alfred Marshall, in his Principles ([1890] 1936, pp. 453–54 and 626) was the first to identify these elements in the context of what he called “composite quasi-rent.” A quasi-rent is the excess above the return necessary to maintain a resource’s current service flow, which can be the means to recover sunk costs. Composite quasi-rent is that portion of the quasi-rent of resources that depends on continued association with some other specific, currently associated resources. Thus, composite quasi-rent is the amount those other currently associated resources could attempt to expropriate by refusing to pay or serve, that is, by holdup.

Marshall offered the example of a steel mill that locates near a public utility and makes an investment, the profitability of which depends on being able to buy power at some given price. Once the steel mill incurs costs that become sunk, the power company could raise power prices. The steel mill would continue to operate so long as the new marginal cost did not exceed marginal revenue, even though the sunk costs are not being recovered. Marshall recognized the danger of parties with sunk costs relying on those in a position to expropriate composite quasi-rents. But he assumed the threat was resolved by “doing what is right” or by “haggling.” So far as we can ascertain, Marshall did not develop the importance of quasi-rent for understanding the organization of the firm.

Composite quasi-rent has been recently rediscovered by others in discussion of “time-irreversibility,” “asymmetric information with uncertainty,” “bilateral monopoly,” “opportunism,” “self-enforcing contracts,” and “principal-agent relationships” to name a few.
If a resource can leave a team without cost or loss of its value, Williamson would say it is independent or is not team-specific, or is "redeployable." But if the remaining resources would lose by its departure, they are dependent (reliant is the term in legal proceedings) on it, and to them, the departing resource is unique because they cannot replace it with no loss. Resources that are mutually dependent are also mutually unique, and vice versa.

A landowner renting land as a site for a skyscraper is an owner of a resource that is unique, but impotent. The landowner cannot cheaply alter the services of the land as a means of extracting some of the quasi-rent of the building from the building owner. But the owner of the building could refuse to pay all of the promised rent, nevertheless secure in the knowledge that the landowner has no feasible alternative use of the land. The landowner's remedies by law limit the expropriation the building owner could extract, but the expense of the remedies limit the protection afforded the landowner through such remedies. An owner of a unique resource will be more tempted to exploit the situation as the composite quasi-rent grows large and as the unique resource’s flow of services become more controlled (for example, by failing to pay the rent, or to show up for work). The more likely and foreseeable is this temptation, the greater is the likelihood that precautionary contractual terms will be sought. Obviously, dependency motivates a desire for reliable services from the unique resource that is relied upon. Services can be obtained by (1) buying the services, or (2a) owning or (2b) renting the unique resource for self-service. Buying services exposes the purchaser to a holdup threat. Hence, the unique resources are more likely to be controlled by (2a) ownership or (2b) rental for self-use. The choice will depend on the ability of the owner to detect abuse or effects on the resource consequent to the way it is used by a renter. With cheap detection, the unique resource is likely to be rented rather than owned by the dependent party. The two features (a) dependency and (b) monitorability of use are important in determining whether the relied upon services will be obtained by ownership or by rental, and hence the degree of integration of ownership of resources in a firm.

B. Moral Hazard and Plasticity

Moral hazard, a form of opportunism, arises in agreements in which at least one party relies on the behavior of another and information about that behavior is costly. The owner of a firm hires a manager and wants the manager to maximize profits. The manager hires employees and want them to follow directions. An investor lends money to a firm and wants the firm to act in that debt holder’s interest. Because it is costly for the principal to know exactly what the agent did or will do, the agent has an opportunity to bias his actions more in his own interest, to some degree inconsistent with the interests of the principal.

The term moral hazard has been unjustly maligned for its moralistic overtones. In the finance literature, especially, it is often called post-contractual opportunism. We resist. First, as we argue here, the events that can give trouble ex post of the contract are not just those associated with moral hazard but also those of holdup, which we believe are neglected compared to moral hazard in most of the principal-agent literature. Second, the term moral hazard correctly implies that if everyone would simply agree to undertake a given standard of effort and abide by the promise, a more efficient outcome would result. The term suggests that people cannot be counted on to do what they say they are going to do, and that failure manifests itself in prices and in contractual arrangements. It also indicates correctly that among the devices used to control such behavior are moralistic aggression and contempt.

Though moral hazard is involved in many transactions, it does not follow that they are regretted. For example, both the insurance market and the labor market experience moral hazard, but all parties are better off having made the transactions than not. The insured pay the full expected losses plus administration to the insurer, so all costs are compensated. The laborers earn their expected marginal product. The moral hazard loss simply measures what could be gained if magically either the cost of monitoring behavior by the insurer/employer were zero, or if people could be counted on to do as they promise.

Though criticizing Williamson for not analytically distinguishing between two types of op-
portunism, we note that the literatures on holdup and moral hazard are almost entirely unintegrated.

The degree to which resources are vulnerable to morally hazardous exploitation depends on what can be called their "plasticity" and on monitoring costs. We call resources or investment "plastic" to indicate that there is a wide range of discretionary, legitimate decisions within which the user may choose. For example, compare a drug research laboratory with a steel manufacturer. There are fewer options for discretionary behavior in steel manufacturing. The technology is largely determined by the nature of the plant. Absentee owners and debt holders have little cause to worry about the managers' turning the resources of the plant into personal consumption, or increasing the riskiness of prospective outcomes. In contrast, a drug research firm could be working on some mundane project with a modest but sure payoff, or on some long shot with a slight change of a high payoff. Research and development firms are plastic.

We conjecture as further illustrations that enterprises with intellectual research and capital, for example, fashion designers, professional service firms such as engineering, law and architecture, and computer software creation, are especially plastic and susceptible to moral hazard. In contrast, industries with less plasticity are railroads, utility services, airlines, petroleum refining (but not exploration), and other activities involving much in the way of "hard" resources. Interestingly, physical resources requiring large sunk costs, and consequently that are vulnerable to holdup, can be implastic and immune to moral hazard.

Cash is among the most plastic of resources, because it can be hastily exchanged for nearly anything. But this example serves well the point that plasticity must be combined with high monitoring costs to result in opportunities for moral hazard. Cash is plastic, but very easy to track as it is used (once the records establish that it is there). Thus, we expect managers handling large cash balances to be subject to considerable controls and review by principals, but we observe that there are many large financial institutions operated by nonowner managers.

By "plastic" we do not simply mean risky. Oil that has only to be pumped and sold is highly implastic. The optimal rate at which to pump the oil depends on the pattern of prices over time, and there is little in the way of possibilities for exploiting an oil well either by increasing the riskiness of its value or by changing its product into personal consumption. But an oil well is a risky asset if the price of oil fluctuates.

C. Two Expeditions Reunited

The early explorers of transactions costs set out on two very different expeditions. One party, guided by the notion of moral hazard and adverse selection, headed off in the direction of insurance and risk, and ventured successfully into generalized principal-agent conflicts. The implications of informational asymmetries and impactedness for behavior and market viability explained various aspects of insurance markets, the assignment of liability, the use of the firm for teamwork, some firm financing decisions, and conflicts of interest between owners and managers, between inside owners and outside owners, and between debt holders and equity holders. In particular, monitoring costs associated with moral hazard became a basis for explaining two different aspects of firms. The first aspect is the role of "management." Some firm members are managers while others are managed, because the party with comparative advantage in deciding what a particular worker should do is not necessarily the worker himself. This relationship is subject to moral hazard, and the desire to control the costs associated with it explains some aspects of firm organization. Second, monitoring costs will also motivate ownership of assets by their user. If the user of an asset is also its owner, the full consequences of how the asset is used falls on the user. When the owner and user are separate parties, the owner bears the costs of the user's behavior, even if the user is careless. Costly detection of care, moral hazard, is thus a source of asset specificity and will drive ownership of an asset by the user. Once the asset is owned by the firm (used by the owner), informational asymmetries regarding that user's effect on the resource keep it in the firm, that is, make it firm-specific.

The other party studying the firm headed off from industrial organization to look for explanations as to why firms vertically integrate. Chapters 4–10 recount Williamson's earlier ve-
vehicle for exploring beyond that territory: holdup and its prevention. Holdup could explain not only common ownership of assets and aspects of contracts, but also contractual constraints on markets such as posted or otherwise inflexible prices, pay or take, first negotiation, and rights of refusal.

From these two different explorations the question arose: Does the essence of the “firm” lie in teamwork or in the nexus of long-term contracts (i.e., agreements restraining the behavior of transactors)? Williamson writes as if he believes teamwork always involves such contracts. We agree, because we can think of neither significant nor interesting cases where teamwork does not create dependencies calling for contractual restraints. The durability of the team-specific investments, especially accumulated team-specific information involved in teamwork, connects teamwork to contracts.

We observe that in most cooperative production (teamwork), people show up for work at the same place every day. (As a matter of fact, people often show up for play in the same locations, too, though less reliably, but for many of the same reasons.) Team members are more productive working together than working separately, and this differential at least partly depends on knowledge of one another’s personal talents. This specific knowledge has lasting value, and consequently departure of part of the team can threaten the team’s value. This special, lasting knowledge makes the members of the team mutually team-specific. They will want assurance of performance and compensation before they will be willing to make any self-financed investments in the team’s efforts—hence, long-term contracts. Teamwork and long-term contracts seldom appear without the other.

III. The Firm

To help explain Williamson’s conception of the firm, we offer a provisional characterization. The classic, paradigmatic private property firm is a coalition among owners of separately owned resources whose value as a team exceeds the sum of the market values each could get separately. Some of this value derives from the durable and costly specialized knowledge they possess about each other. That value depends on continued association; the departure of any of the mutually dependent resources would diminish the value of the team. The desire to protect this value and secure a share of it will motivate contracts among inputs for the continuing services of the resources of the team.

It follows that the team members who own resources (human as well) whose values depend most heavily on the performance of the team (i.e., are the most team-specific) will be willing to pay the most for the right to control the team. By definitional consequence of being owners of such resources, they own the residual value. They are the ones who are called the owners of the firm, though no one literally owns all the resources used in the firm. The owners, or equity holders, as a precaution against moral hazard and holdup, will be common to all contracts with input owners, will possess the right to sell their contractual status, and will bear the residual value of team-specific resources. The more general, less specific, more substitutable resources (some of which are called employees) will be rented, because their value and reward are independent of the fortunes and behavior of the team. They must be, and are, paid no more or less than their opportunity value.

The differences between the value of the team and the summed values all the resources could command outside the team is a return to the entrepreneur’s investment in the search for a successful team. It is, as such, much like the expectational return to investments in search for oil. Absent any team-specific investment on their own part, the other members will earn just their value elsewhere, unless they are able to collude to extract a successful team-assembling entrepreneur’s quasi-rent and profit.

A. Ownership Integration

One form of protection from opportunism is common ownership of the dependent resources as one bundle, that is, ownership integration, sometimes misleadingly called “vertical” or “horizontal” integration, although these adjectives are unnecessarily restrictive. This is obvious enough when all the firm-specific resources are owned by one person. Problems develop if either (a) those resources are owned in common by several people (as in corporations and partnerships), or (b) the firm-specific re-
sources are not all owned in common (as in joint ventures or cooperatives).

B. Joint and Several Ownership in Common

Chapters 11 and 12 apply Williamson’s schema to the corporation. In corporations, stockholders jointly own the firm-specific unique resources. While predetermined shares of ownership preclude disputes over the division of the resulting value (at least in the idealized corporation), the shareholders may have different beliefs about the best choice of action, when dividends ought to be taken, or how much risk to take on. One way to get around this difficulty is to make the shares transferable. Then changes in beliefs and personal consumption plans can be accommodated through buying and selling, without affecting the operation of the firm itself.

Williamson sees the anonymity of stockholders as something of a problem. He describes stockholders as having “investments that are not associated with particular assets” (p. 305) and as “the only voluntary constituency whose relation with the corporation does not come up for periodic renewal” (p. 304). He argues that the “diffuse character of their investments puts shareholders at an enormous disadvantage” with respect to protection from opportunistic exploitation. “The board of directors thus arises endogenously, as a means by which to safeguard the investments of those who face a diffuse but significant risk of expropriation because the assets in question are numerous and ill-defined and cannot be protected in a well-focused, transaction specific way” (p. 306).

Anonymity is derived from limited liability and lowers the transaction costs associated with transferable shares. If shareholders were liable for debts of the firm, both creditors and shareholders would find it in their interest to investigate the wealth of each shareholder. They would also find it in their interest to inhibit sales by rich shareholders to poor ones, for such transactions would leave the remaining shareholders with enlarged liability and creditors with less security. Limited liability insulates both shareholders and creditors from differences in wealth among shareholders. The shareholders become a true société anonyme. These considerations explain why virtually all corporations with publicly traded shares limit shareholders’ liability, and why closely held, non-traded firms often extend the liability of one or more shareholders through personal loan guarantees.

The flexibility and liquidity that enable diffuse ownership and anonymity of shareholders are surely among the great virtues of the modern corporation. But these desirable qualities come at a cost because when ownership is diffuse, the shareholders must delegate operation of the firm. Delegation requires monitoring. The problem is not an absence of firm-specific capital on the part of any stockholder, but simply the fact that the stockholders are numerous. But the potential for “political market failure” due to numerous stockholders is substantially abated by the opportunity for some shareholders to control blocks of shares.

If shareholders opt for concentrated ownership another cost arises: Because shareholders have limited wealth, substantial ownership of wealth in one firm implies that the shareholder with a large block has a less well diversified portfolio with an inferior risk-return trade-off and in some way must be compensated for this by the “outside” shareholders. Other things being equal, people prefer to diversify. Compensation may be in the form of higher salaries for owner-managers, or profits from trading on insider information (Harold Demsetz and Kenneth Lehn 1985). The monitoring costs of different activities should thus explain why some firms are organized with substantial owners being managers (aligning the incentives of management and owners) and others with none. Personal, firm-specific investments made by managers result in both the desire of managers to own stock to protect their firm-specific investments and for various other forms of protection, such as golden parachutes.

We wish Williamson had seized this opportunity explicitly (rather than implicitly) to refute the myth that firms are owned, controlled, and administered by “capital” rather than “labor.” He could have cited labor-owned firms: law, architecture, accounting, engineering, economic consulting, advertising, restaurants, computer software creators—the list is long. These are firms in which the human resources are firm-specific, and “labor” is the owner. To believe that “capital” is in some sense the “boss” and hires “labor” is to fail to understand
the most basic forces that shape the firm: First, the leader of a team (management) is the member with the comparative advantage in deciding what the team and its members should do, and this manager need not be an owner or even part owner in the firm; second, ownership of the team is the residual claimancy on the most team-specific resources, which may be labor or capital. To start an analysis of firms by assuming the presence of “capital” or that capital hires labor is to beg the question of the basis for the existence of a firm.

C. Financing

The partitioning of income to assets owned in common may also take the form of separate debt and equity claims. In a brief but provocative and important paragraph (p. 307), Williamson suggests that the manner in which resources are financed will depend on the attributes of the resources. In other words, debt and equity financing will vary directly with the degree of firm-specificity and redeployability—a view in sharp contrast to theorems of the irrelevance of capital structure. We argue and here attempt to elaborate on that important contention. But the type of firm-specificity is at least as important as the degree. In particular, the distinction between holdup and moral hazard illuminates the issue considerably. Assets that are firm-specific and vulnerable to holdup will affect firm financing very differently from those that are firm-specific and also vulnerable to moral hazard. Moreover, there will be different effects on the degree of inside ownership, the potential for being publicly traded, and the optimal incentives for managers.

If a firm’s assets are plastic and costly to monitor, moral hazard costs arise with debt (which, incidentally, limits the degree to which the tax advantages of debt can be exploited). Once indebted, the equity holders do not bear the full downside losses on projects. Their incentive is to increase risk taking, because the bondholders will bear part of the risks of downside losses, but the equity holders get all of the gains. Bondholders, aware of this incentive, design contracts between themselves and equity holders to control it. Debt contracts usually aim at controlling two potential forms of moral hazard opportunism. First, they often constrain the size of dividends that equity holders can pay to themselves, limiting the ability of equity holders to carry away some of the assets (in the form of a dividend) that secure the debt. Second, debt contracts constrain the degree to which the riskiness of the assets can be increased. For example, they might restrict sales of some assets and purchase of others. But the contracts are not perfect. Bondholders charge equity holders for the expected uncontrollable losses imposed on them. In choosing debt financing, equity holders are trading off the moral hazard costs of debt against the attraction of creditors who are risk averse as well as the tax advantages.

The theory of debt financing rests on the degree of asset plasticity as an explanation of the debt-equity ratio. Compare the opportunities for debt financing for a drug company versus a public utility. The drug company has a much wider range of legitimate choices than does the public utility. The drug company’s activities are more difficult to monitor. This implies that it will be difficult for the debt holders to write a contract with the equity holders to keep from being exploited. The drug company will find debt expensive. But with the public utility, the assets are cheaper to monitor and assess (and the returns to equity are regulated), so the public utility will find that the moral hazard consequences of debt are low.

It is not the riskiness, but the plasticity of the firm’s assets that drives the cost of debt financing. What matters is the degree to which the equity holders can exploit the bondholders ex post by altering the returns to assets. We predict that firms with more plastic assets will have lower debt/equity ratios than firms with less plastic assets. Our earlier example, the oil well, serves well the point that the issue is not risk. The oil well is very implastic but very risky; we deduce that oil recovery will be a highly debt financed business, but that drilling and exploration will not.

Another view is that debt is a device to prevent holdups and to restrain moral hazard. There are two versions of this story. First, Richard Ippolito (1985) suggests that committing income to debt holders keeps strong unions from expropriating it. Second, Michael Jensen (1986) argues that if income is committed to bondholders, the managers cannot spend it, because they have less discretion over interest payments than
over dividends. The important point in both views is that when a quasi-rent is present, because of large sunk costs or a windfall gain, the quasi-rent is vulnerable to expropriation. Managers of a firm with exceptionally large quasi-rents or profits could invest the proceeds toward self-aggrandizing but unprofitable investments. Or the union could strike for higher wages. If, instead, the firm had committed to large interest and debt repayments, these cash flows would have to be channeled to outsiders.

The debt/equity decision can also be seen as a “flow” response to fluctuating financing needs (rather than as an optimal “stock”) which seeks to avoid the suspicion of opportunism. If a firm’s managers sold equity every time it sought additional financing, and bought stock back when it had cash balances, stockholders would find it difficult to distinguish between a true financing strategy and “insider trading.” How can stockholders be certain that managers are not simply selling stock when on the basis of inside information they think the price is too high, and buying when it is too low? Only transacting stockholders would be harmed by this, but knowing this, investors who valued liquidity would be reluctant to buy. Perhaps this is a clue as to why nearly all transient fluctuations in financing of public traded firms are in debt and retained earnings, and almost none with equity issues.

D. Incomplete Integration of Ownership

In Chapter 12 Williamson explores corporate governance, in particular where integration of ownership of firm-specific resources is incomplete. Resources cannot be classified as purely firm-specific and dependent or not, and an absolutely nonoverlapping distinction cannot be made among (a) stockholders, (b) creditors, and (c) employees. If some stockholders own firm-specific resources not shared in common by all other stockholders, or if some nonstockholders have some firm-specific resources in which ownership is not shared with stockholders, conflicts of interest among the otherwise homogeneous interests will arise.

Employees can be both dependent and dependent upon. They may have agreed to make self-financed investments with firm-specific value; to develop skills or knowledge with firm-specific value; to purchase homes whose values depend on the firm’s success; to accumulate rights to subsequent benefits, like pensions; to receive payment later for earlier underpayment when the employee’s actual productivity was difficult to predict. Or, employees may have high transfer or mobility costs to the next best work. If so, they will demand some protection from employer opportunism. And they may seek representation on the board of directors, or at least some control with respect to certain decisions affecting the probability of fulfillment of their contracts, though this incomplete and disproportionate sharing in equity relative to fixed payments will create divisiveness on other issues. Williamson concludes that whether labor serves on the board of directors depends on whether employees have made firm-specific investments.

Dependence does not typically stop at the boundaries of groups of cooperating people in what is conventionally called a “firm.” Not to be ignored are some customers of the firm’s products. A consumer who buys a product, the future performance of which depends on the firm’s continued activity, will have become an owner of a firm-specific resource—much like Marshall’s steel mill, or like a buyer of computers or automobiles for which future spare parts are valuable. When customer dependence becomes dominant, the firm will tend to be organized as a mutual, and the customers will own the firm. A subcontractor whose resource values depend on the prime contractor is a dependent part of a coalition, possibly one involving strong mutual interdependence. Though the resources of a subcontractor and a prime contractor may be separately owned, mutual dependence creates a coalition with contractual relationships similar to those “within” a conventional “firm.” Owners of such firm-specific, but separately owned, resources would want representation or influence on the board, even though, again, this will create divisiveness and conflicts of interest.

Although firm-specificity can extend beyond the traditional boundaries of the firm, this does not imply that it is efficient for all, even slightly, firm-specific parties to be represented on the board of directors. As more parties are added, the commonality of objectives diminishes and the cost of making decisions goes up; the cost of negotiating decisions must be traded off.
against the costs of a decision unfavorable to some parties. From an ex ante perspective, negotiating for control and representation should protect all parties willing to pay for protection before becoming dependent. If parties with repeated transactions unexpectedly find themselves mutually dependent, they will be without a contract to contain opportunism. They will have no choice but to negotiate, and to appeal to social institutions larger than the firm (for example, the law) to determine appropriate protection. But the practice of putting ex ante disinterested parties on the board is insidious. Advocates of membership on the board of directors by ex post interested, or firm-independent resource owners, are, possibly unwittingly, undermining the viability of the corporate form by making it easier for owners of non-firm-specific assets to expropriate quasi-rents from firm-specific assets, thereby reducing the willingness to invest in assets organized as a corporation.

IV. Credible Commitments

Absent integration of ownership of all the interdependent resources, investments can sometimes be protected from opportunism by credible commitments. Williamson emphasizes the important role of credible commitments, as distinct from credible threats, in the entire transaction cost analysis. Indeed it is the major implication and message of this book. These commitments can take several forms with varying degrees of credibility and effectiveness, several of which Williamson explores. At the simplest level, a unique (relied upon) party can post a hostage or a bond forfeitable upon malperformance. Less commonly recognized forms of credible commitments are often misinterpreted as monopolizing, or competitor-obstructing, devices. Some of the following examples are illustrated in Williamson’s two chapters on credible commitments: reciprocity, take-or-pay, duplicative suppliers, product exchanges, posted prices, inflexible prices, most-favored-customer clauses, block-booking, blind-selling, blind-buying, buy-sell agreements, stock options, patent pools, joint ventures, premium profit streams supported by exclusive territories and resale price maintenance, franchise-specific investments, exhibition clearances, athlete trades among teams, and reserve and waiver clauses. All these can serve as means of creating competitive, economical mutual reliance and self-enforcing contracts.

We digress a bit to elaborate on one. Price stability can restrain opportunistic behavior by a resource owner who otherwise could alter the price of its services to extract composite quasi-rent from a reliant party. The resource owner, say a buyer from a dependent supplier (e.g., the only oil field pipeline gathering system buying from several oil well owners, or a tuna or salmon canner buying from fishermen who serve only that canner) would desire to assure those dependent suppliers that the buyer would not engage in opportunistic alteration of prices to expropriate quasi-rents of canner- or pipeline-dependent investments by suppliers for whom there were no other economically readily available buyers.

By publicly posting a price and holding it constant, a buyer can assure suppliers of no last-minute price opportunism. Such posted prices tend to be unresponsive to transient changes in demand and supply, and are changed only when the underlying demand and supply conditions have remained changed for some time. Otherwise, adjustments of price to alleged momentary shifts in demand or supply could mask extensive opportunistic expropriation of dependent supplier’s quasi-rents. This implies that posted or stable prices would prevail where suppliers are in a position of substantial dependence with respect to a buyer. Casual observation seems to support that implication, for example, in pipelines for gathering oil and gas, for fishing boats selling to a cannery, and in restaurants. Even restaurants that print or post menus daily seldom change the price of their offerings in the middle of the evening. Williamson correctly emphasized that a clearer perception and appreciation of the problems that arise when two parties separately own interdependent resources would help redirect the inhospitable gaze of economists and lawyers away from the restraints-of-trade paradigm to one in which reliability of future performance is the focus (Alchian and Woodward 1987).

Williamson’s Chapter 10 on the organization of work contains a withering examination and critique exposing the emptiness of the so-called
radical economics interpretation of power and hierarchy in business firms. It is, however, a very constructive chapter with focus on the organization of employees and on the means whereby people who become reliant on other people, whether on their personal services or capital equipment, protect their investments from holdup. This allows a double purpose of employee unions: to cartelize employees to restrict competition, but also to protect employees’ human firm-specific investments. Tenure, seniority, company “unions,” company towns, layoffs and shutdowns, rigid wage rates, and golden parachutes are some examples of devices or contractual clauses to protect firm-dependent human capital (see Alchian and Woodward 1987).

V. Hierarchical Governance Patterns

Administration, direction, or management involves a flow of information through a chain of decisions, suggestions, orders, and so on, which Williamson calls the governance hierarchy. It takes on many forms, for example, putting-out, inside and outside contracting, federated groups, peer groups, employer-employee authority, and the unitary, multidivisional, or holding company forms of governance which he calls U, M, and H forms. His attempts (in Chapter 11) to bring the U (unitary), H (holding company), and M (multidivisional profit center) organizational forms into line with his main argument are, in our opinion, novel and plausible, but as yet a less successful portion of his published work. He associates them with the extent of firm-specific, nonredeployable resources and the extent of other safeguards against opportunism.

The stereotype U form has a unitary chief over subordinate functional divisions (production, sales, finance, engineering, etc.), of a single product firm, so Williamson associates the U form (unitary control) with greater viability where there is no major complex integration. Williamson suggests that safeguards against opportunism (reflecting degree of specificity of resources) across firms are provided by integrating and using the M form, rather than the U form. The U (unitary) form suffered from inability to collect and utilize all pertinent information at one central headquarters. The M form separated the firm into separate profit centers to reduce the need for information flows across divisions and to permit a more independent ability to use information where it existed (in the divisions). This restrains opportunism of information and physical services across the divisions. The distinctive elements are (a) feasibility of decentralizing the flow of information while providing incentives to use the information to make better decisions and (b) safeguards against opportunistic use or concealment of information. As enterprises increased in complexity and integrated complementary activities to avoid potential opportunistic behavior by outside suppliers, the more integrated firm (whether vertical or horizontal) became more complex and difficult to administer by a central office.

The incentive to use information more effectively is sharpened if responsibilities are separated into semiautonomous profit centers which, in turn, are overseen by managers who monitor and evaluate division performance (instead of managing and administering the divisions), reward or punish division managers, and allocate investable funds among them. Thus, the argument goes, the M form evolved precisely for the same reason the “firm” evolved: to restrain opportunism among controllers of interdependent resources.

The H form (holding company) also divides the group into profit centers, and limits control by the top officer primarily to the amount of reinvestment of each center’s income. The manager of each subordinate division controls its own reinvestment selections. The top holding group primarily retains dividend-like returns. Williamson seems to suggest that an H form is more likely among firms that have less operational and informational interdependence among the several divisions; however, a clearer identification of the resource, product, or production conditions that make one form rather than another the more viable seems to be the next item on the research agenda.

VI. External Government Controls

Government regulatory agencies can help not only to avoid “monopolistic prices,” but also to prevent opportunistic holdups of the type illustrated by Marshall. Chapter 13 contains Williamson’s critical review of regulatory expe-
experiences in some TV cable franchise bidding for natural monopolies. Though precontract award bidding is competitive, the Fundamental Transformation occurs in passing to the post-award stage. Williamson argues convincingly, at least to us, that post-award regulatory authority to restrain opportunism is not effectively or totally displaceable by bidding for franchises with long-term contractual commitments.

Chapter 14 opens with this statement: “Antitrust enforcement has been massively reshaped in the past twenty years” (p. 365). That reshaping is concisely reviewed with respect to merger policy, nonstandard contracting, and strategic behavior, a reshaping in which Williamson’s analysis has had a part, and which is very briefly evaluated in Chapter 14. That chapter is convincing evidence that media pundits and politicians who believe that the policies of the Antitrust Division of the Department of Justice and the Federal Trade Commission have been altered by the Reagan administration, and who long for a return to the old antitrust policies, are wrong and are bound to be disappointed. Not the current administration, but instead the advancing understanding of transactions issues is responsible because it is now the intellectual apparatus of economists. Regardless of the administration, economists will use the best economic understanding available and that, rather than some administration’s biases, will guide future antitrust actions. To believe otherwise is to insult economists in those agencies and, worse, to believe economic understanding has no effect on government.

VII. Definitions and Boundaries of “Firms”?

A. Future Forays into Terra Incognita

The view of organizations arising from the concept of resource interdependence or firm specificity makes the boundary of the firm fuzzy; a bright line distinguishing “inside” and “outside” is missing. The interpretation of the firm as a nexus of long-term contracts among interspecific resources weakens the “firm” as a useful basic unit of analysis. Though we have used the word firm, we believe a better and more useful concept is a coalition: a set of resource owners bound by contractual relations that depend on the degrees of dependence and uniqueness. But definitions are for the taking; none is standard.

The old notion led to an intrafirm versus an interfirm conception of agreements in which “interfirm” relations were viewed inhospitably and with suspicion. An egregious example of the pitfalls of this line of thinking was the procedures in a 1982 antitrust suit within the National Football League. The disputants forced the judge to “determine” whether the 24 teams in the league were “one firm” or separate firms, presuming a clear distinction could be made. As the foregoing emphasizes, this approach ignores the nature and degree of dependence among the involved resource owners and the reasons for contractual arrangements to control free-riding on investments and to restrain opportunist expropriation of dependent quasi-rents. The court “found” the league to be 24 separate firms, and inferred they were therefore subject to court review as to legality of joint action. Instead the court could have ruled they were one firm, with 24 subordinate divisions.

This does not mean that economists could have provided an analysis for the NFL case, or reliably ascertained the effects of different rulings. So far as we are aware, economists have not adequately analyzed many mysterious arrangements in interteam sports, for example, trades rather than sale, or players’ employment contracts. Moreover, the mysteries are not limited to team sports: trades or exchanges of products among members of an industry (e.g., exchanges of petroleum and its products, electricity, aluminum, gypsum, corrugated cardboard, and new automobiles) are not rare. We suspect the explanation for these arrangements lies in avoiding potential expropriative behavior where spot supplies are small, but we are not sure.

The institutions of capitalism are more than just firms, markets, and relational contracting. Dependence occurs in complex ways and motivates a large variety of precautionary arrangements. Joint ventures, mutuals, social clubs, cooperatives, and families, to suggest a few “capitalist” institutions, are, as we understand them, basically similar contractual arrangements in which (a) the joint venturers will be interreliant but the assets, especially human capital will not be owned in common, and (b)
the joint venturers will be dependent on the services of the venture (e.g., research, sociability, pipeline transportation), and (c) alienability of a member's interest is restricted. Without a joint, restrictive arrangement, one venturer could otherwise hold up the other dependent nonowner user of the services.

The country club is a complex and rich example of a cooperative firm. The customers (members) of the country club own the firm—the club. They are also the producers of the firm's product—sociability. The most important assets of the firm are not the grounds and building, but the members themselves, who are now "owned by the firm." The members are mutually reliant on one another to produce the sociability that makes membership in the club worthwhile. Sale of membership is restricted to prevent entry of "undesirables," those whose sociability is not regarded as sufficient to exchange for reciprocal sociability. Members can sell their membership only back to the club, and new memberships can be admitted only upon consent of the group. Outside ownership of such a firm is not viable. The outside owner could admit, for a high price, new members who would destroy the composite quasi-rent created by the members themselves, or could raise the price to existing members to extract the quasi-rent.

VIII. Social Restraints

That contracts are not sufficiently well enforced by resort to the law is emphasized by Williamson. Unique parties who could expropriate dependent quasi-rents resist the temptation, in part, because "it isn't right." Social opprobrium and the feeling of guilt may operate. Actions regarded as "unconscionable" or "unfair" can result in social ostracism or moralistic aggression. We believe it is important to recognize the forces of ethics, etiquette, and "proper, correct, reasonable, moral, etc." standards of conduct in controlling business relationships. We do not believe contracts are observed (e.g., self-enforcing) only so long as the personal economic costs of contract violation exceed expropriable rents obtainable by violations. People do not always violate contracts whenever their own costs are less than their own gains from violation. Temptations of free-riding or stealing are resisted even when the net gains of free-riding or stealing are great. We don't know enough about how such "moral" forces operate to say more than that they exist and should not be ignored in seeking an understanding of how the economic institutions of capitalism, or any other -ism, evolve and operate.

One can see how morally aggressive, pejorative terms like "gouging" with reference to prices make economic sense. A remote auto repair shop servicing an unlucky traveler whose car's fanbelt has unexpectedly failed might charge far more than the full costs of replacing the belt, in order to extract almost the total value of an emergency repair to the unfortunate driver—the value of the service to a customer in dire straits. Or, imagine an ambulance operator or a doctor charging a price reflecting the value of emergency service to a critically injured person. Not without reason is such behavior condemned. It is wicked and reprobate, and it's inefficient.

If such opportunistic expropriation were expected by travelers, they would travel less, or take expensive safety precautions to avoid expropriation. The avoidance costs would exceed the true cost of providing the emergency service, so that society would incur greater costs if people did not act "responsibly, fairly, conscientiously and ethically." Whatever the emotive language, "decent" behavior saves resources and enables greater welfare. (It is another question how such "responsible, nonopportunistic" behavior is induced in society, or why people "waste" their own scarce resources berating gougers.) This is consistent with the usual explanation for why professions (i.e., sellers of services whose buyers are in a position of trust and substantial dependence) typically promote and enforce professional codes of ethics to protect the clients or principals from "expropriative" unethical tactics.

IX. Evaluation

It is hard to decide whether the title or the content of Williamson's book is the more general. Williamson describes explicitly only some of capitalism's institutions. But the forces Williamson describes surely operate in all systems. Regardless, the analysis in Williamson's book will enable a broader, more profound under-
standing of coalitions, institutions, and contract structure, including the allocations and partitioning of property rights, as well as a variety of "institutions" that lack contracts but establish and maintain behavior of a contractual type. A central message of the book bears repeating in Williamson’s words:

Upon observing that humans have a propensity to behave opportunistically, Machiavelli advised his prince that “a prudent ruler ought not to keep faith when by so doing it would be against his interest, and when the reasons which made him bind himself no longer exist ... [L]egitimate grounds [have never] failed a prince who wished to show colourable excuse for the promise.” But ... preemptive opportunism is ... a very primitive response. ... The more important lesson, for the purposes of studying economic organizations, is this: Transactions that are subject to ex post opportunism will benefit if appropriate safeguards can be devised ex ante. Rather than reply to opportunism in kind the wise prince is one who seeks both to give and to receive “credible” commitments. (p. 48)

The message of Machiavelli is to be reversed: Instead of opportunism, offer and seek credible commitments. It is clear Williamson’s view is that the main purpose served by economic organization is not monopoly, efficient “risk bearing, power or the like but is transaction cost economizing, in no small part by use of credible commitments.

Even economists who have read the original articles will find Williamson’s The Economic Institutions of Capitalism provocative, informative, edifying, and very much worth reading.

REFERENCES

Because Williamson’s book contains an extensive bibliography through 1984, we have merely added some other and more recent citations.


Alchian and Woodward: Williamson’s Institutions of Capitalism 79


