
Vertical Integration, Appropriable Rents, and the Competitive Contracting Process

Author(s): Benjamin Klein, Robert G. Crawford and Armen A. Alchian

Source: *The Journal of Law & Economics*, Oct., 1978, Vol. 21, No. 2 (Oct., 1978), pp. 297-326

Published by: The University of Chicago Press for The Booth School of Business, University of Chicago and The University of Chicago Law School

Stable URL: <https://www.jstor.org/stable/725234>

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/725234?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

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

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



JSTOR

, The University of Chicago Law School and The University of Chicago Press are collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Law & Economics*

VERTICAL INTEGRATION, APPROPRIABLE RENTS, AND THE COMPETITIVE CONTRACTING PROCESS*

BENJAMIN KLEIN

University of California, Los Angeles

ROBERT G. CRAWFORD

Brigham Young University

and

ARMEN A. ALCHIAN

University of California, Los Angeles

MORE than forty years have passed since Coase's fundamental insight that transaction, coordination, and contracting costs must be considered explicitly in explaining the extent of vertical integration.¹ Starting from the truism that profit-maximizing firms will undertake those activities that they find cheaper to administer internally than to purchase in the market, Coase forced economists to begin looking for previously neglected constraints on the trading process that might efficiently lead to an intrafirm rather than an interfirm transaction. This paper attempts to add to this literature by exploring one particular cost of using the market system—the possibility of post-contractual opportunistic behavior.

Opportunistic behavior has been identified and discussed in the modern analysis of the organization of economic activity. Williamson, for example, has referred to effects on the contracting process of “*ex post* small numbers opportunism,”² and Teece has elaborated:

Even when all of the relevant contingencies can be specified in a contract, contracts are still open to serious risks since they are not always honored. The 1970's are replete with examples of the risks associated with relying on contracts . . . [O]pen displays of

* We wish to acknowledge useful comments on previous drafts by Harold Demsetz, Stephen Friedberg, Victor Goldberg, Levis Kochin, Keith Leffler, Lynne Schneider, Earl Thompson, and participants at a seminar at the Center for the Study of American Business at Washington University and at Law and Economics Workshops at UCLA and the University of Chicago. Financial assistance was provided by a grant of the Lilly Endowment Inc. for the study of property rights and by the Foundation for Research in Economics and Education. The authors are solely responsible for the views expressed and for the remaining errors.

¹ R. H. Coase, *The Nature of the Firm*, 4 *Economica* 386 (1937), reprinted in *Readings in Price Theory* 331 (George J. Stigler & Kenneth E. Boulding eds. 1952).

² Oliver E. Williamson, *Markets and Hierarchies: Analysis and Antitrust Implications* 26-30 (1975).

opportunism are not infrequent and very often litigation turns out to be costly and ineffectual.³

The particular circumstance we emphasize as likely to produce a serious threat of this type of reneging on contracts is the presence of appropriable specialized quasi rents. After a specific investment is made and such quasi rents are created, the possibility of opportunistic behavior is very real. Following Coase's framework, this problem can be solved in two possible ways: vertical integration or contracts. The crucial assumption underlying the analysis of this paper is that, as assets become more specific and more appropriable quasi rents are created (and therefore the possible gains from opportunistic behavior increases), the costs of contracting will generally increase more than the costs of vertical integration. Hence, *ceteris paribus*, we are more likely to observe vertical integration.

I. APPROPRIABLE QUASI RENTS OF SPECIALIZED ASSETS

Assume an asset is owned by one individual and rented to another individual. The quasi-rent value of the asset is the excess of its value over its salvage value, that is, its value in its next best *use* to another renter. The potentially appropriable specialized portion of the quasi rent is that portion, if any, in excess of its value to the second highest-valuing *user*. If this seems like a distinction without a difference, consider the following example.

Imagine a printing press owned and operated by party A. Publisher B buys printing services from party A by leasing his press at a contracted rate of \$5,500 per day. The amortized fixed cost of the printing press is \$4,000 per day and it has a current salvageable value if moved elsewhere of \$1,000 (daily rental equivalent). Operating costs are \$1,500 and are paid by the printing-press owner, who prints final printed pages for the publisher. Assume also that a second publisher C is willing to offer at most \$3,500 for daily service. The current quasi rent on the installed machine is \$3,000 ($= \$5,500 - \$1,500 - \$1,000$), the revenue minus operating costs minus salvageable value. However, the daily quasi rent from publisher B relative to use of the machine for publisher C is only \$2,000 ($= \$5,500 - \$3,500$). At \$5,500 revenue daily from publisher B the press owner would break even on his investment. If the publisher were then able to cut his offer for the press from \$5,500 down to almost \$3,500, he would still have the press service available to him. He would be appropriating \$2,000 of the quasi rent from the press owner. The \$2,000 difference between his prior agreed-to daily rental of \$5,500 and the next best revenue available to the press once the machine is purchased and installed is less than the quasi rent and therefore is potentially appropriable. If no second party were available at the

³ David J. Teece, Vertical Integration and Divestiture in the U.S. Oil Industry 31 (1976).

present site, the entire quasi rent would be subject to threat of appropriation by an unscrupulous or opportunistic publisher.

Our primary interest concerns the means whereby this risk can be reduced or avoided. In particular, vertical integration is examined as a means of economizing on the costs of avoiding risks of appropriation of quasi rents in specialized assets by opportunistic individuals. This advantage of joint ownership of such specialized assets, namely, economizing on contracting costs necessary to insure nonopportunistic behavior, must of course be weighed against the costs of administering a broader range of assets within the firm.⁴

An appropriable quasi rent is not a monopoly rent in the usual sense, that is, the increased value of an asset protected from market entry over the value it would have had in an open market. An appropriable quasi rent can occur with no market closure or restrictions placed on rival assets. Once installed, an asset may be so expensive to remove or so specialized to a particular user that if the price paid to the owner were somehow reduced the asset's services to that user would not be reduced. Thus, even if there were free and open competition for entry to the market, the specialization of the installed asset to a particular user (or more accurately the high costs of making it available to others) creates a quasi rent, but no "monopoly" rent. At the other extreme, an asset may be costlessly transferable to some other user at no reduction in value, while at the same time, entry of similar assets is restricted. In this case, monopoly rent would exist, but no quasi rent.

We can use monopoly terminology to refer to the phenomenon we are discussing as long as we recognize that we are not referring to the usual monopoly created by government restrictions on entry or referring to a single supplier or even highly concentrated supply. One of the fundamental premises of this paper is that monopoly power, better labeled "market power," is pervasive. Because of transaction and mobility costs, "market power" will exist in many situations not commonly called monopolies. There may be many potential suppliers of a particular asset to a particular user but once the investment in the asset is made, the asset may be so specialized to a particular user that monopoly or monopsony market power, or both, is created.

A related motive for vertical integration that should not be confused with our main interest is the optimal output and pricing between two successive monopolists or bilateral monopolists (in the sense of marginal revenue less

⁴ Vertical integration does not completely avoid contracting problems. The firm could usefully be thought of as a complex nonmarket contractual network where very similar forces are present. Frank Knight stressed the importance of this more than 50 years ago when he stated: "[T]he internal problems of the corporation, the protection of its various types of members and adherents against each other's predatory propensities, are quite as vital as the external problem of safeguarding the public interests against exploitation by the corporation as a unit." Frank H. Knight, *Risk, Uncertainty, and Profit* 254 (1964).

than price). A distortion arises because each sees a distorted marginal revenue or marginal cost.⁵ While it is true that this successive monopoly distortion can be avoided by vertical integration, the results of the integration could, for that purpose alone, be achieved by a long-term or a more detailed contract based on the true marginal revenue and marginal costs. Integrated ownership will sometimes be utilized to economize on such precontractual bargaining costs. However, we investigate a different reason for joint ownership of vertically related assets—the avoidance of postcontractual opportunistic behavior when specialized assets and appropriable quasi rents are present. One must clearly distinguish the transaction and information costs of reaching an agreement (discovering and heeding true costs and revenues and agreeing upon the division of profits) and the enforcement costs involved in assuring compliance with an agreement, especially one in which specialized assets are involved. It is this latter situation which we here explore as a motivation for intrafirm rather than interfirm transactions.

We maintain that if an asset has a substantial portion of quasi rent which is strongly dependent upon some other particular asset, both assets will tend to be owned by one party. For example, reconsider our printing press example. Knowing that the press would exist and be operated even if its owner got as little as \$1,500, publisher B could seek excuses to renege on his initial contract to get the weekly rental down from \$5,500 to close to \$3,500 (the potential offer from publisher C, the next highest-valuing user at its present site). If publisher B could effectively announce he was not going to pay more than, say, \$4,000 per week, the press owner would seem to be stuck. This unanticipated action would be opportunistic behavior (which by definition refers to unanticipated non-fulfillment of the contract) if the press owner had installed the press at a competitive rental price of \$5,500 anticipating (possibly naively) good faith by the publisher. The publisher, for example, might plead that his newspaper business is depressed and he will be unable to continue unless rental terms are revised.

Alternatively, and maybe more realistically, because the press owner may have bargaining power due to the large losses that he can easily impose on the publisher (if he has no other source of press services quickly available), the press owner might suddenly seek to get a higher rental price than \$5,500 to capture some newly perceived increase in the publisher's profits. He could do this by alleging breakdowns or unusually high maintenance costs. This type of opportunistic behavior is difficult to prove and therefore litigate.

⁵ This matter of successive and bilateral monopoly has long been known and explicated in many places. See, for example, Robert Bork, *Vertical Integration and the Sherman Act: The Legal History of an Economic Misconception*, 22 U. Chi. L. Rev. 157, 196 (1954); and the discussion in Fritz Machlup & Martha Taber, *Bilateral Monopoly, Successive Monopoly, and Vertical Integration*, 27 *Economica* 101 (1960), where the problem is dated back to Cournot's statement in 1838.

As we shall see, the costs of contractually specifying all important elements of quality varies considerably by type of asset. For some assets it may be essentially impossible to effectively specify all elements of quality and therefore vertical integration is more likely. But even for those assets used in situations where all relevant quality dimensions can be unambiguously specified in a contract, the threat of production delay during litigation may be an effective bargaining device. A contract therefore may be clearly enforceable but still subject to postcontractual opportunistic behavior. For example, the threat by the press owner to break its contract by pulling out its press is credible even though illegal and possibly subject to injunctive action. This is because such an action, even in the very short run, can impose substantial costs on the newspaper publisher.⁶

This more subtle form of opportunistic behavior is likely to result in a loss of efficiency and not just a wealth-distribution effect. For example, the publisher may decide, given this possibility, to hold or seek standby facilities otherwise not worthwhile. Even if transactors are risk neutral, the presence of possible opportunistic behavior will entail costs as real resources are devoted to the attempt to improve posttransaction bargaining positions in the event such opportunism occurs. In particular, less specific investments will be made to avoid being "locked in."⁷ In addition, the increased uncertainty of quality and quantity leads to larger optimum inventories and other increased real costs of production.

This attention to appropriable specialized quasi rents is not novel. In addition to Williamson's⁸ pathbreaking work in the area, Goldberg's⁹ percep-

⁶ While newspaper publishers generally own their own presses, book publishers generally do not. One possible reason book publishers are less integrated may be because a book is planned further ahead in time and can economically be released with less haste. Presses located in any area of the United States can be used. No press is specialized to one publisher, in part because speed in publication and distribution to readers are generally far less important for books than newspapers, and therefore appropriable quasi rents are not created. Magazines and other periodicals can be considered somewhere between books and newspapers in terms of the importance of the time factor in distribution. In addition, because magazines are distributed nationally from at most a few plants, printing presses located in many different alternative areas are possible competitors for an existing press used at a particular location. Hence, a press owner has significantly less market power over the publisher of a magazine compared to a newspaper and we find magazines generally printed in nonpublisher-owned plants. (See W. Eric Gustafson, *Periodicals and Books*, in *Made in New York* 178, 190 (Max Hall ed. 1959).) But while a magazine printing press may be a relatively less specific asset compared to a newspaper printing press, appropriable quasi rents may not be trivial (as possibly they are in the case of book printing). The magazine printing contract is therefore unlikely to be of a short-term one-transaction form but will be a long-term arrangement.

⁷ The relevance for private investments in underdeveloped, politically unstable, that is, "opportunistic," countries is painfully obvious. The importance for economic growth of predictable government behavior regarding the definition and enforcement of property rights has frequently been noted.

⁸ Oliver E. Williamson, *The Vertical Integration of Production: Market Failure Consid-*

tive analysis of what he calls the "hold up" problem in the context of government regulation is what we are discussing in a somewhat different context. Goldberg indicates how some government regulation can usefully be considered a means of avoiding or reducing the threat of loss of quasi rent. (Goldberg treats this as the problem of providing protection for the "right to be served.") He also recognizes that this force underlies a host of other contractual and institutional arrangements such as stockpiling, insurance contracts, and vertical integration. Our analysis will similarly suggest a rationale for the existence of particular institutions and the form of governmental intervention or contractual provisions as alternatives to vertical integration in a wide variety of cases.

II. CONTRACTUAL SOLUTIONS

The primary alternative to vertical integration as a solution to the general problem of opportunistic behavior is some form of economically enforceable long-term contract. Clearly a short-term (for example, one transaction, non-repeat sale) contract will not solve the problem. The relevant question then becomes when will vertical integration be observed as a solution and when will the use of the market-contracting process occur. Some economists and lawyers have defined this extremely difficult question away by calling a long-term contract a form of vertical integration.¹⁰ Although there is clearly a continuum here, we will attempt not to blur the distinction between a long-term rental agreement and ownership. We assume the opportunistic behavior we are concentrating on can occur only with the former.¹¹

For example, if opportunism occurs by the owner-lessor of an asset failing to maintain it properly for the user-lessee and hence unexpectedly increasing the effective rental price, legal remedies (proving contract violation) may be very costly. On the other hand, if the user owned the asset, then the employee who failed to maintain the asset properly could merely be fired.¹² If the employee

erations, 61 *Am. Econ. Rev.* 112 (Papers & Proceedings, May 1971); and Oliver E. Williamson, *Markets and Hierarchies: Analysis and Antitrust Implications* (1975).

⁹ Victor P. Goldberg, *Regulation and Administered Contracts*, 7 *Bell J. Econ. & Management Sci.* 426, 439-41 (1976).

¹⁰ See, for example, Friedrich Kessler & Richard H. Stern, *Competition, Contract, and Vertical Integration*, 69 *Yale L.J.* 1 (1959).

¹¹ It is commonly held that users of assets that can be damaged by careless use and for which the damage is not easy to detect immediately are more likely to own rather than rent the assets. However, these efficient maintenance considerations apply to short-term contracts and are irrelevant if the length of the long-term rental contract coincides with the economic life of the asset. Abstracting from tax considerations, the long-term contract remains less than completely equivalent to vertical integration only because of the possibility of postcontractual opportunistic renegeing. These opportunistic possibilities, however, may also exist within the firm; see note 4 *supra*.

¹² We are abstracting from any considerations of a firm's detection costs of determining proper maintenance. Ease of termination also analytically distinguishes between a franchisor-

could still effectively cheat the owner-user of the asset because of his specific ability to maintain the asset, then the problem is that vertical integration of a relevant asset, the employee's human capital, has not occurred. For the moment, however, we will concentrate solely on the question of long-term rental versus ownership of durable physical assets.¹³

Long-term contracts used as alternatives to vertical integration can be assumed to take two forms: (1) an explicitly stated contractual guarantee legally enforced by the government or some other outside institution, or (2) an implicit contractual guarantee enforced by the market mechanism of withdrawing future business if opportunistic behavior occurs. Explicit long-term contracts can, in principle, solve opportunistic problems, but, as suggested already, they are often very costly solutions. They entail costs of specifying possible contingencies and the policing and litigation costs of detecting violations and enforcing the contract in the courts.¹⁴ Contractual provisions specifying compulsory arbitration or more directly imposing costs on the opportunistic party (for example, via bonding) are alternatives often employed to economize on litigation costs and to create flexibility without specifying every possible contingency and quality dimension of the transaction.

Since every contingency cannot be cheaply specified in a contract or even known and because legal redress is expensive, transactors will generally also rely on an implicit type of long-term contract that employs a market rather than legal enforcement mechanism, namely, the imposition of a capital loss by the withdrawal of expected future business. This goodwill market-enforcement mechanism undoubtedly is a major element of the contractual alternative to vertical integration. Macauley provides evidence that relatively informal, legally unenforceable contractual practices predominate in business relations

franchisee arrangement and a vertically integrated arrangement with a profit-sharing manager. If cheating occurs, it is generally cheaper to terminate an employee rather than a franchisee. (The law has been changing recently to make it more difficult to terminate either type of laborer.) But the more limited job-tenure rights of an employee compared to a franchisee reduce his incentive to invest in building up future business, and the firm must trade off the benefits and costs of the alternative arrangements. A profit-sharing manager with an explicit long-term employment contract would essentially be identical to a franchisee.

¹³ The problems involved with renting specific human capital are discussed below.

¹⁴ The recent Westinghouse case dealing with failure to fulfill uranium-supply contracts on grounds of "commercial impossibility" vividly illustrates these enforcement costs. Nearly three years after outright cancellation by Westinghouse of their contractual commitment, the lawsuits have not been adjudicated and those firms that have settled with Westinghouse have accepted substantially less than the original contracts would have entitled them to. A recent article by Paul L. Joskow, *Commercial Impossibility, the Uranium Market, and the Westinghouse Case*, 6 J. Legal Stud. 119 (1977), analyzes the Westinghouse decision to renege on the contract as anticipated risk sharing and therefore, using our definition, would not be opportunistic behavior. However, the publicity surrounding this case and the judicial progress to date are likely to make explicit long-term contracts a less feasible alternative to vertical integration in the situations we are analyzing.

and that reliance on explicit legal sanctions is extremely rare.¹⁵ Instead, business firms are said to generally rely on effective extralegal market sanctions, such as the depreciation of an opportunistic firm's general goodwill because of the anticipated loss of future business, as a means of preventing nonfulfillment of contracts.

One way in which this market mechanism of contract enforcement may operate is by offering to the potential cheater a future "premium," more precisely, a price sufficiently greater than average variable (that is, avoidable) cost to assure a quasi-rent stream that will exceed the potential gain from cheating.¹⁶ The present-discounted value of this future premium stream must be greater than any increase in wealth that could be obtained by the potential cheater if he, in fact, cheated and were terminated. The offer of such a long-term relationship with the potential cheater will eliminate systematic opportunistic behavior.¹⁷

The larger the potential one-time "theft" by cheating (the longer and more costly to detect a violation, enforce the contract, switch suppliers, and so forth) and the shorter the expected continuing business relationship, the higher this premium will be in a nondeceiving equilibrium. This may therefore partially explain both the reliance by firms on long-term implicit contracts with particu-

¹⁵ Stewart Macaulay, *Non-Contractual Relations in Business: A Preliminary Study*, 28 *Am. Soc. Rev.* 55 (Feb. 1963).

¹⁶ The following discussion of the market enforcement mechanism is based upon the analysis of competitive equilibrium under costly quality information developed in Benjamin Klein & Keith Leffler, *The Role of Price in Guaranteeing Quality*, *J. Pol. Econ.* (forthcoming 1979), which formally extends and more completely applies the analysis in Benjamin Klein, *The Competitive Supply of Money*, 6 *J. Money, Credit, & Banking* 423 (1974). It is similar to the analysis presented in Gary S. Becker & George J. Stigler, *Law Enforcement, Malfeasance, and Compensation of Enforcers*, 3 *J. Legal Stud.* 1 (1974), of insuring against malfeasance by an employer. This market-enforcement mechanism is used in Benjamin Klein & Andrew McLaughlin, *Resale Price Maintenance, Exclusive Territories, and Franchise Termination: The Coors Case* (1978) (unpublished manuscript), to explain franchising arrangements and particular contractual provisions such as resale price maintenance, exclusive territories, initial specific investments, and termination clauses.

¹⁷ Formally, this arrangement to guarantee nonopportunistic behavior unravels if there is a last period in the relationship. No matter how high the premium, cheating would occur at the start of the last period. If transactors are aware of this, no transaction relying on trust (that is, the expectation of another subsequent trial) will be made in the penultimate period, because it becomes the last period, and so on. If some large lump-sum, final-period payment such as a pension as part of the market-enforcement scheme, as outlined by Gary S. Becker & George J. Stigler, *supra* note 16, this last-period problem is obvious. One solution to this unrecognized last-period problem is the acceptance of some continuing third party (for example, escrow agents or government enforcers) to prevent renegeing on the implicit contracts against renegeing we are outlining. Alternatively, the potential loss of value of indefinitely long-lived salable brand-name assets can serve as deterrents to cheating even where the contract between two parties has a last period. If one party's reputation for nonopportunistic dealings can be sold and used in later transactions in an infinite-time-horizon economy, the firm that cheats in the "last" period to any one buyer from the firm experiences a capital loss. This may partially explain the existence of conglomerates and their use of identifying (not product-descriptive) brand names.

lar suppliers and the existence of reciprocity agreements among firms. The premium can be paid in seemingly unrelated profitable reciprocal business. The threat of termination of this relationship mutually suppresses opportunistic behavior.¹⁸

The premium stream can be usefully thought of as insurance payments made by the firm to prevent cheating.¹⁹ As long as both parties to the transaction make the same estimate of the potential short-run gain from cheating, the quantity of this assurance that will be demanded and supplied will be such that no opportunistic behavior will be expected to occur.²⁰ If postcontractual renegeing is anticipated to occur, either the correct premium will be paid to optimally prevent it or, if the premium necessary to eliminate renegeing is too costly, the particular transaction will not be made.

We are not implicitly assuming here that contracts are enforced costlessly and cannot be broken, but rather that given our information-cost assumptions, parties to a contract know exactly when and how much a contract will be broken. An unanticipated broken contract, that is, opportunistic behavior, is therefore not possible in this particular equilibrium. In the context of this model, expected wealth maximization will yield some opportunistic behavior only if we introduce a stochastic element. This will alter the informational equilibrium state such that the potential cheater's estimate of the short-run gain from opportunistic behavior may be at times greater than the other firm's estimate. Hence, less than an optimal premium will be paid and opportunistic behavior will occur.

The firms collecting the premium payments necessary to assure fulfillment of

¹⁸ Although it may not always be in one's narrow self-interest to punish the other party in such a reciprocal relationship since termination may impose a cost on both, it may be rational for one to adopt convincingly such a reaction function to optimally prevent cheating. R. L. Trivers, *The Evolution of Reciprocal Altruism*, 46 *Q. Rev. Bio.* 35, 49 (March 1971), discusses similar mechanisms such as "moralistic aggression" which he claims have been genetically selected to protect reciprocating altruists against cheaters. Similarly, throughout the discussion we implicitly assume that cheating individuals can only cheat once and thereafter earn the "competitive" rate of return. They may, however, be forced to earn less than the competitive wage if they are caught cheating, that is, take an extra capital loss (collusively, but rationally) imposed by other members of the group. This may explain why individuals may prefer to deal in business relations with their own group (for example, members of the same church or the same country club) where effective social sanctions can be imposed against opportunistic behavior. Reliance on such reciprocal business relationships and group enforcement mechanisms is more likely where governmental enforcement of contracts is weaker. Nathaniel H. Leff, *Industrial Organization and Entrepreneurship in the Developing Countries: The Economic Groups*, 26 *Econ. Dev. & Cultural Change* 661 (1978), for example, documents the importance of such groups in less-developed countries. Industries supplying illegal products and services would likely be another example.

¹⁹ It is, of course, an insurance scheme that not only pools risks but also alters them.

²⁰ As opposed to the analysis of Michael R. Darby & Edi Karni, *Free Competition and the Optimal Amount of Fraud*, 16 *J. Law & Econ.* 67 (1973), the equilibrium quantity of opportunistic behavior or "fraud" will be zero under our assumptions of symmetrical information.

contractual agreements in a costly information world may appear to be earning equilibrium "profits" although they are in a competitive market. That is, there may be many, possibly identical, firms available to supply the services of nonopportunistic performance of contractual obligations yet the premium will not be competed away if transactors cannot costlessly guarantee contractual performance. The assurance services, by definition, will not be supplied unless the premium is paid and the mere payment of this premium produces the required services.

Any profits are competed away in equilibrium by competitive expenditures on fixed (sunk) assets, such as initial specific investments (for example, a sign) with low or zero salvage value if the firm cheats, necessary to enter and obtain this preferred position of collecting the premium stream.²¹ These fixed (sunk) costs of supplying credibility of future performance are repaid or covered by future sales on which a premium is earned. In equilibrium, the premium stream is then merely a normal rate of return on the "reputation," or "brand-name" capital created by the firm by these initial expenditures. This brand-name capital, the value of which is highly specific to contract fulfillment by the firm, is analytically equivalent to a forfeitable collateral bond put up by the firm which is anticipated to face an opportunity to take advantage of appropriable quasi rents in specialized assets.

While these initial specific investments or collateral bonds are sometimes made as part of the normal (minimum-cost) production process and therefore at small additional cost, transaction costs and risk considerations do make them costly.²² We can generally say that the larger the appropriable specialized quasi

²¹ A more complete analysis of market equilibrium by the use of specific capital in guaranteeing contract enforcement is developed in Benjamin Klein & Keith Leffler, *supra* note 16.

²² An interesting example of the efficient creation of such a specific collateral investment is provided in *In re Tastee-Freeze International*, 82 F.T.C. 1195 (1973). In this case the franchisor required the franchisee to purchase all the equipment to make soft ice cream except the final patented feeder mechanism which they would only rent at the nominal price of one dollar per month. This, we believe, served the function of substantially reducing the salvage value of the equipment upon termination and therefore was part of the enforcement mechanism to prevent cheating (for example, intentionally failing to maintain quality) by franchisees. If the feeder were sold, the equipment plus the feeder would have a substantial resale value and would not serve the purpose of assuring contract compliance. Similarly, if the equipment were rented along with the feeder the franchisee would not experience a capital loss if terminated. Since the assets of the franchisee are contractually made specific, a situation is created where the assets are now appropriable by an opportunistic franchisor. Generally, a franchisor will lose by terminating a franchisee without cause since that will produce poor incentives on the remaining franchisees to maintain quality and will make it more difficult for the franchisor to sell franchises in the future. But what prevents the franchisor from an unanticipated simultaneous termination of all franchisees, especially after growth of a chain is "complete"? This is logically equivalent to the last-period problem discussed at note 17 *supra* and is restrained in part by its effects on the salable value of the brand name of the franchisor. While we do not know of any evidence of such systematic franchisor cheating, an analysis of this problem which merely asserts that franchisees voluntarily sign contracts with knowledge of these short-term termina-

rents (and therefore the larger the potential short-run gain from opportunistic behavior) and the larger the premium payments necessary to prevent contractual reneging, the more costly this implicit contractual solution will be. We can also expect the explicit contract costs to be positively related to the level of appropriable quasi rents since it will pay to use more resources (including legal services) to specify precisely more contingencies when potential opportunities for lucrative contractual reneging exist.

Although implicit and explicit contracting and policing costs are positively related to the extent of appropriable specialized quasi rents, it is reasonable to assume, on the other hand, that any internal coordination or other ownership costs are not systematically related to the extent of the appropriable specialized quasi rent of the physical asset owned. Hence we can reasonably expect the following general empirical regularity to be true: the lower the appropriable specialized quasi rents, the more likely that transactors will rely on a contractual relationship rather than common ownership. And conversely, integration by common or joint ownership is more likely, the higher the appropriable specialized quasi rents of the assets involved.

III. EXAMPLE OF APPROPRIABLE SPECIALIZED QUASI RENT

This section presents examples of specialized quasi rents where the potential for their appropriation serves as an important determinant of economic organization. A series of varied illustrations, some quite obvious and others rather subtle, will make the analysis more transparent and provide suggestive evidence for the relevance of the protection of appropriable quasi rents as an incentive to vertically integrate. It also suggests the direction of more systematic empirical work that obviously is required to assess the significance of this factor relative to other factors in particular cases. Where this force towards integration (that is, the economizing on contracting costs necessary to assure nonopportunistic behavior in the presence of appropriable quasi rents) does not appear to dominate, important insights regarding the determinants of particular contracting costs and contract provisions are thereby provided.²³

tion provisions is certainly incomplete (see, for example, Paul H. Rubin, *The Theory of the Firm and the Structure of the Franchise Contract*, 21 J. Law & Econ. 223 (1978)).

This example and much of this section of the paper is based upon a more complete theoretical and empirical analysis of actual contractual relationships developed for an ongoing study by Benjamin Klein of FTC litigation in the area of vertical-distribution arrangements.

²³ It is important to recognize that not only will contracting and enforcement costs of constraining opportunistic behavior determine the form of the final economic arrangement adopted by the transacting parties, but they will also influence the firm's production function. That is, the level of specific investment and therefore the size of the potentially appropriable quasi rent is not an independent "technological" datum in each of these following cases, but is economically determined in part by transaction costs.

A. *Automobile Manufacturing*

An illustrative example is the ownership by automobile-producing companies of the giant presses used for stamping body parts. The design and engineering specifications of a new automobile, for example Mustang for Ford, create value in Ford auto production. The manufacture of dies for stamping parts in accordance with the above specifications gives a value to these dies specialized to Ford, which implies an appropriable quasi rent in those dies. Therefore, the die owner would not want to be separate from Ford. Since an independent die owner may likely have no comparable demanders other than Ford for its product and to elicit supply requires payment to cover only the small operating costs once the large sunk fixed cost of the specific investment in the dies is made, the incentive for Ford to opportunistically renegotiate a lower price at which it will accept body parts from the independent die owner may be large. Similarly, if there is a large cost to Ford from the production delay of obtaining an alternative supplier of the specific body parts, the independent die owner may be able to capture quasi rents by demanding a revised higher price for the parts. Since the opportunity to lose the specialized quasi rent of assets is a debilitating prospect, neither party would invest in such equipment. Joint ownership of designs and dies removes this incentive to attempt appropriation.²⁴

In this context, it is interesting to study in some detail the vertical merger that occurred in 1926 of General Motors with Fisher Body. The original production process for automobiles consisted of individually constructed open, largely wooden, bodies. By 1919 the production process began to shift towards largely metal closed body construction for which specific stamping machines became important. Therefore in 1919 General Motors entered a ten-year contractual agreement with Fisher Body for the supply of closed auto bodies.²⁵ In order to encourage Fisher Body to make the required specific investment, this contract had an exclusive dealing clause whereby General Motors agreed to buy

²⁴ The argument also applies to die inserts which can be utilized to make slight modifications in original dies. The value of die inserts is largely an appropriable quasi rent, and so they will also be owned jointly with the designs and basic dies. Aside from the engineering design of the car, the engine blocks, the exterior shell (and possibly the crankshafts, camshafts, and gearing), no other part of the automobile would appear to possess specialized appropriable quasi rents and therefore necessarily be made exclusively by the automobile company. The integration of Ford into the manufacture of spark plugs—a part which seems to be easily standardizable among different autos—by their merger with Autolite, therefore must be explained on other grounds. See *Ford Motor Co. v. United States*, 405 U.S. 562 (1972).

²⁵ The manufacturing agreement between General Motors and Fisher Body can be found in the minutes of the Board of Directors of Fisher Body Corporation for November 7, 1919.

In addition to this long-term contract General Motors also purchased a 60% interest in Fisher at this time. However, as demonstrated by future events, the Fisher brothers clearly seem to have maintained complete control of their company in spite of this purchase.

substantially all its closed bodies from Fisher. This exclusive dealing arrangement significantly reduced the possibility of General Motors acting opportunistically by demanding a lower price for the bodies after Fisher made the specific investment in production capacity. Since exclusive dealing contractual conditions are relatively cheap to effectively specify and enforce, General Motor's postcontractual threat to purchase bodies elsewhere was effectively eliminated.

But large opportunities were created by this exclusive dealing clause for Fisher to take advantage of General Motors, namely to demand a monopoly price for the bodies. Therefore, the contract attempted to fix the price which Fisher could charge for the bodies supplied to General Motors. However, contractually setting in advance a "reasonable" price in the face of possible future changes in demand and production conditions is somewhat more difficult to effectively accomplish than merely "fixing" required suppliers. The price was set on a cost plus 17.6 per cent basis (where cost was defined exclusive of interest on invested capital). In addition, the contract included provisions that the price charged General Motors could not be greater than that charged other automobile manufacturers by Fisher for similar bodies nor greater than the average market price of similar bodies produced by companies other than Fisher and also included provisions for compulsory arbitration in the event of any disputes regarding price.

Unfortunately, however, these complex contractual pricing provisions did not work out in practice. The demand conditions facing General Motors and Fisher Body changed dramatically over the next few years. There was a large increase in the demand for automobiles and a significant shift away from open bodies to the closed body styles supplied by Fisher.²⁶ Meanwhile General Motors was very unhappy with the price it was being charged by its now very important supplier, Fisher. General Motors believed the price was too high because of a substantial increase in body output per unit of capital employed. This was an understandable development given the absence of a capital cost pass-through in the original contract.²⁷ In addition, Fisher refused to locate their body plants adjacent to General Motors assembly plants, a move General Motors claimed was necessary for production efficiency (but which required a large very specific and hence possibly ap-

²⁶ By 1924 more than 65% of automobiles produced by General Motors were of the closed body type. See Sixteenth Annual Report of the General Motors Corporation, year ended December 31, 1924.

²⁷ Deposition of Alfred P. Sloan, Jr. in *United States v. DuPont & Co.*, 366 U.S. 316 (1961), from complete set of briefs and trial records in custody of General Motors, 186-90 (April 28, 1952). Also see direct testimony of Alfred P. Sloan, Jr. in *United States v. DuPont & Co.*, vol. 5 trial transcript, 2908-14 (March 17, 1953). (The government was attempting to demonstrate in this case that General Motors vertically integrated in order to get Fisher to purchase its glass requirements from DuPont.)

propriable investment on the part of Fisher).²⁸ By 1924, General Motors had found the Fisher contractual relationship intolerable and began negotiations for purchase of the remaining stock in Fisher Body, culminating in a final merger agreement in 1926.²⁹

B. *Petroleum Industry*

Appropriable quasi rents exist in specialized assets of oil refineries, pipelines, and oil fields. This leads to common ownership to remove the incentive for individuals to attempt to capture the rents of assets owned by someone else.

Suppose several oil wells are located along a separately owned pipeline that leads to a cluster of independently owned refineries with no alternative crude supply at comparable cost. Once all the assets are in place (the wells drilled and the pipeline and refineries constructed) the oil-producing properties and the refineries are specialized to the pipeline. The portion of their value above the value to the best alternative user is an appropriable specialized quasi rent. The extent of the appropriable quasi rent is limited, in part, by the costs of entry to a potential parallel pipeline developer. Since pipelines between particular oil-producing properties and particular refineries are essentially natural monopolies, the existing pipeline owner may have a significant degree of market power.

These specialized producing and refining assets are therefore "hostage" to the pipeline owner. At the "gathering end" of the pipeline, the monopsonist pipeline could and would purchase all its oil at the same well-head price regardless of the distance of the well from the refinery. This price could be as low as the marginal cost of getting oil out of the ground (or its reservation value for future use, if higher) and might not generate a return to the oil-well owner sufficient to recoup the initial investment of exploration and drilling. At the delivery-to-refinery end of the pipeline, the pipeline owner would be able to appropriate the "specialized-to-the-pipeline quasi rents" of the refineries. The pipeline owner could simply raise the price of crude oil at least to the price of alternative sources of supply to each refinery that are specialized to the pipeline. Given the prospects of such action, if the pipeline owner were an independent monopsonist facing the oil explorers and a

²⁸ *Id.* It is obvious that long-term exclusive dealing contracts are necessary if such investments are to be made by nonvertically integrated firms. See *In re* Great Lakes Carbon Corp., 82 F.T.C. 1529 (1973), for an example of the government's failure to understand this. Great Lakes Carbon Corporation built plants highly specific to particular refineries to process petroleum coke (a by-product of the refining process) for these refineries and was prosecuted for requiring long-term exclusive dealing contracts with refineries.

²⁹ *United States v. DuPont & Co.*, vol. 1, defendants trial exhibits numbers GM-32, GM-33, GM-34.

monopolist to the refinery owners, everyone (explorers and refiners) would know in advance their vulnerability to rent extraction. Therefore oil-field owners and refinery owners would, through shared ownership in the pipeline, remove the possibility of subsequent rent extraction.³⁰

The problem would not be completely solved if just the oil field or the refineries (but not both) were commonly owned with the pipeline, since the local monopoly (or monopsony) would persist vis-à-vis the other. Prospectively, one would expect the common ownership to extend to all three stages. If several refineries (or oil fields) were to be served by one pipeline, all the refinery (or oil field) owners would want to jointly own the pipeline. A common practice is a jointly owned company which "owns" the pipeline with the shares by producers and refiners in the pipeline company corresponding roughly to the respective shares of oil to be transported.³¹

³⁰ Our argument is distinct from the traditional argument in the oil-business literature that vertical integration occurs to achieve "assurance" of supplies or of markets in the face of implicitly or explicitly assumed disequilibrium conditions. See, for example, P. H. Frankel, *Integration in the Oil Industry*, 1 J. Indus. Econ. 201 (1953); Melvin G. de Chazeau & Alfred H. Kahn, *Integration and Competition in the Petroleum Industry* 102-04 (1959); and Michael E. Canes, *A Theory of the Vertical Integration of Oil Firms* (Oct. 1976) (unpublished manuscript, Amer. Petroleum Inst.). Jerry G. Green, *Vertical Integration and Assurance of Markets* (Oct. 1974) (Discussion Paper No. 383, Harvard Inst. of Econ. Research), similarly argues more formally that price inflexibility in an intermediate market which causes shortages and overproduction is an incentive for vertical integration.

It is also important to distinguish between this risk-reducing reason for joint ownership (that is, the reduction in the risk of appropriation of user-associated specialized quasi rents) and the possible risk reduction from joint ownership when there is negative correlation of changes in values of nonappropriable generalized quasi rents. Joint ownership of assets whose value fluctuations are negatively correlated so that gains in one are offset by losses in the other is said to provide a form of insurance against total value changes of the resources used in the manufacturing process. These changes are not the result of any postcontractual opportunistic behavior but of general economic forces outside the control of the immediate parties. For example, a refinery and an oil-producing property fluctuate in value in opposite directions if a new oil field is discovered. The price of oil will fall but the price of refined products will not fall until additional refineries can process larger amounts of oil into more refined products at essentially constant production costs. Then, some of the oil-field owner's losses in value of crude oil are gained by his refinery. This reduces the fluctuation in values caused by factors unrelated to the efficiency of oil producing, refining, and distributing abilities.

However, diversification can also be achieved by methods other than vertical integration. One way is for the investor to buy stocks in the separate unintegrated firms—in effect integrating their ownership by joint holding of common stocks. Although individual action may not always be as cheap or effective as action through intermediaries, financial intermediaries are available such as mutual funds rather than direct diversification by integrated firms. One possible reason why negatively correlated assets could be worth more combined in a single firm is the reduction in the probability of bankruptcy and hence the probability of incurring bankruptcy costs (such as legal fees). An integrated firm with negatively correlated assets could increase its debt to equity ratio while keeping the probability of bankruptcy constant and therefore decrease the taxes on equity without any additional risk. This may be one of the gains of many conglomerate mergers.

³¹ Jane Atwood & Paul Kobrin, *Integration and Joint Ventures in Pipelines* (Sept. 1977) (Research Study No. 5, Am. Petroleum Inst.), find an extremely high positive correlation

Consider other inputs in the production process. The oil tanker, for example, is specialized to crude oil transportation. But since it is essentially equivalued by many alternative users, the appropriable quasi rent is near zero. So we would expect oil tankers not to be extensively owned by refiners or producers. Similarly, the assets used for refinery construction are not specialized to any single refiner or refinery and they should also not be commonly owned with the refinery.

Preliminary examination of the development of the American petroleum industry in the nineteenth century reveals numerous examples that appear consistent with the hypothesis that as technological change leads to assets involved in production, transportation, refining, and marketing becoming more specialized to other specific assets, joint ownership became efficient as a means of preventing opportunistic behavior.

For example, Rockefeller recognized the importance of the pending technological change implied by the substitution of highly specific long-distance pipelines for the somewhat more general capital of the railroads as the efficient mode of transporting oil and took advantage of it. First, before long-distance pipelines were clearly economical, Rockefeller used his dominant oil-refining position to obtain a price reduction on oil he shipped by rail and also rebates from the railroads on oil shipped by competitive oil producers. We conjecture that Rockefeller obtained these price reductions by threatening to build a pipeline parallel to the railroad. He was therefore able to extract the appropriable quasi rents of the railroads. This explains why the rebates were solely a function of oil shipped and not related to nonoil products such as agricultural goods. It also explains why the discount and rebate to Rockefeller were often of the same magnitude. The payment should be a function of total demand for transporting oil.

The obvious question is why some small oil producer or even a nonoil-producing firm did not similarly threaten the railroads with building a pipeline early (before it was cheaper than rail transport) and demand a payment as a function of total oil shipped. The answer, we believe, is that only a dominant oil

between a firm's crude production and its share of ownership in the pipeline. On the other hand, natural gas pipelines, although apparently economically similar in terms of potentially appropriable quasi rents, do not appear to be vertically integrated. Rather than joint-ownership arrangements with the gas producers, these pipelines are often independently owned. The difference may be due to more effective FPC (and now the Federal Energy Regulatory Commission) regulation (of the wellhead and citygate gas prices and the implied pipeline tariff) compared to the direct Interstate Commerce Commission regulation of oil pipelines as common carriers. Regulation of oil pipeline tariffs could, for example, be easily evaded by opportunistic decreases in the wellhead prices paid for oil. More complete government regulation of gas prices may effectively prevent opportunistic behavior by the natural gas pipeline owners, and thereby serve as an alternative to vertical integration. (See Victor P. Goldberg, *supra* note 9.) Edmund Kitch informs us that the evidence does indicate a much greater degree of vertical integration of natural gas pipelines in the period before FPC regulation.

producer would have credible bargaining power with the railroads in this situation because only a dominant producer would be able to make such a highly specific investment. If a small producer or nonoil-producing firm made such an investment, it could easily be appropriated by the oil-producing firms, especially with an alternative means of transportation available. It was therefore necessary for Rockefeller to gain a dominant oil-producing and refining position in order to make a credible threat to the railroads. Appropriating the quasi rents of the railroads by discounts and rebates not only effectively metered the demand for oil transportation but also made it easier for Rockefeller to gain a monopolistic position in the industry without being forced to buy out rivals at prices that would completely reflect future-discounted monopoly profits.³²

C. *Specific Human Capital*

The previous analysis has dealt with examples of physical capital. When specific human capital is involved, the opportunism problem is often more complex and, because of laws prohibiting slavery, the solution is generally some form of explicit or implicit contract rather than vertical integration.

For example, consider the following concrete illustration from the agricultural industry. Suppose someone owns a peach orchard. The ripened peaches, ready for harvest, have a market value of about \$400,000. So far costs of \$300,000 have been paid and the remaining harvesting and shipping costs will be \$50,000 (\$5,000 transport and \$45,000 labor), leaving \$50,000 as the competitive return on the owner's capital. Assume the laborers become a union (one party to whom the crop is now specialized) and refuse to pick unless paid \$390,000. That would leave \$5,000 for transport and only \$5,000 for the owner

³² Although our preliminary investigation indicates that control of the transportation system and vertical integration of it with the oil fields and refineries were significant, there were many other factors in Rockefeller's success. For example, the unpredictability of the life of oil fields raised the risks of a substantial investment in an integrated pipeline transportation system from one field. That Rockefeller correctly or luckily surmised that the Bradford field in 1874 would be long-lived was surely a source of his success. Also his skill in discovering consumer-preferred retailing methods, achieving lower-cost refining, and correctly assessing the ability to refine sulphurous Ohio crude undoubtedly were additional factors. See, for example, Ralph W. Hidy & Muriel E. Hidy, *History of Standard Oil Company (New Jersey): Pioneering in Big Business 1882-1911* (1955); 1 & 2 Allan Nevins, *John D. Rockefeller: The Heroic Age of American Enterprise* (1940); and Harold F. Williamson & Arnold R. Daum, *The American Petroleum Industry* (1959).

This oil-pipeline analysis of appropriable specific capital may be applicable in many other situations. It should hold, for example, for ore mines and refineries which are specialized to each other. We predict that copper smelters specialized to a single mine will tend to be jointly owned, as will a cement quarry and its nearby smelter (mill). Railroad spur lines (and the land on which the track runs) from ore mines to smelters should likewise be owned by the mine-smelter owner. In addition, we would expect television program producers in an area with a single transmitter tower to be joint owners of the tower.

of the peach orchard, instead of the \$350,000 necessary to cover incurred costs and the cost of capital. If the union had power to exclude other pickers, it could extract all the appropriable quasi rent of that year's crop specialized to that particular labor union's service. The union would be extracting not just the usual monopoly rents involved in raising wages, but also the short-run appropriable quasi rents of the farmer's specific assets represented by the ripened peaches. This gain to the union is a one-period return because obviously the farmer will not make any additional specific investments in the future if he knows it will be appropriated by the union.

To reduce this risk of appropriation, the farmer may have a large clan family (or neighbors of similar farms) do his picking. Because of diseconomies of scale, however, this "cooperative" solution is not generally the lowest-cost arrangement and some reliance on market contracting will be necessary. The individual farmer, for example, may want the labor union to put up a forfeitable bond to compensate him in the event the union under threat of strike asks for more wages at harvest time. Alternatively, but equivalently, the collateral put up by the union could be the value of the brand-name capital of the union, a value which will depreciate if its leaders engage in opportunistic behavior. The farmer would then make a continuing brand-name payment to the union (similar to the premium payment noted above) for this collateral.³³

The market value of the union's reputation for reliability of contract observance is the present-discounted value of these brand-name payments which will be greater than any short-run opportunistic gain to the union leaders that could be obtained by threats at harvest time. These payments which increase the cost to the union of opportunistic behavior would be substantial for a perishable product with a large appropriable quasi rent. It is therefore obvious why producers of highly perishable crops are so antagonistic to unionization of field labor. They would be especially hostile to unions without established reputations regarding fulfillment of contract and with politically motivated (and possibly myopic) leaders.³⁴

³³ If the premium is a payment to the union per unit time, then the arrangement is identical to a collateral-bond arrangement where the union collects the interest on the bond as long as no opportunistic behavior occurs. Because of possible legal difficulties of enforcing such an arrangement, however, the premium may be reflected in the price (that is, a higher wage).

³⁴ It is interesting to note in this context that California grape farmers preferred the established Teamsters Union to the new, untried, and apparently more politically motivated field-workers union organized by Cesar Chavez.

Since unions are not "owned," union leaders will not have the proper incentive to maximize the union's value; they will tend more to maximize returns during their tenure. If, however, union leadership (ownership) were salable, the leaders would have the optimal incentive to invest in and conserve the union's brand-name capital. They therefore would not engage in opportunistic actions that may increase current revenue while decreasing the market value of the union. "Idealistic" union leaders that do not behave as if they own the union may, in fact, produce less wealth-maximizing action than would "corrupt" leaders, who act as if they person-

In addition to implicit (brand-name) contracts, opportunistic union behavior may be prevented by use of explicit contracts, often with some outside arbitration as an element of the contract-enforcement mechanism. Although it is difficult for an outsider to distinguish between opportunistic behavior and good-faith modifications of contract, impartial arbitration procedures may reduce the necessity of explicitly specifying possible contingencies and thereby reduce the rigidity of the explicit long-term contract.³⁵

When the problem is reversed and quasi rents of firm-specific human capital of employees may be opportunistically appropriated by the firm, implicit and explicit long-term contracts are also used to prevent such behavior. Because of economies of scale in monitoring and enforcing such contracts, unions may arise as a contract cost-reducing institution for employees with investments in specific human capital.³⁶

In addition to narrow contract-monitoring economies of scale, a union creates a continuing long-term employment relationship that eliminates the last-period (or transient employee) contract-enforcement problem and also creates bargaining power (a credible strike threat) to more cheaply punish a firm that violates the contract. Even when the specific human-capital investment is made by the firm, a union of employees may similarly reduce the contract-enforcement costs of preventing individual-worker opportunism. There are likely to be economies of scale in supply credibility of contract fulfillment, including the long-term continuing relationship aspect of a union. The existence of a union not only makes it more costly for a firm to cheat an individual worker in his last period but also makes it more costly for an individual worker in his last period to cheat the firm, because the union has the incentive (for example, withholding pension rights) to prevent such

ally own the union. Alternatively, the current members of the union may have control, not in the sense of having directly salable shares, but in the sense that the valuable union asset can be transferred to their children or relatives. If government regulations force union members to give away these rights to future rents (for example, by forcing them to admit minorities and eliminate nepotism), we can expect them to intentionally depreciate or not create the reputation capital of the union by opportunistic strikes. See Benjamin Klein, *supra* note 16, where similar problems with regard to the supply of money by nonprivately owned, nonwealth-maximizing firms are discussed.

³⁵ An interesting legal case in this area is *Publishers' Ass'n v. Newspaper & Mail Del. Union*, 114 N.Y.S.2d, 401 (1952). The union authorized and sanctioned a strike against the New York Daily News although the collective bargaining agreement had "no-strike" and arbitration clauses. The Daily News took the union to arbitration, and the arbitrator found actual damages of \$2,000 and punitive damages of \$5,000 if the union again violated the contract. (The court, however, overturned the punitive damages for technical reasons.) See David E. Feller, *A General Theory of the Collective Bargaining Agreement*, 61 Calif. L. Rev. 663 (1973), for a discussion of the flexibility obtained with arbitration provisions in labor contracts.

³⁶ We should explicitly note that we are not considering unions as cartelizing devices, the usually analyzed motivation for their existence. This force is obviously present in many cases (for example, interstate trucking) but is distinct from our analysis.

an externality on the continuing workers. Therefore unions are more likely to exist when the opportunistic cheating problem is greater, namely, when there is more specific human capital present.³⁷

The first Becker analysis of the specific human-capital problem³⁸ ignored opportunistic bargaining difficulties and implicitly assumed arbitrary contracting costs in particular situations to determine a solution. Becker initially assumed that the firm would cheat the employee if the employee made the specific investment. He then argued that the only reason the firm would not make the entire specific investment is because the quit rate of employees, which is a negative function of wages, would then be greater than optimal. Becker did not consider the completely reciprocal nature of the possibilities for cheating. The opportunistic behavior we are emphasizing suggests the possibility of the employee threatening to quit after the firm makes the specific investment unless the wage rate is readjusted upward. Becker's solution of a sharing of the costs and benefits of the specific investment via an initial lump-sum payment by the employee and a later higher-than-market wage does not eliminate the bilateral opportunistic bargaining problem because the employer may later decrease the wage back to the competitive level (or the employee may demand a higher wage to appropriate the partial specific investment by the employer). If it is assumed that employers will not cheat or break contracts in this way, then the efficient solution would be to merely have the employee make the entire specific investment (and therefore have the optimal quit rate) because the employer can costlessly "guarantee" (by assumption) a higher wage reflecting the increased productivity of the firm. But, more generally, to obtain an equilibrium solution to the problem, the costs of creating credibility of contract fulfillment and the costs of enforcing contracts must be explicitly considered.

One of the costs of using an explicit contract which relies on governmental or other outside arbitration for enforcement—rather than on an implicit contract which relies on depreciation of the value of a firm's brand-name (that is, the loss of future premium payments)—is the likely increase in rigidity. For example, the difficulty of specifying all contingencies in labor contracts and of adjusting to unanticipated conditions is likely to lead to wage rigidity. Because contractual changes tend to create suspicion regarding the purpose of the contract alteration and, in particular, raise the ques-

³⁷ When allowing for this "reverse" effect of employee-specific capital, and therefore higher wages, on the formation of unions, the usual positive effect of unions on wages appears to vanish. See, for example, O. Ashenfelter & G. Johnson, *Unionism, Relative Wages, and Labor Quality in U.S. Manufacturing Industries*, 13 *Int'l Econ. Rev.* 488 (Oct. 1972); and Peter Schmidt & Robert P. Strauss, *The Effect of Unions on Earnings and Earnings on Unions: A Mixed Logit Approach*, 17 *Int'l Econ. Rev.* 204 (Feb. 1976).

³⁸ Gary S. Becker, *Human Capital* 18-29 (1964).

tion of whether a firm is using the changed conditions as an opportunity to seize some of the specific quasi rents, long-term labor contracts may consist of rigid wages and layoff provisions. If in the face of declining demand, a firm must keep wages fixed and lay off workers rather than merely reduce wages, the incentive for it to opportunistically claim a false reduction in demand is substantially reduced.³⁹

The fear of opportunistic behavior leads to price (and often also output) rigidity in all kinds of long-term explicit contracts where specific capital is present. This, in turn, leads to the creation of institutions to encourage increased flexibility in the face of changing market conditions. For example, the prime-rate convention, an announced benchmark in terms of which interest rates of corporate bank loans are stated, may be partially rationalized as a cheap means by which the bank can convey information to borrowers that the bank is not opportunistically raising interest fees to a particular customer. A corporate client who has made a specific investment in the supply of information to the bank regarding its credit worthiness (including its financial record of transactions with the bank) creates some appropriable quasi rents. However, when the price of the loan is stated as, say, prime plus one per cent, unless the bank decides to cheat all customers simultaneously and thereby limit new business, an individual customer can clearly distinguish between general market movements in interest rates and any changes the bank decides to make in the particular customer's credit rating. "Price protection" clauses in contracts, where a price decrease to any customer is guaranteed to be given to all customers, may be explained on similar grounds.

These information-cost-reducing institutions, including the use of impartial arbitrators, are highly imperfect. Therefore contracts involving specific assets, even where a price is not explicitly fixed long term, will consequently involve some price rigidity. The macroeconomic implications of this observation (for example, the employment effects of aggregate nominal demand shocks) are obvious.⁴⁰ But the interaction of macroeconomic considerations and industrial

³⁹ This argument is distinct from the recent argument for the existence of rigid long-term implicit labor contracts as a means of bearing risk. See, for example, D. F. Gordon, *A Neo-Classical Theory of Keynesian Unemployment*, 12 *Econ. Inquiry* 431 (Dec. 1974); and Costas Azariadis, *Implicit Contracts and Underemployment Equilibria*, 83 *J. Pol. Econ.* 1183 (1975). We should also note that although Masamori Hashimoto, *Wage Reduction, Unemployment, and Specific Human Capital*, 13 *Econ. Inquiry* 485 (Dec. 1975), has correctly argued that cyclically flexible wages are more likely when specific human capital is present because both workers and employers will want to minimize the likelihood of job separation and thereby protect future returns on the specific human-capital investment, he ignores the contrary effect of increased specific human capital increasing the potential for opportunistic cheating and therefore increasing wage rigidity. The net theoretical effect is indeterminate. One possible reason that high-ranking corporate executives with a great deal of specific human capital appear to have highly flexible wages is because of the large amount of information about the firm they possess and therefore the shorter lag in detecting opportunism.

⁴⁰ The recent "rational-expectations" approach to business cycles, which relies on consumer

organization may not be that obvious. In particular, an increase in the variance of price-level movements, which increases the expected costs to both parties of price rigidity and thereby increases the acceptable degree of price flexibility, also makes it easier for a firm to cheat by opportunistically raising its price. Increased price uncertainty is therefore likely to lead to increased vertical integration.

Where more trust is present and implicit rather than explicit contracts are used, contract prices including wages are likely to be more flexible. If the variance of the price level increases—which makes it more difficult to detect opportunistic behavior and therefore the short-run gains from such cheating—the equilibrium implicit contract will imply a larger premium stream. The interesting question is what are the economic determinants of the implicit relative to explicit contracting costs which will in turn determine the degree of price flexibility.

One determinant of implicit contracting costs is the anticipated growth of demand for the firm's product. The more rapidly demand is expected to grow, the more likely a firm will rely on an implicit contract with its customers. Creating trust is cheaper for firms facing rapid demand growth compared to firms with stable or declining demand because the loss of future business by customer termination if the firm is found to be cheating implies a relatively larger cost. Therefore a smaller current premium payment is necessary to assure nonopportunistic behavior. Hence the higher the anticipated growth in demand for a firm, the lower the contracting cost of using implicit relative to explicit contracts and the more flexible prices and other contract terms set by the firm can be expected to be.⁴¹

and producer uncertainty regarding whether a particular demand shock is a relative or an aggregate shift (see, for example, Robert E. Lucas, Jr., *Some International Evidence on Output-Inflation Tradeoffs*, 63 *Am. Econ. Rev.* 326 (1973)), implicitly assumes economic agents do not observe current movements of money supply and price level. A more realistic assumption is that economic agents are not "fooled," especially over long periods, about the nature of the shock but rather are bound, either explicitly or implicitly, by long-term contracts that have previously fixed prices.

⁴¹ A crucial determinant of economic organization is therefore the anticipated demand growth compared to the actual demand growth, or the demand growth anticipated at the time of contract and the demand growth actually experienced and therefore anticipated at some later time. For example, one possible reason for the recent movement by oil-refining companies towards vertically integrated retail-marketing operations may be the increased cost of controlling franchised dealers due to the large decrease in the anticipated growth of demand for gasoline in the period since the large OPEC-initiated price increase of crude oil. With demand growing slower than originally anticipated, the initial equilibrium "premium" earned by dealers will now be less than necessary to assure their noncheating behavior. The anticipated decrease in the total number of dealers (that is, the fact that future demand is anticipated to be zero for many dealers in the new equilibrium) will create last-period problems for particular locations that can be largely avoided by employee-operated outlets. See Benjamin Klein & Keith Leffler, *supra* note 16, for a more complete discussion of these issues.

The cost to a growing firm of cheating on laborers, for example, would be higher in terms of the future increased wages (of increased employment) it would have to pay if it cheated. The penalty for not relying on the firm's brand name is then more effective. This may explain why firms such as International Business Machines appear to have highly flexible labor compensation arrangements that are, in fact, quite similar to Japanese wage payments which consist of large, highly variable, biannual bonuses. Our analysis suggests that it is not because of different cultural values that Japanese labor relations rely on much trust, but because the high growth rate of future demand makes it relatively cheap for firms to behave in this way.⁴²

D. *Leasing Inputs and Ownership of the Firm*

Examination of leasing companies should reveal that leases are less common (or too expensive) for assets with specialized quasi rents that could be appropriated by the lessee or lessor. Leasing does not occur in the obvious cases of elevators or the glass of windows in an office building where postinvestment bilaterally appropriable quasi rents are enormous, while the furniture in the building is often rented. In banks, the safe is owned by the bank, but computers (though not the memory discs) are sometimes rented.⁴³ Though this may seem like resorting to trivialities, the fact that such leasing arrangements are taken for granted merely corroborates the prior analysis.

The standard example of leasing arrangements occurs with transportation capital, such as the planes, trucks, or cars used by a firm. This capital is generally easily movable and not very specific. But leasing arrangements are far from universal because some of this capital can be quite specific and quasi rents appropriated. For example, early American steam locomotives were specialized to operating conditions such as high speed, hill climbing, short hauls, heavy loads, sharp corners, as well as types of coal for fuel. Slight differences in engines created significant differences in operating costs. High specialization made it desirable for the rail companies to own locomotives (as well as the land on which water was available for steam). The advent of the more versatile, less specialized, diesel locomotive enabled

⁴² Walter Galenson & Konosuke Odaka, *The Japanese Labor Market, in Asia's New Giant* 587 (Hugh Patrick & Henry Rosovsky eds. 1976); and Koji Taira, *Economic Development and the Labor Market in Japan* (1970), both documented the fact that this highly flexible wage feature of Japanese labor contracts did not become widespread until the postwar period, a time of extremely rapid growth.

⁴³ In addition to computers being less specific and hence possessing smaller appropriable quasi rents than elevators, firms (for example, IBM) that supply computers generally possess extremely valuable brand names per unit of current sales due to a large anticipated growth in demand. Since there are some quasi rents associated with the use of a computer by a bank that could possibly be appropriated by threat of immediate removal, we would expect that if rental contracts existed they would more likely be with highly credible firms with high anticipated demand growth.

more leasing and equipment trust financing. Similarly, Swift, the meat packer and innovator of the refrigerator car for transporting slaughtered beef, owned the specialized refrigerator cars it used.⁴⁴

On the other hand, some capital may be quite specific to other assets in a firm's productive process and yet leased rather than owned. These cases provide useful insights into the nature of the contracting costs underlying our analysis. For example, consider the fact that agricultural land, a highly specific asset, is not always owned but often is rented. Land rented for farming purposes is typically for annual crops, like vegetables, sugar beets, cotton, or wheat, while land used for tree crops, like nuts, dates, oranges, peaches, apricots, or grape vines—assets that are highly specialized to the land—is usually owned by the party who plants the trees or vines.⁴⁵ However, long-term rental arrangements even for these “specialized asset” crops are not entirely unknown.

It is instructive to recognize why land-rental contracts, rather than vertical integration, can often be used without leading to opportunistic behavior. The primary reason is because it is rather cheap to specify and monitor the relevant contract terms (the quality of the good being purchased) and to enforce this particular rental contract. In addition, the landowner generally cannot impose a cost on the farmer by pulling the asset out or reducing the quality of the asset during the litigation process. Note the contrast with labor rental where it is essentially impossible to effectively specify and enforce quality elements (for example, all working conditions and the effort expended by workers) and where the possibility of withdrawal by strike or lockout is real and costly. Therefore, we do observe firms making highly specific investments in, for example, trees or buildings on land they do not own but only rent long term.⁴⁶ This is because credible postcontractual opportunistic threats by the landowner are not possible. However, if the landowner can vary the quality of the land, for example, by controlling the irrigation system to the crops or the electricity supply to a building, then a

⁴⁴ The great bulk of all refrigerator cars are not owned by the railroads, but rather by shipper-users such as packers and dairy companies. See Robert S. Henry, *This Fascinating Railroad Business* 247 (1942).

⁴⁵ While 25% of vegetable and melon farms in California in 1974 were fully owned by the farm operator, 82% of fruit and nut tree farms were fully owned, a significantly different ownership proportion at the 99% confidence interval. Similarly, the ownership proportions of cash grain and cotton farms were 40% and 39%, respectively, both also significantly different at the 99% confidence interval from the proportions of fruit and nut tree farm ownership. See 1 U.S. Dep't of Commerce, Bureau of the Census, *1974 Census of Agriculture, State and County Data*, pt. 5, at tab. 28. Summary by Tenure of Farm Operator and Type of Organization, *id.*, 1974, California, pp. 1-29 to 1-30.

⁴⁶ Rental terms may be related to sales of the firm using the land in order to share the risk of real-value changes and to reduce the risk of nominal land-value changes involved with a long-term contract.

significant possibility of postinvestment opportunistic behavior exists and we would therefore expect vertical integration.⁴⁷

One specific asset that is almost always owned by the firm is its trade-name or brand-name capital and, in particular, the logo it uses to communicate to consumers. If this asset were rented from a leasing company, the problems would be obvious. The firm would be extremely hesitant to make any investments to build up its goodwill, for example, by advertising or by successful performance, because such investments are highly specific to that "name." The quasi rents could be appropriated by the leasing company through increases in the rental fee for the trade name. Not only would the firm not invest in this specific asset, but there would be an incentive for the firm to depreciate a valuable rented brand name. Although these problems seem insurmountable, rental of the capital input of a firm's brand name is not entirely unknown. In fact, franchisors can be thought of as brand-name leasing companies. A franchisee is fundamentally a renter of the brand-name capital (and logo) owned by the franchisor. Because of the specific capital problems noted above, direct controls are placed on franchisee behavior. The rental payment is usually some form of profit-sharing arrangement and, although the franchisee is legally considered to be an independent firm, the situation is in reality much closer to vertical integration than to the standard contractual relationship of the independent market.

Finally, the analysis throws light on the important question of why the owners of a firm (the residual claimants) are generally also the major capitalists of the firm.⁴⁸ As we have seen, owners may rent the more generalized capital, but will own the firm's specific capital. This observation has implications for recent discussions of "industrial democracy," which fail to recognize that although employees may own and manage a firm (say, through their union), they will also have to be capitalists and own the specific capital. It will generally be too costly, for example, for the worker-owners to rent a plant because such a specific investment could be rather easily appropriated from its owners after it is constructed. Therefore it is unlikely to be built. A highly detailed contractual arrangement together with very large brand-name premium payments by the laborers would be neces-

⁴⁷ Coase's example of a monopolist selling more of a durable good, say land, after initially selling a monopoly quantity at the monopoly price is analytically identical to the problem of postcontractual opportunistic behavior. Existing contractual relationships indicate, however, that the land case may be relatively easy to solve because it may not be expensive to make a credible contract regarding the remaining land. But, one of Coase's indicated solutions, the short-term rental rather than sale of the land is unlikely because it would discourage specific (to land) investments by the renter (such as building a house, developing a farm, and so forth) for fear of appropriation. See R. H. Coase, *Durability and Monopoly*, 15 *J. Law & Econ.* 143 (1972).

⁴⁸ We are grateful to Earl Thompson for discerning this implication.

sary to assure nonopportunistic behavior. This is generally too expensive an alternative and explains why capitalists are usually the owners of a firm.⁴⁹

E. *Social Institutions*

Much of the previous analysis has dealt with tangible capital. Contractual arrangements involving such assets are often cheaper than complete vertical integration, even when the assets are highly specific (for example, the land-rental case.) As the discussion on human capital suggests, however, when the specific assets involved are intangible personal assets, the problems of contract enforcement become severe. In addition, when the number of individuals involved (or the extent of the specific capital) becomes very large, ownership arrangements often become extremely complex.

For example, consider country clubs. Golf country clubs are social, in addition to being golfing, organizations. Sociability of a country club involves substantial activities away from the golf course: dinners, dances, parties, cards, games, and general social activities with friends who are members of the club. However, some golf courses are operated with very few social activities for the set of members and their families. The social clubs (usually called "country clubs") are mutually owned by the members, whereas golf courses with virtually no off-course social activity often are privately owned with members paying daily golf fees without owning the golf course.

Mutual ownership is characteristic of the social country club because the specialized quasi rent of friendship is collected by each member whose friendship is specialized to the other members. The members' behavior toward one another constitutes an investment in forming valuable friendships, a congenial milieu, and rapport among the members. Each member has invested in creating that congenial milieu and atmosphere specialized to the other members. And its value could be stolen or destroyed by opportunistic behavior of a party authorized to admit new members.

⁴⁹ Armen A. Alchian & Harold Demsetz, *Production, Information Costs, and Economic Organization*, 62 *Am. Econ. Rev.* 777 (1972), claim that if the owner of the firm also owns the firm's capital it supplies evidence that he can pay for rented inputs, including labor. This appears to be incorrect since the owner could supply credibility by using some of his assets completely unrelated to the production process, such as treasury bonds, for collateral. Michael C. Jensen & William H. Meckling, *On the Labor-Managed Firm and the Codetermination Movement* (unpublished manuscript, Feb. 1977), emphasize the costs of monitoring managerial performance and the maintenance of rented capital, and the problems of efficiently allocating risks in a pure-rental firm. They also note that it is "impossible" for a firm to rent all the productive capital assets because many of them are intangible and therefore "it is impossible to repossess the asset if the firm refused to pay the rental fee" (*id.* at 20). This argument is similar to our analysis of opportunistic behavior. However, rather than asserting that such rentals are impossible, we would merely recognize the extremely high contracting costs generally present in such situations. More importantly, we claim that such an argument also extends to the rental of tangible specific capital.

To see how, suppose the club were owned by someone other than the members. Once the membership value is created by the interpersonal activities of the members, the owner of the club could then start to raise the fees for continuing members. Assuming some costs of the members moving away en masse and forming a new club, the owner could expropriate by higher fees some of the specialized quasi-rent value of the sociability created by the members' specialization to each other in their own group. Alternatively, the owner could threaten to break the implicit contract and destroy some of the sociability capital by selling admission to "undesirable" people who want to consort with the existing members.

Similarly, if the social country club were owned by the members as a corporation with each member owning a share of stock salable without prior approval of existing members (as is the case for the business corporation), a single member could, by threatening to sell to an "undesirable" potential member, extract some value of congeniality from the current members, as a payment for not selling.⁵⁰

An extreme case of this general problem is a marriage. If each mate had a transferable share salable to a third party, there would be far fewer marriages with highly specific investments in affection and children. If a relationship is not one of specialized interest (specialized to a particular other party) or if it required no investment by any member, then the marriage relationship would be more like a corporation. As it is one of highly specific investments, marriages have historically been mutually owned entities, with permission of both parties generally required for alteration of membership. Government arbitration of this relationship to prevent postinvestment opportunistic behavior by either party can contribute toward lower bargaining costs and investments of resources (recoverable dowries) by both parties to improve their respective postinvestment bargaining positions, and, most importantly, create confidence that opportunistic behavior will not be successful. The legislative movement to "no-fault" divorce suggests that modern marriages may have less specific assets than formerly.⁵¹

⁵⁰ The "free-rider" problems of bribing an opportunistic member to prevent sale to an "undesirable" member are obvious. This analysis could be applied to social clubs such as Elks, Masonic Order, and so forth.

⁵¹ Similarly, people whose work is highly specialized to each other will be partners (common ownership). For example, attorneys that have become highly specialized to their coattorneys will become partners, whereas new associates will at first be employees. A small team of performers (Laurel and Hardy, Sonny and Cher) who were highly specialized to each other would be "partners" (co-owners) rather than employee and employer. While it is still difficult to enforce such contracts and prevent postcontractual opportunistic behavior by either party, joint ownership creates an incentive for performance and specific investment not present in an easily terminable employer-employee contract that must rely solely on the personal brand-name reputation of contracting parties. Trust, including the reputation of certifying institutions such as

The importance of mobility costs when many individuals in a group must jointly decide to take action, as in the case of an opportunistic country-club owner, and the importance of government intervention are clearly reflected in the case of the money-supply industry.⁵² The decision regarding what is used as the dominant money (medium of exchange) in society, like many other social agreements and customs, entails a large degree of rigidity on the individual level. A decision to change a social institution, in this case what is used as money, must involve a large subset of the population to be effective. Given this natural monopoly, the cost to an individual or a new entrant of attempting change may be prohibitively costly. Therefore, once a dominant money supplier is established, the potential wealth gain that can be realized through opportunistic behavior by the money issuer (that is, by unanticipated inflation) is enormous. The private implicit contractual solution would therefore entail an extremely high brand-name "premium" payment (seigniorage return) to guarantee that a wealth-maximizing, unregulated, private, dominant money supplier will not cheat by increasing the money supply faster than anticipated. Because this premium payment and therefore the rental price of money will be so high, it is unlikely that a private, implicit contractual solution is the cheapest arrangement.⁵³ Traditional vertical integration would also be extremely costly in this case of a consumer asset used by so many individuals (in fact it is difficult to even understand exactly what it would mean). Some form of government intervention is obviously likely, either in the form of regulation by enforcing an explicit contractual guarantee, or in the form of outright nationalization. Government ownership of the monetary unit is actually close to what one may consider vertical integration on the part of consumers in this particular case.

theatrical agents, law schools, and so on, and the presence of social sanctions against opportunistic partners remain important.

⁵² The following discussion extends the analysis in Benjamin Klein, *supra* note 16.

⁵³ The alternative cost of holding money will be significantly above the marginal cost of producing cash balances (where costs are defined exclusive of the costs necessary to guarantee nonopportunistic behavior), thereby leading to less than "the optimum" quantity of cash balances. See, for example, Milton Friedman, *The Optimum Quantity of Money*, in *The Optimum Quantity of Money and Other Essays* 1 (1969), for the original statement of this supposed inefficiency.

An alternative solution analytically equivalent to the "premium" solution would be the putting up by the dominant money supplier of a large forfeitable collateral bond equal to the value of the possible short-run wealth gain from cheating. This bond would be held in part by each of the demanders of the firm's money in proportion to each particular individual's money holdings and interest received on the bond by each individual would be paid to the firm if cheating did not occur. While this would not create any inefficiencies of price greater than marginal cost as implied by the premium solution, the transaction costs of enforcing such an arrangement among such a large and changing number of individuals would be extremely high. If the government acted as the consumers' agent, the solution would now be similar to a regulated industry, with the potential for opportunistic expropriation of the bond by the government.

IV. CONCLUDING COMMENT

We should emphasize in conclusion that most business relationships are neither likely to be as simple as the standard textbook polar cases of vertical integration or market contract nor as easily explained as some of the above examples. When particular examples are examined in detail, business relationships are often structured in highly complex ways not represented by either a simple rental contract or by simple vertical integration. A timely example is the ownership rights of common services supplied in condominium or "new-town" projects. One solution often adopted is joint ownership of common assets, similar to the joint ownership by petroleum producers and refiners of oil pipeline as noted above. In the condominium case, however, the number of shareowners is sometimes equal to hundreds or even thousands of individuals and the resulting contractual arrangements are closer to a constitution for a local "government" than to the simple paradigm of a two-person market transaction. When governing costs are high, individuals have often opted for a long-term management contract (often with the builder of the housing project) for maintaining the common assets. The possible problems associated with the opportunistic appropriation by the manager of the quasi rents in specialized assets of the individual owners (including specific assets used to furnish each apartment such as carpeting and any specific "friendship capital" from association with other owner occupants) are obvious. The fact that there has been a great deal of litigation in this area is not surprising. The difficulty may be partially due to what appears to be significant economies of scale in supplying confidence concerning contract performance and diseconomies of scale in the actual production and management of housing. Some insurance or franchising arrangement may therefore evolve in this area.

There is a continuing search in this difficult area using market and governmental (regulatory, legislative, and judicial) processes to produce institutional and private contractual innovation that will lead to more economical contractual relations and ownership rights. We have little idea why one solution appears to have been efficient for one condominium project and another solution for another project. This merely indicates that as we move toward more complex ownership relationships the problem of efficiently structuring the economic relationship, either within the firm or via contracts, also becomes highly complex. Stating that the world is complicated is another way of admitting our ignorance. However, explicitly recognizing that contracting costs are not zero, as they are often implicitly assumed to be in economic analysis, and explicitly considering the determinants of these costs (such as the presence of appropriable quasi rents) is the first step in explaining the large variety of contractual and ownership arrangements we observe in the real world.

More generally, we have seen that once we attempt to add empirical detail to Coase's fundamental insight that a systematic study of transaction costs is necessary to explain particular forms of economic organization, we find that his primary distinction between transactions made within a firm and transactions made in the marketplace may often be too simplistic. Many long-term contractual relationships (such as franchising) blur the line between the market and the firm. It may be more useful to merely examine the economic rationale for different types of particular contractual relationships in particular situations, and consider the firm as a particular kind or set of interrelated contracts.⁵⁴ Firms are therefore, by definition, formed and revised in markets and the conventional sharp distinction between markets and firms may have little general analytical importance. The pertinent economic question we are faced with is "What kinds of contracts are used for what kinds of activities, and why?"

⁵⁴ If we think of firms as collections of interrelated contracts rather than the collection of goods operative in the contracts, the question of who "owns" the firm (the set of contracts) appears somewhat nonsensical. It may be useful to think solely of a set of claimants to various portions of the value consequences of the contractual coalition, with no "owner" of the firm.