Transaction Cost Determinants of "Unfair" Contractual Arrangements

By BENJAMIN KLEIN*

Terms such as "unfair" are foreign to the economic model of voluntary exchange which implies anticipated gains to all transactors. However, much recent statutory, regulatory and antitrust activity has run counter to this economic paradigm of the efficiency properties of "freedom of contract." The growth of "dealer day in court" legislation, FTC franchise regulations, favorable judicial consideration of "unequal bargaining power," and unconscionability arguments, are some examples of the recent legal propensity to "protect" transactors. This is done by declaring unenforceable or illegal particular contractual provisions that, although voluntarily agreed upon in the face of significant competition, appear to be one-sided or unfair. Presentation of the standard abstract economic analysis of the mutual gains from voluntary exchange is unlikely to be an effective counterweight to this recent legal movement without an explicit attempt to provide a positive rationale for the presence of the particular unfair contractual term. This paper considers some transaction costs that might explain the voluntary adoption of contractual provisions such as termination at will and long-term exclusive dealing clauses that have been under legal attack.

I. The "Hold-up" Problem

In attempting to explain the complicated contractual details of actual market exchange, I start by noting that complete, fully contingent, costlessly enforceable contracts are not possible. This is a proposition obvious to even the most casual observer of economic phenomenon. Rather than the impersonal marketplace of costlessly enforceable contracts represented in standard economic analysis, individuals in most real world transactions are concerned with the possibility of breach and hence the identity and reputation of those with whom they deal. Further, even a cursory examination of actual contracts indicates that the relationship between transacting parties often cannot be fully described by a court-enforceable formal document that the parties have signed (see Stewart Macauley). While the common law of contracts supplies a body of rules and principles which are read into each contract, in many cases explicit terms (which include these general unwritten terms) remain somewhat vague and incomplete.

Contracts are incomplete for two main reasons. First, uncertainty implies the existence of a large number of possible contingencies and it may be very costly to know and specify in advance responses to all of these possibilities. Second, particular contractual performance, such as the level of energy an employee devotes to a complex task, may be very costly to measure. Therefore contractual breach may often be difficult to prove to the satisfaction of a third-party enforcer such as a court.

Given the presence of incomplete contractual arrangements, wealth-maximizing transactors have the ability and often the incentive to renege on the transaction by holding up the other party, in the sense of taking advantage of unspecified or unenforceable elements of the contractual relationship. Such behavior is, by definition,

*Professor of economics, University of California-Los Angeles. This paper was written while I was a Law and Economics Fellow at the University of Chicago Law School. Armen Alchian, Roy Kenney, Edmund Kitch, Timothy Muris, Richard Posner, and George Priest provided useful comments on earlier drafts.

unanticipated and not a long-run equilibrium phenomenon. Oliver Williamson has identified and discussed this phenomenon of "opportunistic behavior," and my recent paper with Robert Crawford and Armen Alchian attempted to make operational some of the conditions under which this hold-up potential is likely to be large. In addition to contract costs, and therefore the incompleteness of the explicit contract, we emphasized the presence of appropriable quasi rents due to highly firm-specific investments. After a firm invests in an asset with a low-salvage value and a quasi-rent stream highly dependent upon some other asset, the owner of the other asset has the potential to hold up by appropriating the quasi-rent stream. For example, one would not build a house on land rented for a short term. After the rental agreement expires, the landowner could raise the rental price to reflect the costs of moving the house to another lot.[1]

The solution we emphasized was vertical integration, that is, one party owning both assets (the house and the land). Because the market for land is competitive, the price paid for the land by the homebuilder does not reflect these potentially appropriable quasi rents. However, this solution will not necessarily be observed. The size of the hold-up potential is a multiplicative function of two factors: the presence of specific capital, that is, appropriable quasi rents, and the cost of contractually specifying and enforcing delivery of the service in question —the incentive for contract violation and the ease of contract violation. Even where there is a large amount of highly specific capital, the performance in question may be cheaply specifiable and measureable and a complete contract legally enforceable at low cost. Therefore, while a short-term rental contract is not feasible, a possible solution may be a long-term lease. In addition, since the cases we will be considering deal with human capital, vertical integration in the sense of outright ownership is not possible.

[1]This problem is different from the standard monopoly or bilateral monopoly problem for two reasons. First, market power is created only after the house investment is made on a particular piece of land. Such postinvestment power can therefore exist in many situations that are purely competitive preinvestment. Second, the problem we are discussing deals with the difficulties of contract enforcement. Even if some preinvestment monopoly power exists (for example, a union supplier of labor services to harvest a crop), if one can write an enforceable contract preinvestment (i.e., before the planting), the present discounted value of the monopoly return may be significantly less than the one-time postinvestment hold-up potential (which may equal the entire value of a crop ready to be harvested).

II. Contractual Solutions

Since the magnitude of the potential holdup may be anticipated, the party to be cheated can merely decrease the initial price he will pay by the amount of the appropriable quasi rents. For example, if an employer knows that an employee will cheat a certain amount each period, it will be reflected in the employee's wage. Contracts can be usefully thought to refer to anticipated rather than stated performance. Therefore the employee's behavior should not even be considered "cheating." A secretary, for example, may miss work one day a week on average. If secretary time is highly substitutable, the employer can cut the secretary's weekly wage 20 percent, hire 20 percent more secretaries and be indifferent. The secretary, on the other hand, presumably values the leisure more than the additional income and therefore is better off. Rather than cheating, we have a voluntarily determined, utility-maximizing contractual relationship.

In many cases, however, letting the party cheat and discounting his wage will not be an economical solution because the gain to the cheater and therefore his acceptable compensating wage discount is less than the cost to the firm from the cheating behavior. For example, it is easy to imagine many cases where a shirking manager will impose costs on the firm much greater than his personal gains. Therefore the stockholders cannot be made indifferent to this behavior by cutting his salary and hiring more lazy managers. The general point is that there may not be perfect substitutability between quantity and quality of particular services. Hence, even if one knew that an unspecified

element of quality would be reduced by a certain amount in attempting the holdup, an *ex ante* compensatory discount in the quoted price of the promised high quality service to the cost of providing the anticipated lower-quality supply would not make the demander of the service indifferent. Individuals would be willing to expend real resources to set up contractual arrangements to prevent such opportunism and assure high-quality supply.

The question then becomes how much of the hold-up problem can be avoided by an explicit government-enforced contract, and how much remains to be handled by an implicit self-enforcing contract. This latter type of contract is one where opportunistic behavior is prevented by the threat of termination of the business relationship rather than by the threat of litigation. A transactor will not cheat if the expected present discounted value of quasi rents he is earning from a relationship is greater than the immediate hold-up wealth gain. The capital loss that can be imposed on the potential cheater by the withdrawal of expected future business is then sufficient to deter cheating.

In our forthcoming article, Keith Leffler and I develop this market-enforcement mechanism in detail. It is demonstrated that one way in which the future-promised rewards necessary to prevent cheating can be arranged is by the payment of a sufficiently high-price "premium." This premium stream can usefully be thought of as "protection money" paid to assure noncheating behavior. The magnitude of this price premium will be related to the potential holdup, that is, to the extent of contractual incompleteness and the degree of specific capital present. In equilibrium, the present discounted value of the price-premium stream will be exactly equal to the appropriable quasi rents, making the potential cheater indifferent between cheating and not. But the individual paying the premium will be in a preferable position as long as the differential consumer's surplus from high-quality (noncheating) supply is greater than the premium.

One method by which this equilibrium quasi-rent stream can be achieved without the existence of positive firm profits is by having the potential cheater put up a forfeitable-at-will collateral bond equal to the discounted value of the premium stream. Alternatively, the potential cheater may make a highly firm-specific productive investment which will have only a low-salvage value if he cheats and loses future business. The gap between price and salvageable capital costs is analytically equivalent to a premium stream with the nonsalvageable asset analytically equivalent to a forfeitable collateral bond.

III. "Unfair" Contractual Terms

Most actual contractual arrangements consist of a combination of explicit- and implicit-enforcement mechanisms. Some elements of performance will be specified and enforced by third-party sanctions. The residual elements of performance will be enforced without invoking the power of some outside party to the transaction but merely by the threat of termination of the transactional relationship. The details of any particular contract will consist of forms of these general elements chosen to minimize transaction costs (for example, hiring lawyers to discover contingencies and draft explicit terms, paying quality-assurance premiums, and investing in nonsalvageable "brand name" assets) and may imply the existence of what appears to be unfair contract terms.

Consider, for example, the initial capital requirements and termination provisions common in most franchise contractual arrangements. These apparently one-sided terms may be crucial elements of minimum-cost quality-policing arrangements. Given the difficulty of explicitly specifying and enforcing contractually every element of quality to be supplied by a franchisee, there is an incentive for an individual opportunistic franchisee to cheat the franchisor by supplying a lower quality of product than contracted for. Because the franchisee uses a common trademark, this behavior depre-

ciates the reputation and hence the future profit stream of the franchisor.[2]

The franchisor knows, given his direct policing and monitoring expenditures, the expected profit that a franchisee can obtain by cheating. For example, given the number of inspectors hired, he knows the expected time to detect a cheater; given the costs of low-quality inputs he knows the expected extra short-run cheating profit that can be earned. Therefore the franchisor may require an initial lump sum payment from the franchisee equal to this estimated short-run gain from cheating. This is equivalent to a collateral bond forfeitable at the will of the franchisor. The franchisee will earn a normal rate of return on that bond if he does not cheat, but it will be forfeited if he does cheat and is terminated.

In many cases franchisee noncheating rewards may be increased and short-run cheating profits decreased (and therefore franchisor direct policing costs reduced) by the grant of an exclusive territory or the enforcement of minimum resale price restraints (see my paper with Andrew McLaughlin). Franchisors can also assure quality by requiring franchisee investments in specific (nonfully salvageable) production assets that upon termination imply a capital-cost penalty larger than any short-run wealth gain that can be obtained by the franchisee if he cheats. For example, the franchisor may require franchisees to rent from them short term (rather than own) the land upon which their outlet is located. This lease arrangement creates a situation where termination implies that the franchisor can require the franchisee to move and thereby impose a capital loss on him up to the amount of his initial nonsalvageable investment. Hence a form of collateral to deter franchisee cheating is created.[3]

It is important to recognize that franchise termination, if it is to assure quality compliance on the part of franchisees, must be unfair in the sense that the capital cost imposed on the franchisee that will optimally prevent cheating must be larger than the gain to the franchisee from cheating. Given that less than infinite resources are spent by the franchisor to monitor quality, there is some probability that franchisee cheating will go undetected. Therefore termination must become equivalent to a criminal-type sanction. Rather than the usually analyzed case of costlessly detected and policed contract breach, where the remedy of making the breaching party pay the cost of the damages of his specific breach makes economic sense, the sanction here must be large enough to make the expected net gain from cheating equal to zero. The transacting parties contractually agree upon a penalty-type sanction for breach as a means of economizing on direct policing costs. Because contract enforcement costs (including litigation costs which generally are not collectable by the innocent party in the United States) are not zero, this analysis provides a rationale against the common law prohibition of penalty clauses.

The obvious concern with such seemingly unfair contractual arrangements is the possibility that the franchisor may engage in opportunistic behavior by terminating a franchisee without cause, claiming the franchise fee and purchasing the initial franchisee investment at a distress price. Such behavior may be prevented by the

[2]At locations where this incentive is very large, for example, on superhighways where the probability of repeat sales by particular customers is very low, the franchisor may "vertically integrate" and not compensate their employees on any profit-sharing basis. Such fixed wage compensation schemes reduce the incentive to cheat but at the cost of reducing the incentive for workers to supply any effort that is not explicitly specified and measureable by the employer. It is this latter incentive that is harnessed by franchising arrangements.

[3]The initial franchise investment also serves as a means of establishing an efficient compensation mechanism. Because the franchise investment is a saleable asset it provides a market measure of future profit and hence a precise incentive on franchisee efforts to build up the business. While an employee contract can contain a profit-sharing arrangement, and retirement and stock option provisions to reward employee efforts that yield a return far in the future and protect the employee's heirs, it would be extremely difficult to write *ex ante* complete, enforceable (i.e., measureable) contract terms that would as accurately reflect the value of marginal employee efforts.

depreciation of the franchisor's brand name and therefore decreased future demand by potential franchisees to join the arrangement. However, this protective mechanism is limited by the relative importance of new franchise sales compared to the continuing franchising operation, that is, by the "maturity" of the franchise chain.

More importantly, what limits reverse cheating by franchisors is the possible increased cost of operating the chain through an employee operation compared to a franchise operation when such cheating is communicated among franchisees. As long as the implicit collateral bond put up by the franchisee is less than the present discounted value of this cost difference, franchisor cheating will be deterred. Although explicit bonds and price premium payments cannot simultaneously be made by both the franchisee and the franchisor, the discounted value of the cost difference has the effect of a collateral bond put up by the franchisor to assure his noncheating behavior. This explains why the franchisor does not increase the initial franchise fee to an arbitrarily high level and correspondingly decrease its direct policing expenditures and the probability of detecting franchisee cheating. While such offsetting changes could continue to optimally deter franchisee cheating and save the real resource cost of direct policing, the profit from and hence the incentive for reverse franchisor cheating would become too great for the arrangement to be stable.

Franchisees voluntarily signing these agreements obviously understand the termination-at-will clause separate from the legal consequences of that term to mean nonopportunistic franchisor termination. But this does not imply that the court should judge each termination on these unwritten but understood contract terms and attempt to determine if franchisor cheating has occurred. Franchisees also must recognize that by signing these agreements they are relying on the implicit market-enforcement mechanism outlined above, and not the court to prevent franchisor cheating. It is costly to use the court to regulate these terminations because elements of performance are difficult to contractually specify and to measure. In addition, litigation is costly and time consuming, during which the brand name of the franchisor can be depreciated further. If these costs were not large and the court could cheaply and quickly determine when franchisor cheating had occurred, the competitive process regarding the establishment of contract terms would lead transactors to settle on explicit governmentally enforceable contracts rather than rely on this implicit market-enforcement mechanism.

The potential error here is, after recognizing the importance of transaction costs and the incomplete "relational" nature of most real world contracts, to rely too strongly on the government as a regulator of unspecified terms (see Victor Goldberg). While it is important for economic theory to handle significant contract costs and incomplete explicit contractual arrangements, such complexity does not imply a broad role for government. Rather, all that is implied is a role for brand names and the corresponding implicit market enforcement mechanism I have outlined.

IV. Unequal Bargaining Power

An argument made against contract provisions such as termination-at-will clauses is that they appear to favor one party at the expense of another. Hence it is alleged that the terms of the agreement must have been reached under conditions of "unequal bargaining power" and therefore should be invalid. However, a further implication of the above analysis is that when both parties can cheat, explicit contractual restraints are often placed on the smaller, less well-established party (the franchisee), while an implicit brand name contract-enforcement mechanism is relied on to prevent cheating by the larger, more well-established party (the franchisor).

If information regarding quality of a product supplied by a large firm is communicated among many small buyers who do not all purchase simultaneously, the potential holdup relative to, say, annual sales is reduced substantially compared to

the case where each buyer purchased from a separate independent small firm. There are likely to be economies of scale in the supply of a business brand name, because in effect the large firm's total brand name capital is put on the line with each individual sale. This implies a lower cost of using the implicit contract mechanism, that is, a lower-price premium necessary to assure non-breach, for a large firm compared to a small firm. Therefore one side of the contract will be relatively more incomplete.

For example, in a recent English case using the doctrine of inequality of bargaining power to bar contract enforcement, an individual songwriter signed a long-term (ten-year) exclusive service contract with a music publisher for an agreed royalty percentage.[4] Since it would be extremely costly to write a complete explicit contract for the supply of publishing services (including advertising and other promotion activities whose effects are felt over time and are difficult to measure), after a songwriter becomes established he has an incentive to take advantage of any initial investment made by a publishing firm and shift to another publisher. Rather than rely on the brand name of the songwriter or require him to make a specific investment which can serve as collateral, the exclusive services contract prevents this cheating from occurring.

The major cost of such explicit long-term contractual arrangements is the rigidity that is created by the necessity of setting a price or a price formula *ex ante*. In this song publishing case, the royalty formula may turn out *ex post* to imply too low a price to the songwriter (if, say, his cooperative promotional input is greater than originally anticipated.) If the publisher is concerned about his reputation, these royalty terms will be renegotiated, a common occurrence in continuing business relationships.

If an individual songwriter is a small part of a large publisher's total sales, and if the value of an individual songwriter's ability generally depreciates rapidly or does not persist at peak levels so that signing up new songwriters is an important element of a publisher's continuing business, then cheating an individual songwriter or even all songwriters currently under contract by refusing to renegotiate royalty rates will imply a large capital cost to the publisher. When this behavior is communicated to other actual or potential composers, the publisher's reputation will depreciate and future business will be lost. An individual songwriter, on the other hand, does not generally have large, diversified long-term business concerns and therefore cannot be penalized in that way. It is therefore obvious, independent of any appeal to disparity of bargaining power, why the smaller party would be willing to be bound by an explicit long-term contract while the larger party is bound only implicitly and renegotiates terms that turn out *ex post* to be truly divergent from *ex ante*, but unspecified, anticipations.

However, the possibility of reverse publisher cheating is real. If, for example, the songwriter unexpectedly becomes such a great success that current sales by this one customer represents a large share of the present discounted value of total publisher sales, the implicit contract enforcement mechanism may not work. Individuals knowingly trade off these costs of explicit and implicit-enforcement mechanisms in settling upon transaction cost-minimizing contract terms. Although it would be too costly in a stochastic world to attempt to set up an arrangement where no cheating occurs, it is naive to think that courts can cheaply intervene to discover and "fix up" the few cases of opportunistic behavior that will occur. In any event, my analysis makes it clear that one cannot merely look at the agreed upon, seemingly "unfair" terms to determine if opportunism is occurring.

V. Conclusion

Ronald Coase's fundamental insight defined the problem. With zero transaction costs, the equilibrium form of economic organization is indeterminate. However, rather than distinguishing between the crude alternatives of vertical integration and

[4]See *Macaulay v. Schroeder Publishing Co., Ltd.* discussed in M. J. Trebilcock.

market exchange, what we really have to explain are different types of market-determined contractual relationships. I have argued that a particular form of transaction cost based upon the existence of incomplete contracts (due to uncertainty and measurement costs)—a transaction cost I have called the hold-up problem—may be an important reason in many cases for termination-at-will and exclusive-dealing contractual arrangements.

The danger is that a discussion of hold-up-type transaction costs can lead to *ad hoc* theorizing. The discussion here was meant to be suggestive. If economists are to explain satisfactorily the form of particular complex contracts adopted in the marketplace, they must "get their hands dirty" by closely investigating the facts and state of the law to determine hold-up possibilities and contract enforcement difficulties in particular cases. The most useful legal input to obtain knowledge of the institutional constraints on the trading process, is not likely to come from professors of contract law. Rather, we should consider the knowledge accumulated by practicing attorneys familiar with the likely hold-up problems and the contractual solutions commonly adopted in particular industries. When all firms in a particular industry use similar contractual provisions, it is unlikely to be the result of duress or fraud and should not necessarily be considered (as some courts have) as evidence of collusion. Such uniformity suggests the existence of independent attempts within a competitive environment to solve an important common problem and signals the presence of a prime research prospect.

REFERENCES

R. J. Coase, "The Nature of the Firm," *Economica*, Nov. 1937, *4*, 386–405.

V. P. Goldberg, "Toward an Expanded Economic Theory of Contract," *J. Econ. Issues*, Mar. 1976, *10*, 45–61.

B. Klein, R. G. Crawford, and A. A. Alchian, "Vertical Integration, Appropriable Rents and the Competitive Contracting Process," *J. Law Econ.*, Oct. 1978, *21*, 297–326.

_____**and K. Leffler,** "Non-Governmental Enforcement of Contracts: The Role of Market Forces in Guaranteeing Quality," *J. Polit. Econ.*, forthcoming.

_____**and A. McLaughlin,** "Resale Price Maintenance, Exclusive Territories and Franchise Termination: The Coors Case," unpublished manuscript, Univ. California-Los Angeles 1979.

S. Macauley, "Non-Contractual Relations in Business: A Preliminary Study," *Amer. Soc. Rev.*, Feb. 1963, *28*, 55–69.

M. J. Trebilcock, "The Doctrine of Inequality of Bargaining Power: Post-Benthamite Economics in the House of Lords," *Univ. Toronto Law J.*, Fall 1976, *26*, 359–85.

Oliver E. Williamson, *Markets and Hierarchies: Analysis and Antitrust Implications*, New York 1975.