



Vertical Integration as a Self-Enforcing Contractual Arrangement

Benjamin Klein; Kevin M. Murphy

The American Economic Review, Vol. 87, No. 2, Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association. (May, 1997), pp. 415-420.

Stable URL:

<http://links.jstor.org/sici?sici=0002-8282%28199705%2987%3A2%3C415%3AVIAASC%3E2.0.CO%3B2-Y>

The American Economic Review is currently published by American Economic Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aea.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Vertical Integration as a Self-Enforcing Contractual Arrangement

By BENJAMIN KLEIN AND KEVIN M. MURPHY*

A recent wave of large vertical mergers presents a challenge to established theories of vertical integration. The large mergers that have occurred in the pharmaceutical industry between drug manufacturers and companies that manage drug insurance benefits (such as Merck's acquisition of Medco) and in the entertainment industry between program suppliers and network distributors (such as Disney's acquisition of Capital Cities/ABC) do not seem to fit traditional economic theories of vertical integration. The proximate causes for these mergers are fairly obvious. The entertainment mergers have been motivated by regulatory changes that permit TV networks to own the product they distribute, and the drug industry mergers have been motivated by the demonstrated ability of drug insurance managers to influence the sales share of different patented drugs within a therapeutic category. However, it is not obvious why these changes in the market environment led to vertical integration.

To illuminate the economic motivation for these recent vertical mergers, we present an analysis of vertical integration that combines and extends our work on self-enforcing contracts (Klein and Murphy, 1988; Klein, 1996) with earlier work on vertical integration to avoid holdups (Klein et al., 1978). In what follows we first show that competitive, non-free-riding distributors often face a distorted incentive to supply the promotional services desired by manufacturers. We then explain

why the usual contractual solution to this distributor "malincentive" problem is likely to combine court-enforcement and self-enforcement mechanisms. Finally, we outline how vertical integration may facilitate such a self-enforcing contractual arrangement.

I. The Vertical Distortion in Distribution Arrangements

Consider the supply of a product consisting of two stages: an upstream manufacturer and a downstream distributor. It is usually assumed that the manufacturer need not control distribution when there is competition among distributors and no distributor "free-riding." The manufacturer can sell its product to distributors and expect the desired level and type of services to be supplied in the downstream distribution market. However, contrary to current antitrust policy doctrine, a manufacturer may frequently find it necessary to control distribution under these circumstances.

The necessity for manufacturer control of distribution follows from the fact that manufacturers generally face negatively sloped demands for their products. Consequently, the wholesale price they receive is significantly greater than marginal cost. As a result, manufacturers often find it profitable to engage in promotional marketing. Such marketing involves holding the price of the product constant yet moving down the manufacturer's demand curve by supplying services to consumers at a price less than their value (most often free of charge). Such promotional activity thereby lowers the effective price paid by consumers on the margin and induces additional sales.

The obvious economic question is why manufacturers would want to lower price in this way, namely, by supplying services below cost that are often less valuable to consumers

[†] *Discussants:* Scott E. Masten, University of Michigan; Oliver Hart, Harvard University; Bengt Holmstrom, Massachusetts Institute of Technology.

* Department of Economics, University of California, Los Angeles, CA 90024, and Graduate School of Business, University of Chicago, Chicago, IL 60637, respectively.

than a price decrease. The answer is that such promotion "targets" particular consumers. In fact, provision of services below cost is the very definition of promotion; if consumers separately demanded these services and distributors priced them at full cost, the services would be considered part of the product or a separate product, not promotion.¹

Since consumers do not pay the full cost of promotional services, uncontrolled competition among distributors will not lead to the desired quantity of promotion. In particular, distributors will often supply less promotion than is jointly profit-maximizing for the manufacturer and distributor. We illustrated this in Klein and Murphy (1988) with an example of a perfume manufacturer that had a high margin on incremental sales and, therefore, desired that substantial promotional activity be undertaken by its retail distributors. The manufacturer may wish its distributors to hire extra sales clerks willing to take the time to demonstrate its product, or to provide additional shelf space or, more generally, to engage in activities that increase the demand on the margin for the manufacturer's product. Because distributors ignore the substantial additional profit received by the upstream manufacturers when deciding how much of such promotion to supply and because the demand price of promotion is less than cost, competition will not force distributors to provide the desired amount of promotion.

This vertical distortion is not the "successive monopoly" or "double marginalization" problem since the distortion exists even when there is competition in the downstream stage. It is the commonly recognized variable-proportions distortion. As in the standard variable-proportions case, distributors adjust the input they have control over (promotion) to use less of the input on which the manufacturer has increased its margin (the good). However, contrary to the usual case of variable

proportions where the input that distributors control is assumed to be a substitute for the manufacturer's good, the distributor's input here is a complement for the manufacturer's good. Hence, economic adjustment by the distributor involves supplying less promotion.

This vertical distortion regarding distributor supply of promotional activity is especially large in the pharmaceutical and entertainment industries. Firms in these industries generally supply goods that are differentiated and have extremely low marginal costs. Therefore, the manufacturer margin on incremental sales and the demand for promotional activity are very high, resulting in a significant manufacturer incentive to control the vertical distortion.

II. Alternative Solutions to the Vertical Distortion

A. Contractual Solutions

Manufacturers may control distributor promotion by directly contracting for desired promotional services. However, much of the promotion supplied by distributors cannot easily be directly contracted for. While the manufacturer may know after the fact if a distributor has performed as desired, the indications the manufacturer uses to make this determination are complex and difficult to specify contractually. A contract may specify the level of a particular input, such as the number of salespeople, but such contract terms are imperfect measures of true distributor promotional effort.

Moreover, writing contracts based on imperfect measures of performance may actually make things worse by inducing non-performance. Since contract terms are imperfect, they create the opportunity for transactors to use the court to enforce a literal term in a manner that is contrary to the clear intent of the transacting parties. This is illustrated in the Fisher-General Motors case, where Fisher took advantage of the imperfect cost-plus, exclusive-dealing contract, which was designed to prevent the parties from expropriating rents from one another, to adopt an extremely labor-intensive production process and to locate plants far away from the General Motors as-

¹ Although consumers are not willing to pay the full cost of promotion, this does not imply inefficiency (see Gary Becker and Murphy, 1993). Our focus here is not on the socially optimal level of promotion, but on the conditions under which manufacturers find it profitable to control distributor supply of promotion.

sembly plant. The result was inefficiently produced automobile bodies that were highly profitable for Fisher to supply, but very costly for General Motors to buy (Klein et al., 1978).

For this reason, the degree of contractual specification is not determined by the narrow transaction costs of writing an additional term—what one of us has derisively referred to as “ink costs.” Instead, transactors trade off the obvious benefit of being able to use the court to enforce a specific element of performance with the less obvious cost of creating an opportunity to use the court to enforce elements of nonperformance. As the Fisher–General Motors case vividly illustrates, the use of imperfect contract terms solves nonperformance problems in some states of the world but creates nonperformance problems in other states of the world.

These costs of contractual specification can be labeled “rigidity costs,” because an agreement that is formalized in a binding contract cannot cheaply be altered. On the other hand, leaving elements of performance unspecified creates flexibility. If, for example, the Fisher–General Motors understanding had not been formalized in a written contract, the parties would have been able to adjust flexibly or to opt out of the arrangement whenever they wished without being forced by the court to adhere to prespecified terms of a written agreement.

B. Self-Enforcing Solutions

To avoid the rigidity costs associated with contractual specification, transactors will intentionally leave some elements of performance unspecified to be enforced with a self-enforcing mechanism. A self-enforcing mechanism does not have the problems associated with imperfect contract terms because transactors generally can be assumed to know, with a lag, if their transacting partner has performed or not. The threat of termination of the relationship after nonperformance is detected, with the loss of future rents that would be earned in the relationship by the nonperforming party, is what assures performance.

Unfortunately, however, transactors cannot rely entirely on self-enforcement, because the

magnitude of the private sanction that can be imposed for nonperformance is limited. We designate the capital cost that can be imposed for nonperformance by R , the transactor's reputational capital. Transactors deciding whether to perform or not compare the gains from not performing with R . Returning to our distribution example, a distributor can be assumed to choose between two levels of promotional activity, the desired high level of promotion, P_H , or a low minimum level of promotion, P_L , that the parties may be able to specify contractually. The difference between P_H and P_L represents the unspecified elements of distributor performance. The distributor will supply the desired high level of promotion if the gain from nonperformance [i.e., the difference between the expected discounted value of the profit it could earn from supplying the low level of promotion, $\Pi(P_L)$, and the expected discounted profit it could earn from supplying the high level of promotion, $\Pi(P_H)$] is less than the reputational sanction that can be imposed by the manufacturer via termination:

$$(1) \quad \Pi(P_L) - \Pi(P_H) < R.$$

More generally, since future market conditions and hence the future gains from nonperformance are uncertain at the time individuals enter into their contractual relationship, the amount of each transactor's reputational capital can be thought of as more generally defining the self-enforcing range of the contractual relationship. The self-enforcing range measures the extent to which market conditions can change, thereby altering the gains to one or the other party from nonperformance, without precipitating nonperformance (Klein, 1996).

C. Self-Enforcing Contractual Solutions

Because reputational capital and, therefore, the self-enforcing range are limited, transactors often decide to use court-enforced terms (even though they are imperfect) to supplement self-enforcement. The contract terms chosen maximize the self-enforcing range of the relationship but entail the additional risk that conditions may change such

that it becomes wealth-maximizing for one transactor not to perform by taking advantage of the imperfect contract terms. Such a stochastic self-enforcing contractual framework explains what occurred in the General Motors-Fisher Body case.

It is not that General Motors was naive or deceived by Fisher when it agreed to the contract. Rather, it was because of the very large unexpected increase in market demand for closed auto bodies, and hence in Fisher's gain from not performing, that Fisher found itself outside the self-enforcing range. Once Fisher's gain from not performing became greater than the private sanction that could be imposed on Fisher by General Motors, Fisher found it profitable to violate the intent of the contractual understanding by taking advantage of the imperfect terms of the agreement. If the demand for closed auto bodies had not grown so dramatically, the contract Fisher and General Motors had adopted, although imperfect, would have "worked" since the gains to Fisher from taking advantage of the contract would have remained less than Fisher's reputational capital.

The role of contract terms within this framework is very different from the standard principal-agent view of contracts, where contract terms are designed to create optimal incentives for some imperfect court-enforceable proxy for performance. Instead, contract terms are chosen by transactors more generally to facilitate self-enforcement. Facilitation of self-enforcement may be accomplished in either one of two ways. First, contract terms can operate on the left-hand side of equation (1), limiting the expected gains from nonperformance and thereby reducing the amount of reputational capital necessary to make the arrangement self-enforcing. Alternatively, contract terms can operate on the right-hand side of (1), shifting reputational capital between transactors so that each party's reputational capital coincides more closely with its potential expected gain from nonperformance. This second effect is the major rationale for many contractual restrictions found in distribution arrangements, such as resale-price maintenance or exclusive territories, which have the effect of shifting some of the manufacturer's reputational capital to the distributor

and creating a situation where the distributor has more to lose if it does not perform (Klein and Murphy, 1988). In either case, whether the contract terms chosen by transactors control the expected gains from nonperformance or shift reputational capital between the parties, contract terms have the effect of economizing on reputational capital and widening the range of *ex post* market conditions where the contractual arrangement remains self-enforcing.

III. Vertical Integration as a Self-Enforcing Contractual Solution

Vertical integration can be explained as a particular contractual arrangement that often facilitates self-enforcement (Klein and Murphy, 1996). To understand this, we take as our starting point Sanford J. Grossman and Oliver D. Hart's (1986) highly influential model of integration. Grossman and Hart's fundamental insight is that integration minimizes inefficiencies (in particular, investment and managerial inefficiencies, the two elements of performance that they assume cannot be contracted on) by allocating all contractually unspecified or residual rights to the owner. For example, once General Motors owns Fisher Body and Fisher becomes an employee of General Motors, Fisher can no longer take advantage of an imperfect contract to threaten nonperformance because the ultimate power to make most important investment and management decisions, such as the location of the auto-body production plant, is legally transferred to the employer/owner. Employees, even employee managers with authority to make plant-location decisions, do not have the ability to take advantage of such authority because any threat of nonperformance can legally be overridden by the employer.

This reduction in the ability of employees to threaten nonperformance can be seen clearly if we consider the possibility of bribes or side payments by third parties to influence an employee's plant-location decision. While an independent contractor has the right (and even a duty to its shareholders) to consider such payments in making decisions, employees cannot do so. Employees have a fundamental legal obligation of loyalty to their

employer and can be sued for damages if they make decisions on the basis of such payments (Scott Masten, 1988).

While this important insight from Grossman and Hart (1986) regarding the allocation of residual rights to generate optimal performance is correct, it is necessary to build upon the insight to explain how vertical integration operates and to predict where vertical integration is likely to be used. In particular, it is necessary to place the Grossman and Hart insight within a self-enforcing contractual framework.² Without the possibility of self-enforcement, the inefficiencies that Grossman and Hart discuss are likely to be huge. The key to aligning transactor incentives and controlling such inefficiencies is self-enforcement.

Once we are in a self-enforcing contractual framework, it is necessary to take account of the reputational capital possessed by transactors, in addition to their incentives to exploit residual rights to hold up one another, in order to determine who will be the owner/employer and who will be the nonowner/employee. Even if the shift in power from the employee to the employer increased the potential gains and, therefore, the inefficiencies from nonperformance, it would not mean that the wrong transactor was chosen to be the owner/employer. If the employer's reputational capital is larger than the employee's reputational capital, the chosen organizational arrangement could make economic sense. Vertical integration operates by allocating residual rights and thereby shifting transactors' expected gains from nonperformance so that each transactor's expected gain more closely coincides with its reputational capital. By facilitating self-enforcement on both sides of the transaction, the probability that the relationship will remain within the self-enforcing range increases.

Vertical integration also facilitates self-enforcement by dividing assets among the transactors so that it is easier to replace non-

performing transactors. For example, once General Motors owns Fisher Body, it becomes less costly for General Motors to replace Fisher for nonperformance. Because the labor market is highly competitive and specific human-capital investments relatively small, it is likely to be easier for General Motors to find a new plant manager than to find a new substitute auto-body supplier. Alternatively, if Fisher possessed a large amount of human capital that was specific to the auto-body plant, it might be more difficult to replace Fisher for nonperformance than to find a new auto-body supplier, and facilitation of self-enforcement would imply that Fisher would own the auto-body plant. The recognition that the degree of asset specificity can vary depending upon how assets are allocated across transactors is a major improvement over existing models, where asset specificity is assumed to be exogenous with respect to ownership.

Allocating ownership across assets to facilitate self-enforcement by limiting the degree of asset specificity implies obvious "natural" organizational boundaries. For example, if one were starting a firm and deciding which inputs to buy and which to make, an input freely available in the marketplace, such as pencils, would be purchased rather than made because the presence of ready alternatives means that self-enforcement can be used to assure performance. This would be true even if there were a significant performance-measurement problem (so that one could not contractually specify and court-enforce the quality of pencils) and even if pencils were a relatively important input (so that the inefficiencies associated with lower-quality pencils were large). In spite of these conditions, there would be no reason for the firm to own a pencil-manufacturing facility because nonperforming pencil suppliers could easily be replaced.

In addition, vertical integration may increase flexibility by reducing the degree of contractual specification. Within a self-enforcing framework the degree to which performance is spelled out contractually is not exogenous with respect to organizational form. The extent to which transactors use written contract terms is not determined by "ink costs," but by the necessity to supplement self-enforcement. When transactors use an

² A recent paper by George Baker et al. (1996) presents a model that formally integrates self-enforcement with Grossman and Hart's (1986) model. However, the Baker et al. paper does not identify what we consider to be the key aspects of vertical integration that facilitate self-enforcement.

employment relationship, a detailed contract usually is not necessary to supplement self-enforcement because, by legally shifting power from the employee to the employer, a situation is created where the levels of both transactors' reputational capital are substantially greater than their potential gain from nonperformance. As a result, it is unnecessary to control either transactor's behavior contractually. Therefore, an employer/owner can more flexibly "direct" its employees, often instructing them merely to "come to work and then I'll tell you what to do," eliminating the ability of employees to take advantage of imperfect contract terms.

An implication is that vertical integration is more likely to be used when there is significant market uncertainty. Rather than explaining this frequently noted phenomenon by risk-aversion or assured-supply considerations or by the inability to write contract terms that optimally align incentives over time, increased uncertainty leads to vertical integration because it makes it more likely that the alternative of an explicitly specified performance contract, such as the Fisher-General Motors contract, will move outside the self-enforcing range. Vertical integration (and other self-enforcing contractual arrangements such as sharing contracts) are used in such circumstances because these arrangements are more likely to assure the *ex post* distribution of rents between transactors relative to their gains from nonperformance necessary to support self-enforcement in the face of market disturbances.

The key analytical point is that contract terms and organizational structures, including vertical integration, can often be explained not on the basis of explicit contractual alignment of incentives or risk considerations (un-

fortunately, what the economic literature concentrates on), but on the facilitation of self-enforcement.

REFERENCES

- Baker, George; Gibbons, Robert and Murphy, Kevin J. "Implicit Contracts and the Theory of the Firm." Working paper, Harvard University, October 1996.
- Becker, Gary S. and Murphy, Kevin M. "A Simple Theory of Advertising as a Good or Bad." *Quarterly Journal of Economics*, November 1993, 108(4), pp. 941-64.
- Grossman, Sanford J. and Hart, Oliver D. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, August 1986, 94(4), pp. 691-719.
- Klein, Benjamin. "Why Hold-ups Occur: The Self-Enforcing Range of Contractual Relationships." *Economic Inquiry*, July 1996, 34(3), pp. 444-63.
- Klein, Benjamin; Crawford, Robert G. and Alchian, Armen A. "Vertical Integration, Appropriable Rents and Competitive Contracting Process." *Journal of Law and Economics*, October 1978, 21(2), pp. 297-326.
- Klein, Benjamin and Murphy, Kevin M. "Vertical Restraints As Contract Enforcement Mechanisms." *Journal of Law and Economics*, October 1988, 31(2), pp. 265-97.
- . "The Firm as a Self-Enforcing Contractual Arrangement." Unpublished manuscript, University of California-Los Angeles, 1996.
- Masten, Scott. "A Legal Basis for the Firm." *Journal of Law, Economics, and Organization*, Spring 1988, 4(1), pp. 181-98.