

ISSN 0265-8003

THE COSTS AND BENEFITS OF OWNERSHIP:
A THEORY OF VERTICAL AND LATERAL INTEGRATION

Sanford Grossman

and

Oliver Hart

Discussion No. 70
July 1985

Centre for Economic Policy Research
6 Duke of York Street
London SW1Y 6LA

Tel: 01 930 2963

The research described in this Discussion Paper is part of the Centre's research programme in **Applied Economic Theory and Econometrics**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. The CEPR is a private educational charity which promotes independent analysis of open economies and the relations between them. The research work which it disseminates may include views on policy, but the Centre itself takes no institutional policy positions.

These discussion papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

CEPR Discussion Paper No. 70
July 1985

The Costs and Benefits of Ownership:
A Theory of Vertical and Lateral Integration

ABSTRACT*

What determines how integrated a firm is? We emphasize the benefits of "control" when there are difficulties in writing complete contracts. We define the firm as being composed of its assets. We present a theory of costly contracts which emphasizes that contractual rights can be of two types: specific rights and residual rights. When it is too costly to list all specific rights over assets in the contract, it may be optimal to let one party purchase all residual rights. Ownership is the purchase of these residual rights. We show that there can be costs associated with the wrong allocation of residual rights.

JEL classification: O22, 611

Keywords: theory of the firm, vertical and horizontal integration, contracts

Sanford J Grossman
Department of Economics
University of Chicago
1126 East 59th Street
Chicago ILLINOIS 60637
USA
(0101) 312 962 8178

Oliver A Hart
Department of Economics
MIT
Cambridge MA 02138
(0101) 617 253 6930

*Research supported by the National Science Foundation, the Alfred P. Sloan Foundation, and the International Centre for Economics and Related Disciplines at the London School of Economics. This is an April 1985 version of an earlier paper, entitled, "Vertical Integration and the Distribution of Property Rights". We gratefully acknowledge helpful comments from Debra Aron, Richard Epstein, Paul Joskow, Naava Grossman, John Minahan, Andrew Postlewaite, Peter Thistle, Martin Weitzman, Oliver Williamson, Charles Wilson, Toni Zabaiza, and two referees.

NON-TECHNICAL SUMMARY

What is a firm? What are the determinants of how vertically or laterally integrated are the activities of the firm? This paper builds on the foundations laid by Coase (1937), Williamson (1979) and Klein et al (1978) which emphasize the benefits of "control" in response to situations where there are difficulties in writing or enforcing complete contracts. We define the firm as being composed of the assets (e.g. machines, inventories) which it owns. We present a theory of costly contracts which emphasizes that contractual rights can be of two types: specific rights and residual rights. When it is too costly for one party to specify a long list of the particular rights it desires over another party's assets, it may be optimal to purchase all the rights except those specifically mentioned in the contract.

Ownership is the purchase of these residual rights of control. We show that there can be harmful effects associated with the wrong allocation of residual rights. In particular a firm which purchases its supplier, thereby removing residual rights of control from the manager of the supplying company, can distort the manager's incentives sufficiently to make common ownership harmful. We develop a theory of integration based upon the attempt of parties in writing a contract to allocate efficiently the residual rights of control between themselves.

We define a firm to consist of those assets which it owns or over which it has control--we do not distinguish between ownership and control and virtually define ownership as the power to exercise control. Of course, control or ownership is never absolute. For example, a firm which owns a machine may not be able to sell it without the permission of the lenders for which the machine serves as collateral; more generally, a firm may give another firm specific authority over its machines. However, ownership gives the owner all rights to use the machine which he hasn't voluntarily given away or which the government or some other party has not taken by force. We believe that this terminology is roughly consistent with standard usage.

(ii)

It may be extremely costly to write a contract which specifies unambiguously the payments and actions of all parties in every observable state of nature. We assume that integration in itself does not change the cost of writing down a particular contractual provision. What it does change is who has control over those provisions not included in the contract. Consider, for example, a contract between a publisher and a printer for a particular number of copies of a book. If the contract has no provision for an additional print run, but the publisher receives some new information which makes it profitable for another run, then it is obvious that the right to decide whether or not to have the run belongs to the owner of the printing press. This is the simplest possible example of our assumption that the benefit of ownership is the residual right of control, i.e. the right to control all actions that have not been explicitly given away by contract.

It is worth emphasizing that since ownership provides residual rights of control, the usual argument that the feasible set can only become larger under integration fails. If firm 1 buys firm 2, the owner of firm 1 will have the power to intervene in firm 2 in all sorts of ways, some of which may be very undesirable and lead to large efficiency losses. The point is that the owner cannot commit himself to intervene selectively in his subsidiary's operations since by their very definition residual rights refer to powers that cannot be specified in advance (at least in the detail required to make them part of an enforceable contract). It follows immediately from this that integration can impose costs as well as benefits.

1a. General Introduction

What is a firm? What are the determinants of how vertically or laterally integrated are the activities of the firm? This paper builds on the foundations laid by Coase (1937), Williamson (1979) and Klein et al (1978) which emphasize the benefits of "control" in response to situations where there are difficulties in writing or enforcing complete contracts.¹ We define the firm as being composed of the assets (e.g. machines, inventories) which it owns. We present a theory of costly contracts which emphasizes that contractual rights can be of two types: specific rights and residual rights. When it is too costly for one party to specify a long list of the particular rights it desires over another party's assets, it may be optimal to purchase all the rights except those specifically mentioned in the contract. Ownership is the purchase of these residual rights of control. We show that there can be harmful effects associated with the wrong allocation of residual rights. In particular a firm which purchases its supplier, thereby removing residual rights of control from the manager of the supplying company, can distort the manager's incentives sufficiently to make common ownership harmful. We develop a theory of integration based upon the attempt of parties in writing a contract to allocate efficiently the residual rights of control between themselves.

We begin by reviewing some transactions cost based arguments for integration. Conser (1937) suggested that transactions will be organized in the firm when the cost of doing this is lower than the cost of using the market. He added some content to this idea by suggesting that the costs of constant recontracting with an outside firm or manager can be high relative to those of signing a long term contract with an employee in which the employee agrees to carry out the commands of the employer. Williamson and Klein et al added further content to this theory by arguing that a contractual relationship between a separately owned buyer and seller will be plagued by opportunistic and inefficient behavior in situations where there are large amounts of surplus to be divided ex-post, and where, because of the impossibility of writing a complete, contingent contract, the ex-ante contract does not specify a clear division of this surplus. Such situations in turn are likely to arise when either the buyer or seller must make investments which have a smaller value in a use outside their own relationship than within the relationship (i.e. there exist "asset specificities").

While these statements help us understand when the costs of contracting between separately owned firms may be high, they do not elucidate what the benefits are of "organizing the transaction within the firm".² In particular, the transactions cost based arguments for integration do not explain how, given that it is difficult to write a complete contract between a buyer and seller and this creates room for opportunistic behavior, the scope for such behavior changes when one of the self interested owners becomes an equally self interested employee of the other owner. Furthermore, if vertical integration always reduces transaction costs, any buyer A and seller B which have a contractual relationship should be able to make themselves better off as follows: (i) A buys B and makes the previous owner

of B the manager of a new subsidiary; (ii) A sets a transfer price between the subsidiary and itself equal to the contract price which existed when the firms were separate enterprises, and (iii) A gives the manager of B a compensation package equal to the profit of the subsidiary. Given this, however, how can integration ever be strictly worse than non-integration, i.e. what limits the size of the firm?³ 4

A second question raised by the transactions cost based arguments concerns the definition of integration itself. In particular, what does it mean for one firm to be more integrated than another? For example, is a firm which calls its retail force "employees" more integrated than one which calls its retail force "independent but exclusive sales agents"?

Existing theories cannot answer these questions because they do not give a sufficiently clear definition of integration for its costs and benefits to be assessed. By defining integration to be the ownership of assets, not of people, we are able to evaluate the benefits and costs of integration. We will argue that if one party gets rights of control then this diminishes the rights of the other party to have control. To the extent that there are benefits of control, there will always be potential costs associated with removing control (i.e. ownership) from those who manage productive activities.

1b. What is Integration?

We define a firm to consist of those assets which it owns or over which it has control--we do not distinguish between ownership and control and virtually define ownership as the power to exercise control. Of course, control or ownership is never absolute. For example, a firm which owns a machine may not be able to sell it without the permission of the lenders for which the machine serves as collateral; more generally, a firm may give

another firm specific authority over its machines. However, ownership gives the owner all rights to use the machine which he hasn't voluntarily given away or which the government or some other party has not taken by force. We believe that this terminology is roughly consistent with standard usage.⁵

We do not distinguish between employees and outside contractors in the case where the firm provides all the tools and other assets used by the contractor. For example in insurance retailing a firm may use its own employees as commissioned agents, or use independent agents. The important difference between the two forms of retailing is that the employee-agent does not own the list of his clients, while the independent agent does own the list. If the firm owned the list and all the other important assets of the independent agents, then we would say that such a company had the same degree of integration as a company in which the retail sales force was composed of "employees". (A detailed discussion of the insurance industry may be found in Section 4.) As another example, consider vertical integration in shoe manufacturing. In the 18th century much of the manufacturing of shoes switched from the "putting out" system where the worker sewed the upper and lower halves of the shoe at home, to factory work where the factory owner's machines were used by the worker to put the shoes together.⁶ Even if workers are paid by the piece in both cases, the firm is more integrated in the latter case because it owns more of the machines used in production.

The above examples illustrate that the issue of ownership can be separated from the issue of contractual compensation. A firm may pay another firm or person by the piece or a fixed amount (salary), irrespective of the ownership of the machines. As Coase points out, the benefits of integration must surely be more than the ability to choose a new payment method. We assume that a payment method, whether it be salary compensation to an employee in the integrated company, or a price for goods to be delivered

between companies, is some function of the observable states of nature and the observable performance of the parties to the contract. We further assume that integration in itself does not make any new variable observable to both parties. Any audits which an employer can have done of his subsidiary are also feasible when the subsidiary is a separate company.⁷

It may be extremely costly to write a contract which specifies unambiguously the payments and actions of all parties in every observable state of nature. We assume that integration in itself does not change the cost of writing down a particular contractual provision.⁸ What it does change is who has control over those provisions not included in the contract. Consider, for example, a contract between a publisher and a printer for a particular number of copies of a book. If the contract has no provision for an additional print run, but the publisher receives some new information which makes it profitable for another run, then it is obvious that the right to decide whether or not to have the run belongs to the owner of the printing press. This is the simplest possible example of our assumption that the benefit of ownership is the residual right of control, i.e. the right to control all actions that have not been explicitly given away by contract.

It is worth emphasizing that since ownership provides residual rights of control, the usual argument that the feasible set can only become larger under integration fails. If firm 1 buys firm 2, the owner of firm 1 will have the power to intervene in firm 2 in all sorts of ways, some of which may be very undesirable and lead to large efficiency losses. The point is that the owner cannot commit himself to intervene selectively in his subsidiary's operations since by their very definition residual rights refer to powers that cannot be specified in advance (at least in the detail required to make them part of an enforceable contract). It follows immediately from this that integration can impose costs as well as benefits.

1c. Introduction to the Model

In order to be more specific about the costs and benefits of integration, it is necessary to set up a formal model of the relationship between two firms. This is done in Section 2. For simplicity, the relationship, which may be either vertical or lateral, is assumed to last two periods.⁹ In the first (i.e., the ex-ante) period the manager of each firm makes relationship-specific investments, while in the second (i.e., the ex-post) period some further production decisions are taken and the benefits from the relationship are realized. A basic assumption of the model is that the production decisions, represented by q , are sufficiently complex that they cannot be specified completely in an initial contract between the firms. We have in mind a situation where it is prohibitively difficult to think about, and describe unambiguously, in advance how all the potentially relevant aspects of the production allocation should be chosen as a function of the many states of the world. For example, the dimensionality of q might exceed the dimensionality of the state space, with a different dimension of q being relevant in each different state of the world. In fact, to simplify matters, we go further and suppose that no aspect of q is ex-ante contractible.¹⁰ The noncontractibility of q creates the need to allocate residual rights of control, since if it is not specified how q will be chosen, there must be some implicit or explicit default which allows some party to choose the relevant components of q in the second period. We assume that the owner of each asset has the right to control that asset in the case of a missing provision. For example, if a contract between a printing press owner and a publisher fails to specify the extent to which the printer is obliged to change the ink used in the event that a new color becomes popular, the owner of the printing plant has the power to choose the ink color, and if

possible to force the publisher to pay more for the new color. However, if the publisher owns the printing press, he "controls the missing provision" and can enforce the color of his choice.

Although q is ex-ante noncontractible, we suppose that, once the state of the world is determined, the (small number of) relevant aspects of q become clear and the parties can negotiate or recontract over these (costlessly). That is, q is, ex-post, contractible. Since the parties are assumed to have symmetric information, recontracting will always lead to an ex-post efficient allocation, whatever the initial allocation of ownership rights is.¹¹ The distribution of ex-post surplus, however, will be sensitive to ownership rights. For example, in the case of the printer and the publisher, while it may always be efficient to shift to the new ink color, the printer will extract more surplus if he owns the printing plant and can therefore refuse to change color if negotiations fail.

Through their influence on the distribution of ex-post surplus, ownership rights will affect ex-ante investment decisions. That is, although ex-post efficiency (relative to investment decisions) is guaranteed under any ownership structure, each ownership structure will lead to a (different) distortion in ex-ante investment. The ex-ante investments that we are referring to are those which are not verifiable, and hence are not reimbursable, e.g., they represent the effort the manager devotes to setting up a well functioning firm. We suppose that the parties allocate ownership rights in such a way that the ex-ante investment distortions are minimized. The implications this has for the desirability of integration are the main focus of the paper, and are analyzed in Section 3.

Since there are features of our theory which lack quantitative completeness, in Section 4 we show how the theory can be applied to a particular industry, the insurance industry. Finally, Section 5 contains

conclusions.

2. The Model

Consider two firms, 1 and 2, which are engaged in a relationship, which for simplicity we suppose lasts two periods. The firms sign a contract at date 0 and soon after make relationship-specific investments, denoted by a_1 , a_2 . At date i , some further actions q_1 , q_2 are taken and the gains from trade are realized. We write firm i 's benefit from the relationship at date i , net of investment costs, as

$$(2.1) \quad B_i(a_i, \phi_i(q_1, q_2)).$$

We are adopting the convention that all costs and benefits are measured in date i dollars. We will often interpret the relationship as a vertical one where upstream firm 2 supplies downstream firm 1 with an input. In this case $B_2 < 0$ may be a cost. However, another interpretation is that the relationship is a lateral one, e.g. between two retail stores with adjacent locations. For technical reasons, we have assumed that B_i depends on some function ϕ_i of q_1 , q_2 , and is increasing in ϕ_i .¹²

Each firm is run by a manager who we suppose is hired by the initial owner at date 0. Further, a date 0 contract specifies the ownership rights which are delegated to each manager. As we explained in the introduction, ownership rights over assets have meaning only as residual rights of control over these assets (i.e., rights which are not ex-ante contractible). The q_i 's represent rights of control which are not contractible as of date 0, but which are controllable by an owner as of date i . For example, if firm 2 is a coal mine, and firm 1 is an electricity generating plant located adjacent to firm 2, then $\phi_1(q_1, q_2)$ may represent the quality of the coal delivered.

Suppose that the boiler firm 1 installs to burn coal does not function well if the coal supplied has high ash content. Then if firm 1 owns firm 2, it can exercise its rights of control over firm 2's assets to direct that the coal should be taken from a deposit with low ash content (i.e. firm 1 chooses a subvector of q_2). On the other hand, if firm 2 owns firm 1, it can exercise its right of control over firm 1's assets to direct that the boiler should be modified to accept high ash content coal.

An alternative to ownership in this example is a full contingent contract which makes the coal mine's remuneration a function of the ash content of the coal supplied. Our assumption is that ash content is just one of many characteristics of the coal, and a contract which covers each characteristic is too costly to write. A second alternative to ownership by firm 1 is a contract which gives firm 1 the specific right to direct the areas of the mine in which coal is dug out. This would clearly be reasonable for any one particular right of control. However we assume that there are many aspects of the upstream firm's operations, each of which may be important in a different contingency, and thus the costs of assigning specific rights of control ex-ante are very much higher than the costs of assigning generalized control via rights of ownership.¹³ (See also the discussion in Footnote 20.)

We assume that although (a) the q_i 's are too complex to be included in the date 0 contract; (b) the (relevant components of the) q_i 's are contractible at date 1 once it is clear which contingency has occurred (in the above example, the contingency is one that makes ash content important). This means that the parties can write a new contract at date 1 to specify how the q_i 's should be chosen.

We turn now to the choice of the ex-ante investments a_i . We assume that a_i , although observable to the manager of firm j , is not verifiable and hence

is not ex-ante contractible (a_i might refer, for example, to the effort devoted to setting up a well functioning firm). Similarly $B_i(a_i, \phi_i)$ refers to the personal (and hence non-contractible) benefit manager i receives when he makes the investment decision a_i ex-ante, and ϕ_i is determined ex-post through the choice of q_1, q_2 .

We assume that under any ownership structure, separate managers are needed to choose a_i and a_j . The contrary assumption that integration is useful because it substitutes one manager for two managers has been advanced by Aron (1985), Mann and Wissink (1984).

We summarize our assumptions so far as follows:

- (2.2) The manager of firm i chooses a_i and receives the net benefit B_i . This benefit is personal to manager i , i.e., it does not appear in the accounts.¹⁴
- (2.3) The action a_i is observable to firm j 's manager but is not verifiable. It follows that payments from firm i to firm j cannot be made conditional on the a 's.¹⁵
- (2.4) q_i is ex-ante noncontractible, but ex-post contractible. The owner of firm i has the right and power to choose q_i at date 1. No special skills are required to choose q_i .¹⁶

We make the further assumption that there is a competitive market at date 0 in identical potential trading partners (who are owners of assets prior to the date 0 contract which re-allocates ownership) and identical potential managers for type 1 firms and for type 2 firms. This market determines the ex-ante division of the surplus between the initial owners of firms 1 and 2 and the managers they hire. Further, we suppose that all ownership rights of control are delegated to the managers of firms 1 and 2. That is, the initial owners of assets prior to date 0 can be ignored once the date 0 contract is signed. Thus, if firm i owns firm j the owner-manager of

firm i will choose q_j . We will sometimes just state that the owner of firm j chooses q_j . We let the market clearing payoffs of firm 2's initial owner and of the managers of firms 1 and 2 be $\bar{V}_2, \bar{U}_1, \bar{U}_2$ respectively. An optimal contract can then be thought of as being chosen by firm 1's initial owner to maximize his payoff subject to these constraints. The timing of events is as follows:

Date 0 Contract signed which allocates ownership rights among managers. Then a_1 and a_2 are chosen simultaneously and noncooperatively by managers 1 and 2.

Date 1 If the original contract is not renegotiated, the owner of firm i (i.e., the manager who has been given the ownership rights of control) chooses q_i . The choices of different owners and managers are made simultaneously and noncooperatively. However, the contract may be renegotiated. Then B_1, B_2 are realized.

One strong assumption that we have made is that no date 1 variables are contractible at date 0. In a vertical relationship we might expect the parties at least to be able to contract on the date 1 level of trade. For instance, in the case of the coal mine, the ex-ante contract might be a "requirements" contract with a specified minimum and maximum delivery (i.e., the electricity generating plant chooses the level of coal it wants and the mine must deliver). One interpretation of our model is that we are taking this transfer of goods contract as given, and q_1, q_2 then represent noncontractible choices which affect each party's benefit from the completion of trade relative to this contract. Of course, a problem with this interpretation is that we would expect the form of the transfer of goods contract to be sensitive to the structure of ownership. We should emphasize that we believe that the flavor of our results concerning the role of

ownership will continue to hold when this sensitivity is modelled.¹⁷

In a nonvertical relationship, the assumption that there are no contractibles is more palatable (nothing concrete may be traded). In the case of physically adjacent retail stores, for example, the q_i might represent ex-post choices regarding advertising, product lines, hours the stores are open, etc. The number of customers visiting store i , ϕ_i , will depend on the choice of q_j of store j , as well as on q_i .

In the next section, we analyze the optimal contract between firms 1 and 2, focusing particularly on how ownership rights should be allocated.

3. Analysis of the Optimal Contract, including the Allocation of Ownership Rights.

Given that there is no uncertainty and that monetary transfers are available, an optimal contract must maximize the total ex-ante net benefits or surplus of the two managers,

$$(3.1) \quad B_1(a_1, \phi_1(q_1, q_2)) + B_2(a_2, \phi_2(q_1, q_2)).$$

It is useful to consider as a benchmark the first-best where contrary to our assumptions above a_1, a_2 are verifiable and q_1, q_2 are ex-ante contractible.

Definition. Let $a_1^*, a_2^*, q_1^*, q_2^*$ be the (assumed unique) maximizers of $B_1 + B_2$ subject to $a_i \in A_i, q_i \in Q_i$ ($i = 1, 2$).

The first-best contract would state that manager i must choose a_i^* at date 0 and q_i^* at date 1 (if not he must pay manager j a large penalty), and would specify monetary transfers so as to guarantee that the payoffs of firm 1's manager, firm 2's manager and firm 2's initial owner are $\bar{U}_1, \bar{U}_2, \bar{V}_2$ respectively.

Given that q_1, q_2 are not ex-ante contractible, the first-best cannot generally be achieved, as we shall now see.¹⁸ Under our simplifying assumption that no date 1 variables are contractible as of date 0, the only element of the contract is the allocation of ownership rights. That is, a contract will consist simply of an allocation of ownership rights and transfer payments between the initial owners of the firms and their managers. There are three interesting cases to consider. The first is where the firms remain separately owned or nonintegrated, i.e., manager 1 owns and controls firm 1's assets and manager 2 owns and control firm 2's assets. The second is where firm 1 owns firm 2, i.e., manager 1 owns and controls the assets of both firms (we call this firm 1 control). And the third is where firm 2 owns firm 1 (we call this firm 2 control). There is a fourth case where manager 1 owns firm 2's assets and manager 2 owns firm 1's assets. This case, which is the mirror image of case 1, appears less interesting than the others since it seems likely in practice to give a much lower level of surplus than case 1. We therefore ignore it in what follows.¹⁹

Case 1: Nonintegration -- Manager 1 chooses q_1 and Manager 2 chooses q_2 .

It is useful to start at date 1 and work backwards. At date 1, a_1, a_2 are predetermined and the only question concerns the choices of q_1, q_2 . If no further negotiation takes place, q_1, q_2 will be chosen simultaneously and noncooperatively by managers 1, 2 to maximize $\phi_1(q_1, q_2)$ and $\phi_2(q_1, q_2)$, respectively. We assume:

(3.2) There exists a unique pair (\hat{q}_1, \hat{q}_2) satisfying: $q_1 = \hat{q}_1$ maximizes $\phi_1(q_1, \hat{q}_2)$ subject to $q_1 \in Q_1$, and $q_2 = \hat{q}_2$ maximizes $\phi_2(\hat{q}_1, q_2)$ subject to $q_2 \in Q_2$.

In other words the game where manager 1 maximizes ϕ_1 has a unique Nash equilibrium.

Of course, given a_1, a_2 the noncooperative equilibrium (\hat{q}_1, \hat{q}_2) is unlikely to be ex-post efficient in the sense of maximizing

$$(3.3) \quad B_1(a_1, \phi_1(q_1, q_2)) + B_2(a_2, \phi_2(q_1, q_2)).$$

Therefore the two parties can gain from writing a new contract at date 1 which specifies that $q_1 = q_1(a_1, a_2)$, $q_2 = q_2(a_1, a_2)$, where these are the maximizers of (3.3) (if there are several maximizers, choose any pair). We will use the notation $\hat{q} \equiv (\hat{q}_1, \hat{q}_2)$ and $q(a_1, a_2) \equiv (q_1(a_1, a_2), q_2(a_1, a_2))$. The new contract is feasible since q_1, q_2 are ex-post contractible. It will specify a transfer price p which serves to allocate the gains from negotiation. Because we don't want to get into the details of contract renegotiation, we shall simply assume that the parties split the increase in total surplus 50:50, i.e., the transfer price p satisfies

$$(3.4) \quad B_1(a_1, \phi_1(q(a_1, a_2))) - p = B_1(a_1, \phi_1(\hat{q})) + (1/2)* \\ [B_1(a_1, \phi_1(q(a_1, a_2))) + B_2(a_2, \phi_2(q(a_1, a_2))) - B_1(a_1, \phi_1(\hat{q})) - B_2(a_2, \phi_2(\hat{q}))] \\ \equiv \psi_1(a_1, a_2, \hat{q})$$

$$(3.5) \quad p + B_2(a_2, \phi_2(q(a_1, a_2))) = B_2(a_2, \phi_2(\hat{q})) + (1/2)* \\ [B_1(a_1, \phi_1(q(a_1, a_2))) + B_2(a_2, \phi_2(q(a_1, a_2))) - B_1(a_1, \phi_1(\hat{q})) - B_2(a_2, \phi_2(\hat{q}))] \\ \equiv \psi_2(a_1, a_2, \hat{q}).$$

This is in fact the Nash bargaining solution. Note that most bargaining solutions will yield an ex-post Pareto optimal outcome given our assumptions that the parties have the same information and that bargaining, i.e.,

contract renegotiation, is costless (see e.g. Rubinstein (1982)). It should be clear from what follows that our results will generalize to many other divisions of the surplus.

We assume that a_1 and a_2 are chosen noncooperatively by each agent at date 0. A Nash equilibrium in date 0 investments is a pair $(\tilde{a}_1, \tilde{a}_2) \in A_1 \times A_2$ such that

$$(3.6) \quad \psi_1(\tilde{a}_1, \tilde{a}_2, \hat{q}) > \psi_1(a_1, \tilde{a}_2, \hat{q}) \text{ for all } a_1 \in A_1,$$

$$(3.7) \quad \psi_2(\tilde{a}_1, \tilde{a}_2, \hat{q}) > \psi_2(\tilde{a}_1, a_2, \hat{q}) \text{ for all } a_2 \in A_2.$$

The total ex-ante surplus from the relationship in this equilibrium is then

$$(3.8) \quad B_1(\tilde{a}_1, \phi_1(q(\tilde{a}_1, \tilde{a}_2))) + B_2(\tilde{a}_2, \phi_2(q(\tilde{a}_1, \tilde{a}_2))).$$

A sufficient condition for the existence of a Nash equilibrium in date 0 investments is that A_i is convex and ψ_i is concave in a_i ($i = 1, 2$).

We have seen how to compute total surplus in the case of nonintegration.²⁰ This will generally be less than the first-best level of surplus since the ex-ante investments will be inefficient. To see this, note that that part of ψ_i which depends on a_i is given by

$$(3.9) \quad 1/2 B_i(a_i, \phi_i(\hat{q})) + 1/2 B_i(a_i, \phi_i(q(a_1, a_2))) \\ + 1/2 B_j(a_j, \phi_j(q(a_1, a_2))).$$

Hence the first order conditions for a Nash equilibrium are

$$(3.10) \quad 1/2 \frac{\partial B_i}{\partial a_i} (a_i, \phi_i(\hat{q})) + 1/2 \frac{\partial B_i}{\partial a_i} (a_i, \phi_i(q(a_1, a_2))) = 0 \quad (i = 1, 2),$$

where we are using the envelope theorem to eliminate remaining terms involving the ex-post efficient $q(a_1, a_2)$. This contrasts with the first-order conditions for the solution of (3.1),

$$(3.11) \quad \frac{\partial B_i}{\partial a_i} (a_i, \phi_i(q(a_1, a_2))) = 0 \quad (i = 1, 2).$$

The inefficiency arises, then, because manager i puts 50% weight on the noncooperative outcome \hat{q} , which is generally ex-post inefficient, instead of all the weight on the cooperative outcome, which is ex-post efficient. To the extent that the marginal and total benefits of a_i move in the same direction, this can substantially distort the choice of a_i . It is worth emphasizing that in this model all the inefficiency is due to the wrong choice of ex-ante investment levels. The assumption of costless renegotiation ensures that there is no ex-post inefficiency and so if ex-ante investments (more precisely nonreimbursable ones, as emphasized in the introduction) are unimportant, the first-best can always be achieved.²¹

Case 2: Firm 1 control, i.e. Firm 1 owns Firm 2 and so Manager 1 chooses q_1 and q_2

At date 1, manager 1 will now choose (q_1, q_2) to maximize ϕ_1 . We assume that there is a unique pair (\bar{q}_1, \bar{q}_2) such that:

$$(3.12) \quad (\bar{q}_1, \bar{q}_2) \text{ solves: Maximize } \phi_1(q_1, q_2) \text{ subject to } (q_1, q_2) \in Q_1 \times Q_2.$$

The pair (\bar{q}_1, \bar{q}_2) will generally not be ex-post Pareto optimal and so recontracting at date 1 will lead to the pair $q_1(a_1, a_2), q_2(a_1, a_2)$, as in the case of nonintegration. Given the 50% sharing rule, manager i's final payoff is as in (3.4)-(3.5) with (\bar{q}_1, \bar{q}_2) replacing (\hat{q}_1, \hat{q}_2) . The date 0 Nash equilibrium in investments and the final level of surplus are also defined as in the case of nonintegration (see (3.6)-(3.8)), again with (\bar{q}_1, \bar{q}_2) replacing (\hat{q}_1, \hat{q}_2) . Firm 1 control will generally lead to inefficient ex-ante investments, since $(\bar{q}_1, \bar{q}_2) \neq (q_1(a_1, a_2), q_2(a_1, a_2))$ (See (3.10)-(3.11)).

Case 3: Firm 2 control, i.e. Firm 2 owns Firm 1, and so Manager 2 chooses q_1 and q_2

Now, at date 1, manager 2 will choose (q_1, q_2) to maximize ϕ_2 . We assume that there is a unique pair $(\underline{q}_1, \underline{q}_2)$ such that:

$$(3.13) \quad (\underline{q}_1, \underline{q}_2) \text{ solves: Maximize } \phi_2(q_1, q_2) \text{ subject to } (q_1, q_2) \in Q_1 \times Q_2.$$

This case is the same as the previous one with $(\underline{q}_1, \underline{q}_2)$ replacing (\bar{q}_1, \bar{q}_2) everywhere. Again ex-ante investments will generally be inefficient.

We consider now which of the above three cases represents the optimal ownership structure. It is worth emphasizing that we are talking about final ownership after the date 0 contract is signed, and that any ownership structure which does not maximize total surplus can be Pareto dominated by one that does, with an appropriate sidepayment. We saw in (3.10) that the inefficiency in the a's is due to the fact that manager i puts 50% weight on

the noncooperative solution (which equals (\hat{q}_1, \hat{q}_2) under nonintegration, (\bar{q}_1, \bar{q}_2) under firm 1 control and (q_1, q_2) under firm 2 control) instead of 100% on the cooperative solution $(q_1(a_1, a_2), q_2(a_1, a_2))$. It is clear therefore that which ownership structure is optimal depends in some sense on which of the pairs (\hat{q}_1, \hat{q}_2) , (\bar{q}_1, \bar{q}_2) , (q_1, q_2) is closest to $(q_1(a_1, a_2), q_2(a_1, a_2))$. In some special situations, one of the pairs might actually equal $(q_1(a_1, a_2), q_2(a_1, a_2))$, in which case the associated ownership structure will achieve the first-best. For example, suppose $\phi_1(q_1, q_2) = \phi(q_1, q_2) \equiv -\phi_2(q_1, q_2)$ represents "quality" of firm 2's supply (where firm 1 likes high quality and firm 2 likes low quality). Assume that there are just three quality levels, high, medium and low, given by $\phi(\bar{q}_1, \bar{q}_2)$, $\phi(\hat{q}_1, \hat{q}_2)$ and $\phi(q_1, q_2)$ respectively. Then it might be that high quality, say, is always ex-post efficient whatever a_1, a_2 are, in which case firm 1 control achieves the first-best.

Of course if the q 's and a 's are continuous variables, and B_1, B_2 are differentiable functions, the ex-post efficient solution will almost always be sensitive to the a 's, in which case no one of the noncooperative pairs (which are constants) can be equivalent to it. However, there are some important special cases in which the equivalence holds approximately, and we now investigate these. We consider first a situation in which ϕ_i depends primarily on q_i , and show that nonintegration, generally, strictly dominates any form of integration. Thus, suppose

$$(3.14) \quad \begin{aligned} \phi_1(q_1, q_2) &= \alpha_1(q_1) + \epsilon_1 \beta_1(q_2), \\ \phi_2(q_1, q_2) &= \alpha_2(q_2) + \epsilon_2 \beta_2(q_1), \end{aligned}$$

where $\epsilon_1, \epsilon_2 > 0$ are small. In the nonintegrated solution, at date 1,

manager 1 chooses $q_1 = \hat{q}_1$ to maximize $\alpha_1(q_1)$ and manager 2 chooses $q_2 = \hat{q}_2$ to maximize $\alpha_2(q_2)$. At date 0, by (3.9), manager 1 chooses $a_i = \hat{a}_i$ to maximize

$$(3.15) \quad 1/2 B_i(a_i, \alpha_i(\hat{q}_i)) + \epsilon_i \beta_i(\hat{q}_i) + 1/2 B_i[a_i, \alpha_i(q_i(a_1, a_2)) + \epsilon_i \beta_i(q_j(a_1, a_2))] + 1/2 B_j(a_j, \alpha_j(q_j(a_1, a_2)) + \epsilon_j \beta_j(q_j(a_1, a_2))).$$

It is clear that in the limit $\epsilon_1 = \epsilon_2 = 0$, \hat{q}_i is ex-post efficient, i.e., $(q_1(a_1, a_2), q_2(a_1, a_2)) = (\hat{q}_1, \hat{q}_2)$ for all a_1, a_2 . Hence (3.15) implies that in the limit $a_i = \hat{a}_i$ maximizes $B_i(a_i, \alpha_i(\hat{q}_i))$, and so $a_1 = \hat{a}_1$ and $a_2 = \hat{a}_2$ are ex-ante efficient. Therefore, by continuity, for ϵ_1, ϵ_2 small, nonintegration achieves approximately the first-best.

Firm 1 or firm 2 control, in contrast, may lead to great inefficiencies. Under firm 1 control, manager 1 chooses $q_1 = \hat{q}_1$ to maximize $\alpha_1(q_1)$ and $q_2 = \bar{q}_2$, say, to maximize $\beta_1(q_2)$. This means that in the limit $\epsilon_1 = \epsilon_2 = 0$, a_1 is chosen to maximize

$$(3.16) \quad 1/2 B_1(a_1, \alpha_1(\hat{q}_1)) + 1/2 B_1(a_1, \alpha_1(\hat{q}_1)) = B_1(a_1, \alpha_1(\hat{q}_1)),$$

while a_2 is chosen to maximize

$$(3.17) \quad 1/2 B_2(a_2, \alpha_2(\bar{q}_2)) + 1/2 B_2(a_2, \alpha_2(\hat{q}_2)).$$

The choice of a_1 is efficient, but that of a_2 is generally not, since \bar{q}_2 may be very different from \hat{q}_2 . Similarly under firm 2 control, a_2 is chosen efficiently while a_1 is not. Thus, under (3.14), where firm 1 has negligible benefits from residual rights of control over firm j, i's ownership of j causes the use of these rights to distort manager j's ex-ante expenditures in

setting up firm j .

We have seen that if ϕ_1 depends significantly on q_1 , but hardly at all on q_2 , nonintegration is optimal. We show next that if, say, firm 2's benefit does not depend very much on the allocation of residual rights of control, but firm 1's benefit is sensitive to these, firm 1 control is optimal. Suppose

$$(3.18) \quad \phi_2(q_1, q_2) = \gamma_2 + \epsilon_2 \delta_2(q_1, q_2),$$

where γ_2 is constant. It is clear that when ϵ_2 is small, firm 2 cares very little about q_1, q_2 and so if firm 1 has control over these it will make an approximately ex-post efficient choice. This in turn will lead to approximately ex-ante efficient choices of a_1, a_2 . Hence firm 1 control achieves approximately the first-best. Nonintegration or firm 2 control, on the other hand, will lead to inefficient choices of the q 's in the noncooperative equilibrium, and hence inefficient choices of the a 's. On the other hand, if

$$(3.19) \quad \phi_1(q_1, q_2) = \gamma_1 + \epsilon_1 \delta_1(q_1, q_2),$$

of course, we get exactly the opposite result -- now firm 2 control is optimal.

We summarize these results in the following proposition.

Proposition 1

- (1) If (3.14) holds, it is the case that for $\epsilon_1, \epsilon_2 > 0$ small, nonintegration yields approximately the first-best, while firm 1 and firm 2 control generally do not.
- (2) If (3.18) holds, then for $\epsilon_2 > 0$ small, firm 1 control yields

approximately the first-best, while nonintegration and firm 2 control generally do not.

- (3) If (3.19) holds, then for $\varepsilon_1 > 0$ small, firm 2 control yields approximately the first-best, while nonintegration and firm 1 control generally do not.

Proposition 1 says that if the noncontractibles q_2 have a small effect on firm j 's benefit B_j , it is efficient for firm 1 to control them. Note that there will be a significant loss in surplus from 1 not controlling them if $\frac{\partial}{\partial q_2} \frac{\partial B_1}{\partial a_1}(a_1, \phi_1(q))$ is large, i.e. if the marginal product of a_1 is sensitive to q_2 (if $B_1(a_1, \phi_1(q)) = f_1(a_1) + \phi_1(q)$, say, there is no loss at all). To put it another way, Proposition 1 only tells us that a particular ownership structure is optimal -- it doesn't quantify the costs of being at a suboptimal structure. However, by choosing $\frac{\partial}{\partial q} \frac{\partial B_1}{\partial a_1}$ appropriately, we may easily construct examples where this loss is extremely large.

Proposition 1 deals with the case where the noncontractibles are important to one party, but not to another. In general, both parties will care about the noncontractibles and, as a result, each ownership structure will lead to a distortion in ex-ante investments. The crucial question then is which ownership structure leads to the least significant distortion. In order to analyze this, we make some further simplifying assumptions. Recall that $B_1(a_1, \phi_1(q_1, q_2))$ is increasing in ϕ_1 . We now suppose

- (3.20) A_1, A_2 are convex subsets of R (so that investment decisions are scalars),

$$(3.21) \quad \frac{\partial^2 B_i}{\partial q_i \partial a_i} (a_i, \phi_i(q_1, q_2)) > 0 \quad (\text{marginal benefit is high when average benefit is high}),$$

$$(3.22) \quad \frac{\partial^2 B_i}{\partial a_i^2} (a_i, \phi_i(q_1, q_2)) < 0,$$

(3.23) The maximizers $q_1(a_1, a_2)$, $q_2(a_1, a_2)$ of (3.3) are independent of (a_1, a_2) in the relevant range; we write them as q_1^* , q_2^* .

The fourth of these is a strong assumption. It says that the ex-post efficient choice of the noncontractibles is independent of ex-ante actions. The assumption is not reasonable if B_1 , B_2 are differentiable functions of the q 's. However, it may hold if the q 's take on only discrete values. In any case the argument that follows can be generalized to the case where (3.23) is violated, at the cost of additional complexity.

The first-order conditions for the choice of ex-ante investment by the managers are given by (3.10) where we replace $\phi_i(\hat{q})$ by $\tilde{\phi}_i$. Here $(\tilde{\phi}_1, \tilde{\phi}_2) = (\phi_1(\tilde{q}_1, \tilde{q}_2), \phi_2(\tilde{q}_1, \tilde{q}_2))$, where $(\tilde{q}_1, \tilde{q}_2)$ are the choice of the q 's in the event that renegotiation does not occur (i.e. $(\tilde{q}_1, \tilde{q}_2) = (\hat{q}_1, \hat{q}_2)$ under nonintegration, $(\tilde{q}_1, \tilde{q}_2) = (\bar{q}_1, \bar{q}_2)$ under firm 1 control and $(\tilde{q}_1, \tilde{q}_2) = (\underline{q}_1, \underline{q}_2)$ under firm 2 control). Given (3.23), (3.10) therefore becomes

$$(3.24) \quad 1/2 \frac{\partial B_i}{\partial a_i} (\bar{a}_i, \tilde{\phi}_i) + 1/2 \frac{\partial B_i}{\partial a_i} (\bar{a}_i, \phi_i^*) = 0.$$

The first-best investment decisions, in contrast, are characterized by

$$(3.25) \quad \frac{\partial \phi_1}{\partial a_1} (a_1^*, \phi_1^*) = 0, \text{ where } \phi_1^* = \phi_1(q_1^*, q_2^*).$$

Proposition 1 dealt with the case where one of the ownership structures gave rise to a $(\tilde{q}_1, \tilde{q}_2)$ very close to (q_1^*, q_2^*) . Our concern now, however, is with cases where (\hat{q}_1, \hat{q}_2) , (\bar{q}_1, \bar{q}_2) and $(\underline{q}_1, \underline{q}_2)$ are all quite "far" from (q_1^*, q_2^*) . We may illustrate the situation in the following diagram.

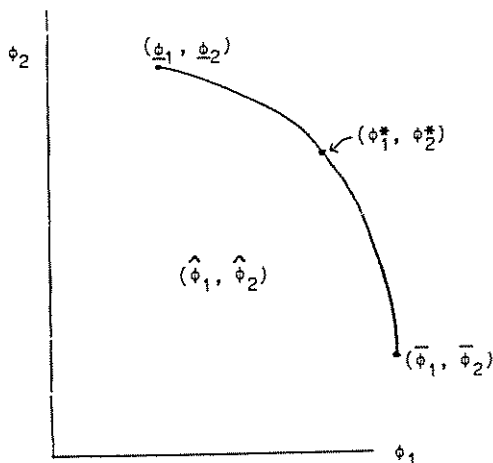


FIGURE 1

The curve represents the efficient $\phi_2 - \phi_1$ combinations, where we write $\underline{\phi}_1 = \phi_1(\underline{q}_1, \underline{q}_2)$, $\hat{\phi}_1 = \phi_1(\hat{q}_1, \hat{q}_2)$, $\bar{\phi}_1 = \phi_1(\bar{q}_1, \bar{q}_2)$. We have drawn it to be continuous, but it could equally well be a set of discrete points. If firm 1 or firm 2 has control, $(\tilde{\phi}_1, \tilde{\phi}_2)$ will lie on the efficiency frontier since one party controls q_1 and q_2 . Under nonintegration, in contrast, the noncooperative outcome $(\hat{\phi}_1, \hat{\phi}_2)$ may well be highly inefficient due to the uncoordinated choice of (q_1, q_2) .

We may now determine the nature of the investment distortions corresponding to the different ownership structures. Using the assumption that $\frac{\partial^2 B_i}{\partial a_i \partial \phi_i} > 0$, we see that the left-hand side of (3.24) is positive (resp. negative) at $a_i = a_i^*$ if $\tilde{\phi}_i > \phi_i^*$ (resp. $< \phi_i^*$). Hence since $\frac{\partial^2 B}{\partial a_i^2} < 0$, $\tilde{a}_i > a_i^*$ as $\tilde{\phi}_i > \phi_i^*$.

Now $\bar{\phi}_1 > \phi_1^*$, $\bar{\phi}_2 < \phi_2^*$ (see Figure 1), and so $\bar{a}_1 > a_1^*$, $\bar{a}_2 < a_2^*$, i.e. under firm 1 control, firm 1 overinvests relative to the first-best and firm 2 underinvests. On the other hand, under firm 2 control, firm 2 overinvests relative to the first-best and firm 1 underinvests. Nonintegration is more complicated since the nature of the distortion depends on the relationship of $(\hat{\phi}_1, \hat{\phi}_2)$ to (ϕ_1^*, ϕ_2^*) . However, if the outcome $(\hat{\phi}_1, \hat{\phi}_2)$ is highly inefficient -- which seems plausible in a number of cases -- it will quite likely lie to the south-west of (ϕ_1^*, ϕ_2^*) , i.e. $\hat{\phi}_1 < \phi_1^*$, $\hat{\phi}_2 < \phi_2^*$. Hence in this case $\hat{a}_1 < a_1^*$, $\hat{a}_2 < a_2^*$, that is, nonintegration leads to underinvestment by both firms.

The trade-offs should now be fairly clear. Firm 1 control will be desirable when firm 1's ex-ante investment is much more important than firm 2's (so that firm 2's underinvestment under firm 1 control is relatively unimportant) and when the distortion due to overinvestment by firm 1 is less severe than that which would arise from underinvestment, as in e.g., the nonintegrated solution. Firm 2 control will be desirable when firm 2's investment decision is much more important than firm 1's and when overinvestment by firm 2 is a less severe problem than underinvestment. Finally, nonintegration is desirable if a_1, a_2 are both "important" in some sense, so that it is preferable to have both of them at a medium level, than to have one very high and the other very low as under integration (note that

if $\hat{\phi}_2 > \bar{\phi}_2$, $\hat{\phi}_1 > \underline{\phi}_1$, then $\hat{a}_2 > \bar{a}_2$, $\hat{a}_1 > \underline{a}_1$, i.e. firm 1's investment under nonintegration is greater than under firm j control).

As an example, suppose we have a pure vertical relationship, that is, $B_2 = B_2(a_2, \phi_2(q_2))$ doesn't depend on q_1 , and suppose that $a_2 \equiv 0$. Then nonintegration and firm 2 control both lead to the same value of ϕ_2 , i.e. $\hat{\phi}_2 = \underline{\phi}_2$. It follows that $\hat{\phi}_1 < \phi_1^*$ (see Figure 1), and hence $\hat{a}_1 < a_1^*$, that is, a_1 is too low under nonintegration. This may lead to considerable inefficiency. One way to raise a_1 is to give firm 1 control. However, under firm 1 control, a_1 overshoots a_1^* and so whether inefficiency is reduced depends on whether the distortion due to overinvestment is smaller than that due to underinvestment.

We see that, even in the case where firm 2's investment is irrelevant, we cannot conclude in general that firm 1 control is desirable. This is because firm 1 may overinvest given that it ignores the cost imposed on firm 2's manager from its choice of q . This effect results from our assumption that the benefits B_1 accrue to manager 1 and are inalienable from him. A slight variant of our model, however, is where B_2 , say, is perfectly alienable in the sense that there is a way for manager 1 to capture B_2 as long as he controls firm 2's assets. Let $B_2 = f_2(a_2, \phi_2(q_1, q_2)) - C_2(a_2)$ where f_2 is a date 1 variable benefit and C_2 is a sunk investment cost. We have in mind a situation where firm 2's manager can be costlessly replaced by another (equally skilled) manager at date 1 (training is unimportant) and this new manager can be offered a contract which pays him $-f_2$ (for simplicity, we suppose that the opportunity costs of both the old and new manager are zero). This means that if firm 1's manager has control his benefit becomes $B_1 + f_2$, i.e. the benefit f_2 is transferred. The fact that manager 2 will not receive f_2 will, of course, have a very adverse effect on

his date 0 incentives, but, in the case where manager 2's date 0 investments are relatively unimportant, it is intuitively clear that firm 1 control will be optimal. The desirability of integration when a_2 is not important is shown in the following proposition.

Proposition 2

Suppose that $B_2 = f_2(a_2, \phi_2(q_1, q_2)) - C_2(a_2)$, and $\frac{\partial f_2}{\partial a_2} \equiv 0$. Suppose further that the ex-post benefits f_2 can be acquired by the owner of firm 1 via its right of control q_2 and that $f_2 > 0$. Then firm 1 control yields the first-best while nonintegration and firm 2 control generally do not.²²

Proof. If firm 1 has control, it will, in the noncooperative solution, face the benefit function $B_1(a_1, \phi_1(q_1, q_2)) + f_2(a_2, \phi_2(q_1, q_2))$, since when $f_2 > 0$ firm 1 takes f_2 from manager 2. Further manager 1 receives $-C_2(a_2)$ overall (plus some transfer payment) and chooses a_2 to maximize this. Since f_2 doesn't depend on a_2 , this leads to the socially correct investment decision by firm 2. Also firm 1 maximizes the social objective function at both dates and hence also chooses the socially correct investment decision.

Q.E.D.

We conclude this section with two remarks:

- (1) One simplifying assumption we have made is that when firm i owns firm j , it can control all the residual rights, q_j . In reality, a subvector q_j^i of q_j may always remain under the control of manager j , say because manager j is the only person with the ability to control this particular aspect of the firm's operation. Our analysis can easily be generalized to this case. The main difference is that, even under integration, the pre-renegotiation

choice of (q_i, q_j) will involve a lack of coordination by firms 1 and 2. Note that ownership rights are likely to be less important the more components of q_j remain under manager j 's control. For example, suppose firm j is a law firm with a single lawyer and firm 1 is firm j 's single client. Then if the client buys the law firm he may no more be able to get the lawyer to provide a special service than if the lawyer was in private practice. That is, the value of controlling firm j 's assets may be very small in this case.

- (2) We have assumed that a_i affects only B_i . A generalization of the model is to the case where a_i affects also B_j ; e.g., a_2 , as well as reducing firm 2's variable cost at date 1, may also influence the quality of firm 2's supplies. Both Propositions 1 and 2 can be extended to this case.

This completes our analysis of the costs and benefits of integration. In the next section, we apply this analysis to the case of the insurance industry.

4. An Application

The main results of section 3 may be summarized as follows (for the case where investment decisions are scalars and (3.20)-(3.23) hold):

Firm i will own Firm j

- if 11. (a) values of a_i below a_i^* , the first-best level of firm i 's investment, cause a considerable loss of surplus while values of a_i above a_i^* do not (alternatively, a_i is very sensitive to marginal incentives when $a_i < a_i^*$ but not so sensitive when $a_i > a_i^*$); and (b)

a_j does not affect surplus very much relative to a_1 (alternatively, a_j is not very sensitive to marginal incentives);
or I2. q_1, q_2 are very important to firm 1, but not to firm j.

No integration will occur

if N1. Very low values of either a_1 or a_2 cause a considerable loss of surplus, while medium values of a_1, a_2 do not;
or N2. q_1 is very important to firm 1 but not to firm 2, and q_2 is very important to firm 2, but not to firm 1.

We now apply these results to the insurance industry. Any real industry is, of course, far more complex than our model. One important difference is that in practice some variables will be contractible at date 0. We will therefore interpret our model with considerable latitude in what follows.

In the insurance industry some firms have a sales force which sell primarily their own company's products.²³ These companies are called direct writers, and their sales force may include employees (with virtually no ownership rights to office equipment) or agents who are independent contractors (who may own their office equipment, and the building housing their agency office). Aside from the ownership of some office equipment there are no major differences between employees and nonemployees; both are on commissions and the differences in commission between the two types just reflect in an obvious way the differences in who bears office expenditures. However in all cases direct writers are distinguished by the fact that the insurance company and not the agent owns the list of policyholders. Ownership of the list of policyholders entitles the insurance company to sell insurance to the policy holder if the agent terminates the relationship with

the insurance company. Insurance company ownership of the list also means that the agent has no right to renew the insurance policy with a different company; he cannot leave the company and take his clients with him.

Insurance companies which are not direct writers sell insurance through independent agents and brokers (who we will lump together as independent agents in distinction to the "captive" agents discussed above). The independent agents are distinguished by the fact that they, rather than the insurance company, own the list. An independent agent can sell any insurance company's product to his client. If the agent terminates his relationship with a particular insurance company that company has no right to solicit the business from the agent's list. Even without termination of the relationship, if the agent thinks that a client would be happier with the insurance of another company, the agent can encourage the client to change companies.

An insurance company has a number of expenditures which, given characteristics of the (contractible) commission structure to be explained below, can create ex-post surplus between the insurance company and its agents and/or brokers. These expenditures include training of agents, client list building expenditures (such as advertising), product development and policy holder services. An insurance agent can have similar expenditures. To the extent that the efforts of the parties in generating these expenditures are not verifiable, they cannot be reimbursed directly without the creation of moral hazards. Instead the contract between the parties will specify payments as a function of observable, e.g., commissions to the agent for policies produced for the insurance company.

We will use our framework to analyze the determinants of who owns the list of policyholders. (We assume that the agent does not want to own the whole insurance company.) Note that since there is only one asset here

(namely the client list), the choice is, in the language of our model, between firm 1 control and firm 2 control. Nonintegration has no meaning. To proceed, we must provide a model of the insurance industry. Space limitations permit only the simplest model. We assume that the agent devotes effort which is not verifiable to acquiring and keeping clients. The greater this effort, the more likely it is that a typical client will renew his insurance in the future, i.e. that he will be persistent. Examples of such effort are the care with which the agent tailors the initial policy to the client's needs and the efficiency with which he deals with a claim once the policy is in force. Note that it is important for what follows that this effort yields dividends in the future, not just at the time when it is incurred; e.g., a claim dealt with speedily today is likely to encourage the client to renew next year and the year after. To simplify the exposition, we assume that the agent can either "work" and produce only persistent clients or "not work" and produce only temporary clients, and that, if effort were verifiable, the insurance company would be prepared to compensate the agent for the extra effort of delivering persistent clients. An immediate implication of these assumptions is that if the agent is paid a commission for the initial acquisition of the client and no later commission as a function of the persistence of the client, then the agent will deliver only temporary clients, and this is inefficient relative to the first-best. (Note that similar incentive problems will arise if some clients are naturally more persistent than others and the agent must devote extra effort to finding the more persistent clients; the analysis below applies also to this case.)

In order to induce the agent to produce persistent clients, the commission structure must be back-loaded to reward the agent's initial effort costs. Specifically, the agent must get an initial commission somewhat lower than the acquisition cost of a client, but get renewal commissions which are

in excess of the agent's servicing costs associated with obtaining the renewal, i.e., the renewal premium must have some component of a reward for the effort of delivering persistent clients.

The back-loading of commissions, in and of itself, has no particular implication for who owns the list, unless there are noncontractibles. We will be concerned with two kinds of non-contractibles which could interfere with the above commission structure: (1) Non-contractibles which can hurt the agent if the company owns the list, and (2) non-contractibles which can hurt the company if the agent owns the list.

Important examples of (1) have to do with the fact that the insurance company can make the product it is selling less competitive (e.g. by raising its price or lowering the quality of its services relative to other insurance companies) and hence make the client more likely to want to switch insurance companies. For example, an insurance company can decide that it does not want to insure automobiles in a particular region, so it raises its prices or lowers the quality of its services in that region, or the insurance company can change the type and quality of its advertising which affects the likelihood that a client will renew his policy. It is very difficult for an insurance company to write a contract with agents which specifies all the relevant ways in which, and contingencies under which, the company will support the competitive position of its particular products, i.e., these actions really are non-contractible. Such non-contractibles can seriously distort the agent's effort decision if the firm owns the list and the commission structure is back-loaded. In particular once the commission structure is back-loaded, the agent will lose the renewal premium, and thus be unable to recover his cost of delivering persistent clients when the company takes acts which lead the client to want to switch insurance

companies. On the other hand if the agent owns the list then the back-loading of the commission structure does not distort the agent's action because the agent can switch the client to another company when the first company is a bad match for the client. In the notation of our model this is a case where the "q" of the insurance company is very important for the agent's ex-ante effort.

There are also type 2 non-contractibles, i.e. non-contractibles which can hurt the company if the agent owns the list. First, if the company develops an unanticipated new insurance product, then the agent's clients cannot be solicited without the agent's permission when the agent owns the list. Second, when the agent owns the list, he can encourage his clients to switch to other companies if this seems advantageous (to him or to them). In some states of the world, such a switch may be efficient, but in other states it will merely increase the agent's profits at the expense of those of the company. The ability of the agent to switch customers in this way will distort the company's ex-ante investments. Important examples of such investments are the ex-ante expenses which the company bears in building the list, such as advertising; the general training component of any company training provided to the agent, which cannot be reimbursed directly by the agent (say for risk-sharing reasons); and development of new products which can be sold to existing policy holders.²⁴ |

So the trade-off between the different ownership structures is as in Section 3. As in that section, we suppose that marginal benefits are small when average benefits are small. It follows that, if the company owns the list, the agent will have an insufficient incentive to deliver persistent clients, i.e. he will underinvest in this activity. The company, on the other hand, will have at least the socially correct incentive to invest in list building and similar activities, i.e. it will if anything overinvest in

these activities. In contrast, if the agent owns the list, the company will underinvest in list building, but the agent will work hard to deliver persistent clients.

Further understanding about list ownership can be gained by considering what would happen if the reason for the back-loading of commissions disappeared. Recall that the back-loading was necessary because (a) the agent devoted non-verifiable effort to the servicing of clients, and (b) the persistence of the client was sensitive to this effort. Much can be explained by noting that some kinds of insurance policies are more likely to be renewed than others, and this can make (b) much less of a factor. An example is "whole life" life insurance. A life insurance policy will involve a longer term contract than automobile insurance or fire and casualty insurance because a short term policy gives very little protection to a person against the event that he will be sick, but not die during the term of the life insurance policy and then be uninsurable thereafter. As a result, a life insurance customer has less of a tendency to switch insurance companies than does an automobile insurance customer. Moreover, to the extent that life insurance renewals do not occur, it is not because the agent has given the customer bad service on his claims!! When renewals are relatively insensitive to the agent's actions, the commission structure need not be as back-loaded, and hence the argument for the agent to own the list is weakened. Further, even with some back-loading, to the extent that one company's q's do not affect the desire of a client to switch given that his insurance is a long term contract, the agent has less need to own the list.

Our analysis therefore predicts that in products where the renewal is not guaranteed and is sensitive to the agent's actions, the agent will be more likely to own the list whereas in products where the renewal is more certain and is less sensitive to the agent's actions, the company will be

more likely to own the list. We now argue that these predictions are consistent with facts characterizing the insurance industry.

One important fact is that about 65% of the premiums in property-casualty insurance are generated by agents who own the client list, while in life insurance about 12% of the premiums are generated by agents who own the list.²⁵ Most property-casualty insurance is sold for a shorter term than most life insurance. Table 1 gives a more detailed breakdown for life insurance. It can be seen that term insurance is sold far more often by agents who retain list ownership than is whole life insurance.

Another important fact is that there is great variation regarding list ownership among products in the property-casualty product area. For example, independent agents have a 47% share of the market for private passenger automobile liability insurance, while they have a 96% share of surety insurance.²⁶ Marvel (1982) has shown that there is a positive correlation between the market share of independent agents, and the size of an agent's client acquisition costs (as measured by advertising and other acquisition expenses). We think that this is some support in favor of our conclusion that the agent will own the list when the agent's marginal incentives are relatively important in generating the renewal.²⁷ Table 1 is suggestive of a similar point for life insurance. The selling of substandard insurance and group/pension insurance involves substantial effort on the part of the agent to find an insurer which is a good match for the client. The willingness of the client to maintain his insurance coverage with the agent depends on the quality of the match. Hence the ownership of the list by the agent provides him with more protection from the non-contractible acts of the insurance company, than he would receive with company list ownership and the back loading of the commissions.

Table 1

Life Insurance

<u>Product</u>	<u>Percent of Agents Who Use Brokerage Companies to Sell the Indicated Product</u>
Substandard Insurance	55.9%
Term Insurance	46.2%
Group/Pension	43.1%
Whole Life Insurance	19.4%

Source: Czepiec (1984), Table 1

Note: The percentage figure refers to agents who "frequently" place their clients with a brokerage insurance company. A brokerage insurance company is an insurance company which uses independent "agents" who are called brokers because they do not have a legal agency relationship with the company but instead represent the client.

5. Conclusions

When two parties enter into a relationship in which assets will be used to generate income, the parties can, in principle, contractually specify exactly who will have control over each dimension of each asset in each particular future contingency. We have argued that there is often a low cost alternative to contracts which allocate all specific rights of control. In particular, when it is too costly for one party to specify a long list of the particular rights it desires over another party's machines, then it may be optimal to purchase all rights except those specifically mentioned in the contract. Ownership is the purchase of these residual rights of control. Vertical integration is the purchase of the assets of a supplier (or of a purchaser) for the purpose of acquiring the residual rights of control.

It should be noted that asset control is only important in situations where the quantity and quality variables which are directly relevant for the parties are ex-ante noncontractible. That is, if a buyer and seller who are engaged in a vertical relationship can specify in sufficient detail how the quantity and quality of the supplier's product should vary with the state of the world, then they can achieve the first-best under nonintegration. Under these conditions, there is no reason for either party to care about how the other party is using its assets. If such a detailed specification of quantity and quality is not possible, however, asset control becomes important since asset use is then a proxy for quantity and quality decisions.

The literature on transactions costs has emphasized that incomplete contracts can cause a nonintegrated relationship to yield outcomes which are inferior to those which would have been achieved with complete contracts. It is implicitly assumed that integration yields the outcome which would have arisen under complete contracts. We argue that the relevant comparison is

not between the nonintegrated outcome and the complete contract outcome, but instead between a contract which allocates residual rights to one party and a contract which allocates them to another. Residual rights have value to both parties, so when the buyer owns the seller, the manager of the buyer's subsidiary loses residual rights relative to those he would have in the non-integrated situation. The negative efficiency consequences of the loss of residual rights by one party must be weighed against whatever benefits are generated when those rights are given to the other party.

We have presented a model which analyzes the implications of different assignments of residual rights. This model emphasizes the distortions in ex-ante investments that are caused by contractual incompleteness. It should be noted, however, that in general other distortions may also be important. For instance, even if all ex-ante investments can be verified and hence are reimbursable, residual rights may matter if the ex-post distribution of the surplus is important for other reasons, e.g., due to the risk aversion of the parties. An example is where party one has an investment project, but does not wish to finance it entirely himself since he would then bear all the risk. One possibility is to raise the funds externally from the market, which is risk neutral, say. The outside investors, who we suppose are led by manager 2, should then receive as their return a sizeable fraction of the project's benefits. If manager 1 retains control of the project, however, he may be able to divert these benefits ex-post from the investors to himself through his choice of noncontractibles, and, knowing this, the investors may withhold some of their funds. In order to encourage outside investment, therefore, manager 1 may have to hand over some control to manager 2, e.g., by giving him ownership rights over some of the assets.²⁸

It should also be noted that if there is some barrier to ex-post renegotiation, due, e.g., to the presence of transaction costs or asymmetric

information, control of residual rights will be important in affecting the size of the ex-post surplus as well as the distribution (even in the absence of ex-ante investments). An analysis of the costs and benefits of ownership in this case may be found in Grossman and Hart (1984). A related idea is discussed by Farrell (1985).

It is worthwhile to consider which of the assumptions of the "Coase Theorem" we drop in order to reach the conclusion that the distribution of ownership rights has efficiency consequences. The model of Sections 2 and 3 permits ex-post bargaining of the type suggested in Coase (1960), but the ex-ante efficiency of the relationship between the two parties will depend on how residual rights of control are allocated. The impossibility of ex-ante bargaining over all aspects of the product to be delivered, i.e. the incompleteness of the contract, is the source of our conclusion that the distribution of property rights has efficiency consequences.

We have developed a model to illustrate the idea that ownership allocates residual rights of control in situations where contracts contain missing provisions. We have emphasized the symmetry of control, namely that when residual rights are purchased by one party they are lost by a second party, and this inevitably creates distortions. That is, integration shifts the incentives for opportunistic and distortionary behavior, but it does not remove these incentives.

FOOTNOTES

1. See also Teece (1980), Williamson (1971), Williamson (1983), Williamson, Wachter and Harris (1975).
2. Coase is quite specific about the benefit of integration: it creates a master-servant relationship out of what otherwise would have been a buyer-contractor relationship. This seems more a theory of the ownership of people than of assets. If we assume that an employee is as self-interested as an outside contractor, it is unclear what the difference between a master-servant relationship and a buyer-contractor relationship is.
3. See Evans and Grossman (1983) for an elaboration of the above critique of the transactions cost based arguments for integration.
4. Coase states that the size of the firm is limited by the managerial capacity of the single owner to manage many activities. As noted in the text, this is unconvincing, since the owner could always hire another manager. The other authors do not give any clear statement as to what limits the size of the firm, but appear to accept Coase's view that integration transforms a hostile supplier into a docile employee, and thus the contracting problems associated with independent ownership are greatly diminished. (However, there are some references to increased bureaucracy, and its associated cost. See Levhari and Keren (1983), Rosen (1982), Williamson (1967) and Waldman (1984) for specific models of how the number of people involved in production affect the overall cost of production. None of these papers makes any distinction between the activities carried out via contract to separate owners, and the activities carried out in a single ownership unit. That is, the theories are equally valid descriptions of how a firm can use hierarchies of outside contractors, as they are theories of employment within the firm.)
5. Richard Posner, whose opinion on the legal definition of ownership we solicited, has referred us to the following statement by O. W. Holmes (1881):

"But what are the rights of ownership? They are substantially the same as those incident to possession. Within the limits prescribed by policy, the owner is allowed to exercise his natural powers over the subject-matter uninterfered with, and is more or less protected in excluding other people from such interference. The owner is allowed to exclude all, and is accountable to no one but him".

O.W. Holmes, Jr., The Common Law 246 (1881).
6. See Chandler (1977, p. 54).
7. Arrow (1975) has analyzed the benefits of vertical integration based upon the assumption that without integration it is more costly for one firm to communicate information to another than with integration. We do not see why any new method of communication becomes feasible under integration. The incentives of people to lie may change if their incentive structure changes, but Arrow does not explain how integration changes the set of feasible incentive structures.

8. Williamson (1983) gives an example (pp. 523-524) of a contract written between non-integrated firms where there is no penalty for cancellation. He assumes that under vertical integration, or via the use of hostages, it is possible to extract a penalty from the buyer when he fails to take delivery of the seller's product. To the extent that there are artificial legal barriers to cancellation penalties, the same effect can be achieved by the deposit of a large lump sum payment at the time the contract is signed which will only be partially paid back to the buyer if he refuses to take deliver.
9. We model the relationship as a "once and for all" event. To the extent that there will be a long term relationship, the incentives for vertical integration may be different from those we have given here. See Tellaer (1981) and Kreps (1984) for the role of reputation in long term relationships as an enforcement device, and Williamson (1979) for arguments on the role of repetitive idiosyncratic purchases in providing a cost to nonvertical integration. None of these papers deals with the influence of reputation on the ownership of assets. To the extent that reputation helps mimic the outcome of a complete contract, it would seem to have no implication for the ownership of residual rights of control.
10. See Grossman and Hart (1984) for models where some components of q are contractible, while others are not contractible.
11. In a more complex model ex-post inefficiencies will also appear in conjunction with costs of renegotiation. See Grossman and Hart (1984, Section 2) for a model of ownership where ex-post inefficiencies rather than ex-ante inefficiencies are analyzed.
12. Here a_i , q_i are vectors in compact subsets of Euclidean spaces A_i and Q_i respectively, and B_i and ϕ_i are continuous functions.
13. A detailed analysis of contracts between coal mines and electricity generating plants may be found in Joskow (1984).
14. In a more general model, accounting data such a profit might be publically observable. Each manager would presumably then be under some incentive scheme chosen by the owner. B_i might then represent remuneration from this incentive scheme net of effort costs.
15. The assumption that manager j observes a_i can be justified on the grounds that since managers i and j are involved in a common project they are likely to have good information about each other.
16. This means that the owner of firm i can contract with a subordinate to implement the choice of q_i ; moreover, since there are many subordinates available, none is in a position to refuse to carry out the owner's wishes or to argue about terms.
17. Elsewhere we have considered the effect of date 0 contractibles for the special case where no revisions of the date 0 contract are permitted at date 1. See Grossman and Hart (1984). In the present model, however, where revisions are allowed, the introduction of contractibles

complicates matters greatly. With a contractible, not only can the parties agree on a schedule relating the payment from firm i to firm j to the contractible, but also they can agree on a way of revising this price schedule at date 1 according to messages manager i and manager j send reflecting the choice of the sunk investments a_1, a_2 (for an analysis of this in a special case, see Hart and Moore (1985)). With no contractibles, the payment from firm i to firm j at date 1 is just a constant, and any attempt to make it sensitive to the environment will fail since price revisions are a zero sum game from the point of view of the buyer and seller.

18. Note that it would be possible to achieve the first-best if the q_i were ex-ante contractible, even given that the a_i are not. For if the date 0 contract specifies that $q_i = q_i^*$, party i has an incentive to choose a_i to maximize $B_i(a_i, \phi_i(q_i^*, q_j^*))$, i.e. to set $a_i = a_i^*$.
19. There is also a class of more complicated contracts that make asset ownership at date 1 a function of messages the managers of firms i and j send after they have observed each other's investment decision. An example of this is an option to own contract. Our results are not affected by the existence of such contracts and so, for simplicity, we ignore them.
20. The reader may be concerned about our assumption that the manager can think clearly enough about q to solve (5.6), but that it is too costly to contract for q or design a mechanism to implement a particular q . This assumption can be understood if we imagine that the non-contractible represents a special service which will be required of a firm at date 1, and that the type of service which is appropriate depends on the realization of a state of nature. Let there be N states of nature. The states are defined in such a way that state s requires the choice of activities from an M dimensional space denoted by Q_s . The idea is that different activities are required for different states, i.e., while elements of $Q_s, Q_t, s \neq t$, are both M dimensional Euclidean vectors, their coordinates refer to entirely distinct activities (different machines, for example). Further in state s , the benefit function B is assumed to depend on the noncontractibles only through the chosen element q_s in Q_s , say $B = B(a, q_s; s)$; if in state s , some vector of activities in Q_t is chosen, $t \neq s$, no benefits are derived. Suppose in addition that we can normalize the spaces of activities so that $B(a, q_s; s) = B(a, q)$ where q lies in a single space Q (where the coordinates of q , of course, continue to refer to different activities in different states). Then, from an ex-ante point of view, the manager, taking each s as equally likely, thinks of his objective as $B(a, q)$, where q is a typical value assigned to the vector q_s . Further, any element q in Q is contractible ex-post (so that ownership has some value). However, to make q ex-ante contractible, it would be necessary to specify different coordinates of q for each of the N states, and we assume that this is too costly (note that even with a small number of states, it may be difficult to do this if it is hard to describe the states objectively in advance).
21. The result that the conflict over the division of surplus at date 1 can lead to a distortion in investment at date 0 is similar to the finding of Groot (1984). In Groot's model, however, investment expenditure is observable, there are no noncontractibles, and the inefficiency in ex-

ante investment results from the assumed impossibility of writing binding contracts.

22. In practice, the replacement of a manager may well be publically observable and hence contractible on. So that we can stick with our framework where there are no contractibles, we suppose that a replacement involves a move from one job to another in the company, which may not be verifiable (the manager may be "kicked upstairs" to a job with no perquisites, for example). That is, a replacement is part of the noncontractible q_2 . Note that if $f_2 < 0$, then the owner will not replace manager 2 but instead impose the cost $-f_2$ on him. Proposition 2 can be generalized to this case if there is a noncontractible action under manager 2's control which forces the owner to hire another manager and compensate him for the amount $-f_2$ (see Remark 1 below). Then the owner's net benefit again becomes $B_1 + f_2$.
23. Our statements about the structure of the insurance industry are based upon Webb et. al., (1984) Strickler, (1981) and conversations with professionals in the Insurance industry. We are very grateful to Naava Grossman for her help in finding general information and data sources, and for providing general information herself. We would also like to thank Peter Thistle.
24. In each of the examples of non-contractibles, we have indicated acts which each party could take that put the other party at a disadvantage. In some of these cases, for example the insurance company changing the support it provides to a given product, the non-contractible act does not involve direct manipulation of the item of which we are trying to explain the ownership, namely the client list. We have taken as given that the insurance agent does not want to own the whole insurance company. Hence the relevant variable which will allow the agent to increase his control over the renewal premiums to be generated by a particular client is the ownership of the list, rather than direct control over the insurance company's marketing and product support program.
25. The property-casualty number comes Webb et. al, p. 85; the life insurance number is from LIMRA (1977) p. 9 and is the fraction of premiums written by insurance brokers (as opposed to captive agents) in 1977 for the United States. The LIMRA study also estimates that brokers tend to specialize somewhat in term policies, rather than whole life policies.
26. See Webb et. al., pp. 85-88.
27. Marvel (1982) offers an alternative explanation for the correlation. He argues that there are situations where it is more efficient for the company to advertise than for the agent. In these situations, the insurance company helps bring the client to the agent. According to Marvel, an agent who did not have an exclusive dealings contract with the insurance company could then switch the customer to another insurance company which does not advertise and thus can pay higher commissions. This argument faces the following difficulty. First, if the company advertises the specific benefits of its product, why should the customer allow the agent to switch him to another insurance company? Marvel seems to be assuming that the agent uses a "bait and switch" sort

of tactic against his customers. Second, if the insurance company convinces the customer about the general benefits of insurance, then how does an exclusive dealings contract protect the insurance company? The customer will just go to a cheap company which advertises somewhat less (which he can find in the Yellow Pages, rather than through television).

Another piece of evidence which Marvel brings in favor of his argument is that exclusive dealings companies tend to spend more on advertising, than do companies without exclusive dealings contracts. This correlation is consistent with our explanation as well. If, for any reason, a company is assured of more policies per customer it acquires, then it may spend more on acquisition costs. Further, Marvel does not explain the fact that life insurance tends to be sold through captive agents far more frequently than property casualty insurance. He also does not distinguish ownership of the renewal from exclusive dealings. A company can own the renewal without having an exclusive dealings contract. Industry sources are emphatic in pointing out that "...The most important characteristic of the independent agency system in comparison with the exclusive agency system is the independent agent's 'ownership of renewals or expirations'..." Strickler (1981, p. 294). The renewal plays no role in Marvel's argument. Nevertheless Marvel's argument can be modified to supplement ours as follows. First, an exclusive dealings contract is one method of enforcing list ownership rights. Second, if unlike Marvel we assume that the company is advertising the specific high quality of its agent force (e.g., "Your State Farm agent is always available"), then an exclusive dealings contract would be a method of recovering the expenditures from its agents. Note that an insurance company which is involved in selling for a longer time than any one agent or customer has reputational incentives to choose agents of high quality, so that its advertising is to some extent truthful. We would then argue that companies will have exclusive dealings contracts when they are better able to convince customers of the agent's quality than is the agent, and/or they are able to select agents of high quality.

28. We would like to thank John Minahan for a helpful discussion about this example.

REFERENCES

- Aron, D., "Ability, Moral Hazard, and Firm Diversification, Part I," University of Chicago, mimeo, May 1984.
- Arrow, K. J., "Vertical Integration and Communication," Bell Journal of Economics, Vol. 6, No. 1, (Spring 1975), pp. 173-85.
- Coase, Ronald H., "The Nature of the Firm," Economica n.s. 4, (1937): 386-405; reprinted in George J. Stigler and Kenneth E. Boulding, eds., Readings in Price Theory. Homewood, IL: Richard D. Irwin, Inc. 1952.
- _____, "The Problem of Social Cost," Journal of Law and Economics, Vol. 3, (October 1960): 1-44.
- Chandler, A. Jr., The Visible Hand, Harvard University Press.
- Evans, D., and Grossman, S., "Integration" in Breaking Up Bell, Essays in Industrial Organization and Regulation (David Evans, ed.) North Holland Publishing Co., (New York: 1983), 95-126.
- J. Farrell (1985), "Allocating and Abrogating Rights: How Should Conflicts be Resolved Under Incomplete Information?", mimeo.
- Goldberg, Victor P., and Erickson, John E., "Long-Term Contracts for Petroleum Coke," Department of Economics Working Paper Series No. 206, University of California, Davis, (September 1982).
- Grossman S. and O. Hart (1984), "Vertical Integration and the Distribution of Property Rights," mimeo.
- Grout, P., "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," Econometrica, Forthcoming.
- Hart, O. and J. Moore, (1985), "Incomplete Contracts and Renegotiation," M.I.T. Working Paper #367.
- Holmes, O. W., The Common Law 246, 1881.
- P. Joskow (1984), "Vertical Integration and Long-Term Contracts: The Case of Coal Burning Electric Generating Plants." MIT Working Paper #361.
- Keren, M., and Levhari, D., "The Internal Organization of the Firm and the Shape of Average Costs," Bell Journal of Economics, Vol. 14, No. 2, (Autumn 1985): 474-486.
- Klein, Benjamin. "Transaction Cost Determinants of 'Unfair' Contractual Determinants," American Economic Review, 70, (May 1980): 356-362.
- _____, Crawford, R. A., and Alchian, A. A., "Vertical Integration, Appropriate Rents, and the Competitive Contracting Process," Journal of Law and Economics, 21 (October 1978): 297-236.
- Kreps, D., "Corporate Culture and Economic Theory" unpublished mimeo (1984).

- LIMRA, "Brokerage Ordinary Life Insurance Production by Ordinary Agencies and Branch Offices, United States and Canada in 1977", Research Report 1979-3, Life Insurance Marketing and Research Association, Farmington, Connecticut.
- Mann, D. and Wissink, J., (1984), "Inside vs. Outside Production: A Contracting Approach to Vertical Integration," Center for the Study of Organizational Innovation, University of Pennsylvania.
- Marvel, H., "Exclusive Dealing," Journal of Law and Economics, Vol. XIV, April 1972, pp. 1-25.
- Rosen, S., "Authority, Control, and the Distribution of Earnings," Bell Journal of Economics, Vol. 13, No. 2, (Autumn 1992): 311-323.
- Rubinstein, A., (1982), "Perfect Equilibrium in a Bargaining Model," Econometrica, 50, 97-109.
- Strickler, Nancy, editor, Marketing Life and Health Insurance, (Life Office Management Association: 1981).
- Teece, David J., "Economics of Scope and the Scope of the Enterprise," Journal of Economic Behavior and Organization, 1, (September 1980): 223-245.
- Teiser, L. G., "A Theory of Self-Enforcing Agreements," Journal of Business, 53, (February 1981): 27-44.
- Waldman, M., "Worker Allocation, Hierarchies and the Wage Distribution," Review of Economic Studies, 51, (January 1984): 95-109.
- Webb, B. L., Laumie, J. J., Rokes, W. P., and Baglioni, N. A., Insurance Company Operations, Vol. 1, 3rd ed., (American Institute for Property and Liability Underwriters: 1984).
- Williamson, Oliver E., "Hierarchical Control and Optimum Firm Size," Journal of Political Economy, 75, (April 1967): 123-138.
- _____, "The Vertical Integration of Production: Market Failure Considerations," American Economic Review, 61, (May 1971): 112-123.
- Williamson, O. M. Wachter, and J. Harris, "Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange," Bell Journal, (Spring 1975): 250-61.
- _____, "Credible Commitments: Using Hostages to Support Exchange," American Economic Review, 83, No. 4, (September 1983): 519-540.