Incomplete Contracts, Control Rights and Integration Decisions in Economic Organisations

Philip Iestyn Williams

London School of Economics

A thesis submitted for the PhD degree, University of London.

March 1996

Abstract

This thesis comprises an introduction and four distinct chapters. Its central theme is the role played by the allocation of asset ownership rights in motivating asset-specific investment, when contracts are incomplete.

Chapter 1 considers the debt financing of an entrepreneurial project. To encourage asset-specific investment and loan repayment, debt structure should minimise both (voluntary) strategic default and liquidation following (unavoidable) liquidity default. Liquidation incentives are critical and shown to depend crucially on creditor characteristics. In general, borrowing from multiple creditors with contrasting attributes is found optimal. The benefits of borrowing from a creditor also undertaking project trade are explored.

In Chapter 2 the relationship between asset ownership and investment specificity is examined. Asset control encourages efficient, asset-specific investment by owners. However, lock-in fears lead non-owners to choose widely applicable but less effective investment. The interactions between asset ownership, firms' technology choices and workers' investments are considered. In particular, it is found that the costs and benefits of individual integration decisions are sensitive to overall industry structure.

The specificity framework is extended in Chapter 3 to model a retailer's product choice. Vertical merger encourages investment in integrated supply and foreclosure of non-integrated manufacturers. An anti-competitive as opposed to an efficiency interpretation depends delicately on the trade-off between the benefits of supplier-specific investment and multi-product retailing. Where retailers compete, it is shown that vertical integration implements effective competition-reducing differentiation strategies.

In Chapter 4 vertical integration, through the incentive effects of asset ownership, is shown to amount to a specialisation decision. The attractions of encouraging investment in input as opposed to final good production depend on the effectiveness of investment at each manufacturing stage, and the scale benefits of input sales to generally rivalrous downstream firms. These benefits are sensitive to downstream competitive pressures, yielding a potentially non-monotonic relationship between competition and integration.

Contents

Acknowledgements

During the long period it has taken to complete this research I have accumulated an enormous debt to my supervisors, John Moore and Patrick Bolton. I would like to thank them both for their time, tolerance and encouragement - it has not been an easy task for them! To have benefitted from their supervision has been an enormous privilege. My intellectual debt is apparent throughout this dissertation.

For one year John Sutton took on the role of formal supervision. However, his support and advice has continued over a far greater period. He also found me a home with the Industry Group in STICERD, for which I am extremely grateful. During his stay at the LSE in 1992-93, Oliver Hart gave freely of his time in discussing many of the ideas incorporated in the thesis. This was far more than I could have expected. Since I began work at the University of Southampton, Geoff Stewart has also offered support and encouragement.

I have been extremely fortunate in having excellent economists as fellow students at LSE. Much of the work in this thesis was first presented at the Economics Research Students' Seminar. The perceptive comments of the participants there has greatly improved the final output. More directly, Chapters 3 and 4 of my thesis are based on joint work with two of these students. I owe Frank Bickenbach and Maija Halonen respectively a debt of gratitude. Much of what I have learned during this project was learned with or from them.

I gratefully acknowledge financial support from the ESRC.

I would also like to thank Aeronwen Griffiths. This work has imposed an enormous burden on her. The value of her constant support throughout this undertaking cannot be overstated.

Introduction

At the heart of this thesis is the idea that ownership matters. A casual example of its importance is provided by Margaret Thatcher's policy of selling council-owned housing to tenants. My own grandparents were among the first to take advantage of this opportunity. Ignoring the political dimension and focusing on the personal, there can be no doubt that few other events in the last twenty years have had as great an impact on their lives. No one entering their house, both before and after this reallocation of property rights, could fail to notice the influence a change in ownership has had on their actions. This relationship between the allocation of ownership rights and individual incentives will be our fundamental concern. Specifically, we will focus on the central role played by ownership in motivating asset-specific investment. In particular, we will focus on the organization of the firm.

Intuitively, ownership matters because it confers authority over assets. It was my grandparents decision alone to install double-glazed windows, to knock down an internal partition, to install heating in all rooms. Crucially, ownership provides some control over the relationships that involve the asset. In my grandparents case, nothing was more important than securing their absolute right to occupy the house. In recent years it has become increasingly apparent, of course, that initial property rights are not a guarantee of perpetual control. Where house purchase is based on mortgage finance, for instance, default on contracted repayments may result in a transfer of ownership and control rights to creditors. This is an important issue, to which we will return. My grandparents were fortunate in this respect - Mrs Thatcher's *generosity* allowed them to buy their house outright.

A formalisation of these ideas was provided by the incomplete contracts approach to ownership formulated by Grossman and Hart, in their classic 1986 JPE paper [Grossman and Hart (1986)]. They offered a rigorous re-statement and extension of ideas that had evolved in a literature extending back to the seminal work of Ronald Coase on the nature of the firm [Coase (1937)]. At the heart of the Grossman-Hart perspective is the identification of asset ownership rights with residual control rights. These residual rights yield control of the asset wherever restrictive contracts agreed by the asset owner are silent, or unenforceable. As

Coase recognised, such general, broadly specified powers of authority are a defining feature of organization within the firm.

Of course, residual rights have operational importance only where contracts are incomplete i.e. where the enforceable terms of contractual agreements are not all-encompassing. Indeed, as Oliver Williamson has eloquently argued [Williamson (1985), p.30-32], an authority-based governance structure is relevant only where contractual incompleteness coincides with opportunistic behaviour and asset specificity. When complete, definitive contracts can be written, residual control rights are irrelevant. Without opportunism, the power of a broadly defined "promise" is sufficient to guarantee efficiency. Finally, in the absence of specificity, asset ownership is no longer critical. Instead, market based transactions can be relied upon to generate desirable outcomes. Again, Coase foreshadowed the emphasis specificity places on long-term, idiosyncratic relationships in his focus on contracting issues in the 1937 Economica article.

It is the combined assumptions of incomplete contracts, opportunism and asset specificity that provide the link between ownership and investment incentives in the Grossman-Hart framework. If investment is asset-specific, then by definition the relationship between the investor and relevant assets is critical to value creation. However, if complete, definitive contracts can be written, this only implies a minimal relationship between the investor and the asset *owner*. The precise nature of the investment, the terms of access to the asset and any division of the returns to investment can be enforceably specified once-and-for-all, in advance of any asset-investor interaction. Yet if contractual incompleteness bites, the allocation of the residual control rights of ownership will define the nature of an on-going interaction. Once asset-specific investment has been undertaken, the investor is then essentially locked-in to a long-term relationship with the asset owner.

In particular, an opportunistic owner may manipulate the incompleteness of contracts to abrogate any initial agreement. New terms for the asset-investor relationship must then be renegotiated *after* investment has been undertaken.

In general, this will enable the owner to secure a share in the returns to investment.[1] Consequently, the investor will only receive a fraction of those returns. Foreseeing this eventuality, the investor is likely to underinvest in the asset specific relationship.

An obvious remedy to the underinvestment problem cited is to allocate asset ownership rights to the investing party. The investor is then guaranteed the full return on investment, ensuring efficiency. However, in general, any asset will be involved in numerous asset-investor relationships. Any particular allocation of residual control rights, encouraging investment by the owner, will then inevitably imply inefficient underinvestment by non-owning parties.[2] Indeed, the observation that the costs and benefits of ownership "*can be understood as two sides of the same coin*" [Hart and Moore (1990), p.1120] represents the fundamental insight of the Grossman-Hart framework.

As stressed, the particular assumptions of incomplete contracts and asset specificity are fundamental to the approach highlighted. With respect to the first of these key assumptions - the pervasive impact of contractual incompleteness - little work has been undertaken to add substance to its superficial merits.[3] In part, this reflects the elusiveness of a formal model of bounded rationality, which underlies much of the motivation for the difficulty in writing incomplete contracts [see e.g. Williamson (1985)]. Recent work by Segal (1995) represents a notable preliminary attempt to place the foundations of the incomplete contracts assumption on a firmer footing. In essence, one must ask why can definitive contracts not be written? In a number of recent papers, sophisticated

---

[1] Of course, an opportunistic investor too may seek to use contractual incompleteness to renege on the terms of the initial agreement.

[2] Note that allocating residual control rights jointly to several agents will not eliminate the general underinvestment problem. For consistency, some form of decision-making mechanism regarding these rights would have to be instituted e.g. veto powers, majority voting. Inevitably, such processes would again involve a sharing of returns to individual asset-specific investment.

[3] A discussion of the foundations of the incomplete contracts assumption is provided in Hart (1995), Chapter 4.

mechanisms have been employed to push back the bounds on contractability. For example, it is widely assumed in the incomplete contracts literature that, though not *verifiable* to a court of law, actions are *observable* to the interacting agents. If mechanisms can be designed such that these knowledgeable agents truthfully reveal whether the terms of contracted agreements have been met, then a complete contract environment can be effectively restored.

While accepting the weakness of an inadequate understanding of this cornerstone in our approach to ownership, we firmly believe that the general principles of the incomplete contracts perspective are robust. We will not undertake a rigorous analysis of its foundations here. Rather, we will (tentatively) take the pervasiveness of incomplete contracts as given in the analysis to follow.

In contrast, the other critical assumption underlying our approach - asset specificity - will be considered in some detail. In general, investment will not be wholly specific to a single asset. There are two dimensions to this. Firstly, an investment will rarely require access to a unique asset to generate any value at all i.e. few assets are *essential*, as defined in Hart-Moore (1990). Secondly, an investment will rarely generate maximum value in conjunction with a single asset alone i.e. there are generally complementarities between assets. Likewise, assets are rarely investment-specific.

Once these aspects of specificity are recognised, a multilateral framework for analysis is illuminating, indeed inevitable. In particular, an investor's return and therefore incentives to invest will depend on the value of alternative asset configurations, and the distribution of asset ownership rights. For example, if an investment requires access to many assets, and ownership of those assets is widely dispersed, then the return on the investment must be shared with many owners. Significant underinvestment is then likely.

Hart and Moore (1990) makes a notable contribution in this multilateral direction. They emphasise the coordination role of collecting asset ownership rights in the hands of a single owner. The firm can therefore be thought of as a nexus of ownership rights. Their key insight is that such integration can improve the investment incentives of non-owning workers, in addition to those of owner-investors. In reducing the number of parties involved in asset access negotiation, and hence the division of investment returns, concentrating ownership can encourage asset-specific investment by such workers.

A second significant contribution in a multilateral framework is provided by Bolton and Whinston (1993). They model an environment where a monopoly input supplier, serving multiple manufacturers, may be capacity constrained. In effect, when capacity constraints do not bind, the supplier has assets (inputs) which are specific to each buyer's investment, since there is then no competition for input. On the other hand, when capacity constraints bind, the asset or assets are no longer totally buyer specific. In this capacity constrained context a vertically integrated input supplier may overinvest at the downstream manufacturing stage. By raising the value of the supplier's internal supply option, such investment reduces the extent of any supplier lock-in to the non-integrated manufacturer. In turn, this raises the supplier's share of the non-integrated buyer's investment returns.

As highlighted, at the heart of the transaction cost approach is the view that when investment is asset specific, markets will not operate effectively. The distribution of surplus will therefore be driven by individual negotiating positions. In a multilateral environment, with a myriad of actual and potential interactions, determining the outcome of this overall bargaining problem would be an extremely complex task. Furthermore, the appropriate structure for this problem will not, in general, be unambiguously determined. Precise outcomes are likely to depend delicately on the details of the specific institutional environment. Hart-Moore and Bolton-Whinston, for example, adopt very different approaches to the bargaining structure. Hart-Moore adopt a cooperative approach, based on the Shapley value, while Bolton-Whinston explicitly model a non-cooperative setup. It is reassuring that the broad pattern of results, at this *first cut*, appears relatively robust. Clearly, an ad hoc approach to the bargaining problem is, fundamentally, undesirable. Nevertheless, the bargaining structures we adopt in our analysis are more-or-less arbitrary. We believe significant insight, at minimum complexity cost, can be achieved in this manner.

The two multilateral papers cited above differ significantly in their approach to the determination of asset ownership. The Hart-Moore analysis is conducted in the efficiency spirit of the traditional transaction cost literature. Extending the Grossman-Hart approach to a multilateral environment, they (implicitly) invoke the Coase Theorem, in considering ownership rights allocations that maximise overall efficiency. This contrasts with the framework adopted by Bolton-Whinston.

Bolton and Whinston stress that where the *overall* pattern of asset ownership is not (or cannot be) definitively and enforceably contracted upon, then externality effects may arise in the equilibrium allocation of ownership rights. They argue that, in precisely the incomplete contract environment considered, it may be impossible to detail optimal ownership arrangements for every conceivable contingency. Adopting crude and rigid ownership rules may then be extremely inefficient. Furthermore, in an industry setting, anti-trust regulations may limit the range of enforceable inter-firm ownership agreements. In undertaking bilaterally attractive trade in ownership rights, the two agents involved will not then internalise the impact of their decisions on third parties. In the Bolton-Whinston context, generally inefficient integration outcomes may result.

In Chapter 2 of the thesis we consider the specificity of investment explicitly. The more specific an investment, the more effective it will be in a particular application. General investment will be less effective in any given use, but have wider applicability. Unlike earlier models we focus not only on human capital investment by agents, but also on physical capital investment by firms. Indeed, we will pay particular attention to the interaction between these investment decisions.

Where ownership and investor roles are separated, fear of lock-in will discourage efficient, specific investment. For instance, non-owning workers will be reluctant to develop firm-specific skills, preferring general human capital that maximises employment options. Integrated outcomes will ensure close linkages between assets and between owners and assets, but lead to poorly tailored skills for non-owning workers. Non-integrated structures, on the other hand, in fragmenting ownership encourage more widespread ties between workers and assets, at the expense of weaker linkages between assets.

We will assume ownership trade is driven by *bilateral*, rather than overall efficiency, concerns. Externality effects may then emerge. Integrated firms are likely to develop idiosyncratic technologies. Consequently, in highly integrated settings the market for any particular skills is likely to be thin. Investment by non-owning workers in given specific skills is therefore discouraged, increasing the mismatch costs of integration. However, in deciding to integrate, individual firms will not internalise the impact of their decisions on the industry-wide costs of integration.

This last point can be illustrated using the housing example with which we began. My grandparents live on a large estate of identical houses. Any neighbourhood-specific (e.g. employment), or moveable house-specific (e.g. furniture) investment could therefore be easily re-deployed, at little cost, to an alternative house. However, while the whole estate was council-owned, the possibility of a neighbourhood house transfer would not prevent lock-in to the local authority.

In contrast, suppose a number of tenants had purchased their homes on the estate already, and a healthy housing market existed. My grandparents would then no longer be locked-in to the council, since the option of a private house purchase would then always be available. Efficient, house-specific investment would be safeguarded, without the need for my grandparents to undertake any property purchases themselves. Of course, in making their individual ownership decisions, other tenants will not consider the impact on my grandparents.

The recognition that bilateral considerations alone may drive ownership structure is essential, if a theory of the firm is to contribute significantly to our understanding of *industry-wide* organisation. This relationship between organisation *within* the firm, and the interaction *between* firms is surely of fundamental importance. [4] The potential for bilaterally motivated integration allows not only the existence of externalities, but also for the possibility that integration is undertaken for strategic reasons.

At one level, of course, it is a logical conclusion of the Coasian argument that the processes that determine the boundaries of the firm must also define the nature of market interaction between firms. Stated simply, markets begin where firms end. However, important strategic implications may emerge even when ownership decisions do not affect inter-rival transactions directly. Our central thesis is that ownership has a significant impact on individual incentives. Vickers (1985) and Bonanno and Vickers (1988), among others, have recognised the strategic role of firm structure, in manipulating such incentives. However, only recently have strategic considerations been married to an explicit theory of ownership and the boundaries of the firm. In Chapters 3 and 4 we will adopt such an approach.

In Chapter 3 we model a retailer, who must decide between rival suppliers and between single-product and multi-product retailing. As in Chapter 2, the retailer has an investment specificity decision to make. In the model, this takes the form of a discrete location choice. Vertical integration may encourage supplier-specific location, where an independent retailer would choose general investment. Indeed costly merger will only take place if a change in the retailer's location decision is induced.

Retailer supply patterns may, but need not, be sensitive to the location decision. Integration can therefore result in foreclosure of the non-integrated supplier. Whether this foreclosure is efficient or not depends delicately on the inter-relationship between the relative merits of specific versus general investment, and single versus multiproduct retailing. In the context of this simple model we explore the interaction between efficiency and competition motives for vertical

[4] Indeed, in exploring the birth of the modern corporation Alfred Chandler links innovation in the internal organization of the firm with the emergence of oligopolistic industry structure.

integration.

We extend the basic model to consider the impact of vertical integration on competition between *retailers.* Integration, in encouraging supplier-specific investment, allows retailers to adopt product differentiation strategies, ameliorating the effects of competition. We show that retailer competition concerns may motivate otherwise unattractive merger. For moderate levels of competition, a partially integrated industry structure then results, as integration by a subset of firms is sufficient to reduce competition significantly. For stronger competition levels, a coordinated integration process is required to realise the benefits of reduced competition. In this case, a fully vertically integrated industry results. The integration process then has the characteristics of an integration chain.

In the final chapter of the thesis, Chapter 4, we again consider the industry-wide effects of vertical integration. In the spirit of Grossman-Hart, we explicitly model the investment trade-off inherent in any ownership allocation. In the vertical context, integration effectively amounts to a strategy of specialisation at a particular production stage. If asset ownership is concentrated in the hands of downstream managers, strong downstream investment is encouraged but there will be little investment at the upstream level. In contrast, a non-integrated structure encourages moderate investment at both production stages. The resolution of this trade-off will, in part, reflect the relative effectiveness of investment at each stage in the overall value-creation process.

In a multilateral industry environment this trade-off will also depend on the possibilities of trade with other firms. In our particular context, a non-integrated upstream supplier may make additional input sales to other downstream firms. Non-integration is then encouraged by scale effects on upstream investment - cited in the traditional transaction cost literature as a factor limiting the extent of vertical integration.

As competition between downstream firms intensifies, the attractions of this input supply role are ambiguous. For weak competition, supplying rivals with input may have a negative effect on overall profits - the impact on downstream profits outweighing any added profits from input sales. However, as final product market competition intensifies, the profits of the input supply operation become increasingly attractive, as downstream profits are increasingly dissipated. A non-monotonic relationship between the intensity of final market competition and industry structure may therefore be observed.

In the preceding discussion we have assumed that, where a trade of ownership rights is bilaterally attractive, such trade will always take place, even when contracting restrictions prevent the socially optimal ownership configuration from emerging. Of course, those parties giving up ownership rights to valuable assets will require compensation. Yet it is a fundamental tenet of our incomplete contracts - residual control rights approach that such compensation cannot be credibly offered directly from future returns to efficient asset-specific investment. Therefore, if initial asset ownership is separated from agents with *potentially* significant efficiency-enhancing investments, *actual* investment by these agents may not take place, unless they have access to sufficient initial finance. The power of the Coase Theorem here relies critically on an absence of wealth constraints. There is little reason to believe the distribution of asset-specific skills will be reflected in the distribution of initial wealth. This is especially true given the incomplete contracting emphasis on idiosyncratic, non-marketable investment. Concentrating control rights in the hands of those with initial wealth is therefore unlikely to be efficient.

For efficiency, mechanisms are required for the effective transfer of initial cash endowments, from wealthy agents to those with valuable investment opportunities. Of course, such project funding will only be forthcoming if lenders can secure adequate repayment, at least in expected terms. Yet in precisely the incomplete contract environment outlined, where we have argued that asset ownership matters most, guaranteeing repayment may also be most difficult.

Returns will accrue in the first instance to the agents in control of project assets. Where these returns are non-contractible and the controlling agents are able to divert project returns for private benefits [see Hart-Moore (1989), and Hart (1995)], repayments must be voluntary (ruling out standard equity). Project cash can be extracted from controlling managers only if future access to valuable assets is made credibly contingent on such repayment. This debt-like relationship between project finance and contingent asset control was highlighted in the path-braking analysis by Aghion and Bolton (1992). Repayment-inducing benefits of future asset control may include continued access to multi-period project returns, as in Hart-Moore (1989), or repeated finance for new project opportunities, as in Bolton-Scharfstein (1990).

Of course, the credibility of any threat to deprive the manager of asset control in the event of default, depends critically on the assets generating significant value for the creditor in alternative use. This value will, for instance, depend on the extent to which the asset-manager relationship is idiosyncratic, and also on the buoyancy of asset resale markets. We will return to this issue in greater detail below..

Where project returns are driven by entrepreneurial asset-specific investment, the financier's problem therefore involves a fundamental tension. To maximise project cash extraction, the threat to seize control of entrepreneurial assets in the event of default must be credible. Yet as we have seen in earlier discussion, security of asset control is essential if the entrepreneur is to undertake asset-specific investment in incomplete contract environments.

This problem is exacerbated when project returns are uncertain. Our incomplete contracting environment may prevent scheduled debt repayments responding *directly* to the state of nature. However, one would ideally like the structure of the debt agreement to reflect the contingent nature of project returns. In particular, an ideal debt contract would ensure maximum feasible repayments in every state of nature, while minimising the threat of *unavoidable* default, and loss of project control, by the entrepreneur. As stated in Bolton and Scharfstein (1993): "an optimal contract balances the benefits of deterring strategic defaults against the costs of inefficient liquidation following liquidity defaults".

One approach to this problem is to adopt a standard incentive design framework. The magnitudes of cash transfers from the entrepreneur are likely to be simply verifiable. By ensuring repayments are state dependent, a renegotiation-proof, incentive compatible debt contract allows the state of the world to be effectively revealed by the entrepreneur. By making the extent of the liquidation threat contingent on repayment, it is then (indirectly) dependent on the state of the world. Of course, the need for incentive compatibility, in general leads to inefficiency.

Alternatively, if the extent to which an initial debt contract can be renegotiated is state-dependent, then *equilibrium* repayments too may be made contingent on the state of the world. Creditors' incentives to renegotiate debt after default will reflect their liquidation values for project assets. Now if the levels of renegotiated repayments can be structured to accurately reflect the pattern of state contingent project returns, then maximum cash extraction can be achieved, while ensuring minimum inefficient liquidation. This approach to debt and renegotiation was pioneered in Hart and Moore (1989). We adopt their framework in Chapter 1 of the thesis.

A creditor's liquidation value for a project is determined by one of two factors - the strength of the resale market for project assets, and the creditor's own private value for the project. Clearly, the magnitude of this last factor will reflect the creditor's abilities to extract value from the project i.e. it will depend on creditor characteristics. In Chapter 1 we argue, in the spirit of Shleifer and Vishny (1992), that the resale market for project assets is likely to be weak in precisely those states where the project entrepreneur is cash constrained, and likely to default on debt repayments. After all, our incomplete contracts approach has been based on a foundation of asset specific, idiosyncratic investment. Those agents with high value for such project assets are therefore likely to come from the same industry as the original entrepreneur, and will suffer many of the same shocks.

As a result, potential buyers for project assets are likely to be cash constrained precisely when liquidity default occurs, in bad states of nature. The extent of debt renegotiation in bad states of the world will therefore be driven by the creditor's own value for project assets. Consequently, creditor characteristics will be important for the structure of debt and for the total finance that can be raised. The question of *from whom* entrepreneurs borrow matters.

Suppose we represent a creditor's private value for liquidated assets by a factor L. The ability of low L creditors to extract project cash will be limited in the bad state of the world. Nevertheless, this does act as a commitment to renegotiate when the entrepreneur is cash constrained, allowing the face value of debt to be set high. Low L creditors are therefore able to extract maximum project cash in the *good* state. High L creditors, on the other hand, are effective at cash extraction in the *bad* states of the world - potentially too effective - since they do not rely on asset resale markets. To prevent excessive liquidation of project assets, such creditors may have to limit the face value of debt, in turn restricting their cash extraction powers in the good state. Significantly, these contrasting attributes suggest the value of multiple creditor borrowing. This possibility is explored in Chapter 1. We also explore the extent to which additional creditor linkages with the entrepreneurial project, through trade for example, can be used to advantageously manipulate L. Where a creditor depends heavily on project trade, liquidation incentives are likely to be diluted.

As stated at the outset, the central theme of this thesis is that ownership matters. In incomplete contract environments, ownership confers control of assets. Control, in turn, will drive incentives for asset-specific investment. In the following chapters we will explore the consequences of this basic mechanism for project finance, for the optimal boundaries of the firm, and for industry structure and competition.

References

Aghion, P. and P. Bolton (1992): "An Incomplete Contract Approach to Financial Contracting". Review of Economic Studies, 59:473-494.

Bolton, P. and D. Scharfstein (1990): "A Theory of Predation based on Agency Problems in Financial Contracting". American Economic Review, 80:93-106.

Bolton, P. and D. Scharfstein (1993): "Optimal Debt Structure with Multiple Creditors". LSE Financial Markets Group Working Paper 161.

Bolton, P. and M. D. Whinston (1993): "Incomplete Contracts, Vertical Integration, and Supply Constraints". Review of Economic Studies, 60: 121-148

Bonanno, G. and J. Vickers (1988): "Vertical Separation". Journal of Industrial Economics, 36: 257-265.

Coase, R.(1937): "The Nature of the Firm". Economica, 4: 386-405.

Chandler, A. D. (1959): "The Beginnings of 'Big Business' in American Industry". Business History Review, 33: 1-31.

Grossman, S. and O. D. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration". Journal of Political Economy, 94: 691-719.

Hart, O. (1995): "Firms, Contracts, and Financial Structure", Clarendon Press, Oxford.

Hart, O. and J. Moore (1989): "Default and Renegotiation: A Dynamic Model of Debt", LSE Working Paper.

Hart, O. and J. Moore (1990): "Property Rights and the Nature of the Firm". Journal of Political Economy, 98: 1119-1158.

Segal, I. (1995): "Complexity and renegotiation: A Foundation for Incomplete Contracts", Mimeo, Harvard University.

Shleifer, A. and R. Vishny (1992): "Liquidation Values and Debt Capacity: A Market Equilibrium Approach", Journal of Finance, 47:1343-1366.

Vickers, J. (1985): "Delegation and the Theory of the Firm". Economic Journal (Supplement), 95: 138-147.

Wlliamson, O. E. (1985): "The Economic Institutions of Capitalism". Free Press, New York.

Chapter 1

Project Development, Liquidation Values and Multiple Creditor Borrowing

1.1 Introduction

In this paper we consider a situation where a project, initiated by an entrepreneur, requires a cash investment F to go ahead. The entrepreneur has no capital resources of her own and therefore she must obtain external finance if the project is to proceed. A financier will only lend her the required capital if he can ensure sufficient repayment to break even on the loan.

The initial project returns accrue in the first instance to the entrepreneur and cannot be allocated directly to the creditor, since return streams are taken to be unverifiable and hence non-contractible. As a result the entrepreneur must be provided with adequate incentives to voluntarily repay the creditor. The structure of an optimal debt contract provides such incentives.

A key feature of an effective debt contract is the right of the creditor to seize project assets if the borrower (the entrepreneur) defaults on contracted repayments. The potential loss of project control rights discourages voluntary (strategic) default on loan agreements by the entrepreneur. However in poor states of the world low project returns may result in unavoidable (liquidity) default. Loss of control in such circumstances may have adverse effects on entrepreneurial effort investment.

As illustrated in e.g. Grossman and Hart (1986), idiosyncratic asset specific investment may be sensitive to the allocation of asset ownership and control rights when contracts are incomplete. In the current context the potential unavoidable loss of project control in the event of enforced liquidity default will discourage investment in long-term project returns by the entrepreneur. As stressed in a number of recent papers (e.g. Hart and Moore (1989)) future return streams play a crucial role in extracting debt repayments from an entrepreneur. The decrease in future project value through diminished effort investment leads to reduced incentives to meet debt obligations, and possible increases in voluntary (strategic) default.

An ideal debt contract would be sufficiently flexible to permit contract renegotiation and continued entrepreneurial control of project assets in the event of low project returns and liquidity default, and yet rigid enough to secure maximum feasible repayment from the entrepreneur in all states of nature. In the paper we examine how the structure of the optimal debt contract attempts to reconcile these objectives, and in particular how it responds to changes in a key creditor characteristic - the ability of the creditor himself to manage and generate value from liquidated project assets.

When a creditor has limited ability to manage project assets successfully himself the attraction of liquidation depends on the strength of the resale market for project assets. In good states of the world cash rich industry managers will form a healthy market for liquidated project assets. Any limited personal value the creditor has for the assets is then irrelevant. Strong asset resale markets enable the creditor to extract maximum debt repayment in good states, where the entrepreneur is cash rich. Conversely in bad states of the world the market for project assets will be weak since potential buyers are likely to be cash constrained - a point stressed in Shleifer and Vishny (1992). The incentives to liquidate on the part of the creditor are consequently weak, encouraging renegotiation of the original contract with the entrepreneur. The creditor is thus unable to extract significant debt repayment from the entrepreneur in these states. However, in precisely these (bad) states the creditor has little cash to offer in any case. The creditor's commitment to renegotiation when cash is tight enables him to set high debt face value, securing maximum repayment in the good state (when renegotiation is unattractive) while easily preventing liquidity default in the bad state.

The value of project assets for a creditor with considerable ability to manage those assets himself is less dependent on the strength of the market. This enables such a creditor to extract repayment from the entrepreneur even when market conditions are poor. However, entrepreneurial cash constraints are likely to be tight in such conditions and the creditor's incentives to wrest control from the entrepreneur may in fact be too strong. To avoid the consequences of enforced liquidity default (i.e. reduced entrepreneurial effort) the strong creditor may have to lower the face value of the debt, or else provide the entrepreneur with sufficient cash reserves to avert default when project returns are low. Such provision is however costly, reducing the funds available for project finance.

To recapitulate, low creditor project value acts as a commitment to renegotiation when cash is limited, enabling high debt face values to be set. In turn this allows large repayments to be extracted in good states of the world. High creditor asset value enables cash to be extracted in bad states, when project asset markets are weak. But the need to prevent liquidity default limits cash extraction in the good state. These contrasting qualities suggest the merits of borrowing from multiple creditors.

We find that multiple creditor finance dominates single creditor finance when borrowing from creditors with diverse values for project assets is combined. The priority assigned to the various debts is also found to be important. The optimal debt contracts combine low and inflexible senior debt owed to creditors with considerable value for project assets with high, renegotiable junior debt owed to creditors with low value for project assets. The (non-negotiable) senior debt ensures maximum repayment in bad states where returns are low and project asset markets are weak, while the junior debt allows maximum repayment to be extracted in good states (when markets are strong and renegotiation unattractive) but is negotiated to manageable levels in bad, cash constrained states. Such a combination of debt contracts elicits maximum feasible repayment from the creditor in all states of the world, and hence maximises the funds available for project finance.

Those creditors with high values for liquidated assets are likely to be closely involved with the entrepreneur's industry - allowing them to develop the skills to be able to generate some return from the assets. Yet for these creditors spare cash may well be in short supply during bad states of the world (since their industry loans are unable to repay in such states). If these creditors are senior claimants then if the entrepreneur defaults on both loans the assets will pass to them. Their high value for project assets makes renegotiation post-default unattractive (given that in the bad states the entrepreneur has little cash to offer), and thus deters voluntary default. On the other hand if creditors with low liquidation value for the assets are senior claimants the assets pass to them in default, and the high liquidation value creditors get nothing (Cash constraints mean that high liquidation value creditors are unable to "buy out" the senior debt). Low liquidation value creditors are keen to renegotiate the initial debt contracts since they have little value for project assets. Voluntary default is therefore attractive. To extract maximum cash in bad states (and increase feasible lending) creditors with high values for project assets should be senior claimants, discouraging voluntary default. Such debt should however be low, to prevent unavoidable default and liquidation.

Finally the paper considers the merits of obtaining project finance from a trading partner in contrast to situations where the trading and creditor roles are distinct. It is found that the adverse knock-on effects of liquidation (i.e. loss of entrepreneurial project skills) on trading partner profits reduces liquidation incentives where the trading partner is also the creditor. When creditor value for project assets is high, such weakening of liquidation incentives is beneficial - reducing the costs of avoiding enforced, cash constrained default. In such cases the creditor and trading partner roles should be combined, with the creditor taking as large a share of project trade as possible. However for creditors with a low value for liquidated project assets a further weakening of liquidation incentives reduces entrepreneurial debt repayment and hence restricts viable project financing to an even greater extent. In such circumstances the creditor and trading partner roles should remain separated.

In many circumstances a trading partner's ability to manage project assets will increase with the extent of his interaction with the entrepreneur. In this case the greater a creditor's share of project trade the greater the value he can extract from project assets in the event of liquidation. Such an influence may dominate the trade profit effect. It will then be optimal for a creditor with limited intrinsic ability to manage liquidated assets to take on a maximal share of project trade

- reversing the earlier conclusion. The experience effects of increased project interaction then boost rather than diminish the creditor's liquidation value.

A growing number of papers have considered the relationship between optimal debt structure and debt renegotiation in an incomplete contract setting (see e.g. Hart and Moore (1989), Aghion and Bolton (1992)). A number of recent papers have also considered the role of multiple creditors in such a setting e.g. Bolton and Scharfstein (1993), Berglof and von Thadden (1994). None of these papers however has considered the inter-relationships between creditor liquidation values, the resale market for project assets and the propensity of creditors to renegotiate debt contracts when borrowers are in financial distress. Nor do these papers consider the implications of debt structure for long-term human capital investment - a central theme in this paper. This last issue is addressed in a different context by von Thadden (1995).

Hart and Moore (1990) stresses the importance of debt seniority structure, but unlike our analysis they do not focus on the relationship between seniority and post-default debt renegotiation incentives.

Finally, several recent papers have explored the links between finance provision, investment and the allocation of property rights e.g. Aghion and Tirole (1994), Gertner, Scharfstein and Stein (1994). In particular Aghion and Tirole examine the ability of a creditor to extract returns from an entrepreneur when the creditor is also a trading partner (customer) - a theme which we address in this paper too, though in a somewhat different context.

1.2 An Introductory Example

The paper's principal arguments are illustrated in the following simple example.

We will consider an entrepreneurial project that, if initiated, can generate returns over two periods. Date 1 project returns depend on the state of nature. In bad states of nature (probability 1/2) £20 is generated, while in the good state (probability 1/2) the return is £60. The date 2 project return is £100, irrespective of the state of nature.

Project initiation requires a cash injection F at date 0. The entrepreneur has no wealth so the cash required must be borrowed from the (competitive) date 0 finance market. Project returns are non-verifiable, therefore financiers cannot be offered equity shares in the project.

For this simplified example we will assume that entrepreneurial control of project assets must be *feasible* at all dates, in all states of nature. [5]

Feasible debt contracts involve the repayment of debt with face value D at the end of date 1. In the event of default the creditor can take control of project assets, generating liquidation values L_G and L_B in the good and bad states respectively. We assume that the entrepreneur can avoid post-default liquidation by offering the creditor a repayment equal to his liquidation value.

In the cases below we will consider the maximum project finance available when there is (i) a single class of low L_B creditors (Case A), (ii) a single class of high L_B creditors (Case B), and (iii) both high and low L_B creditor classes (Case C).

Case A (Low L_B Creditor Borrowing):

Suppose that the creditor can generate a return of £16 at date 2 from the project assets in the bad state, and £80 from the project in the good state i.e.

L_B = £16

L_G = £80

The optimal debt face value is then D = £60, generating maximum project finance of F = £38.

---

[5] The idea driving this assumption is that the prospective benefits of project control are essential in motivating entrepreneurial activity. Of course, the feasibility of securing project control does not imply that the entrepreneur will always choose to do so.

In the bad state the project provides the entrepreneur with £20 in cash at date 1. Since the debt has a face value of £60 the entrepreneur has no option but to default on the loan agreement. However, to avoid post-default liquidation the entrepreneur need offer the creditor £16 only. This is the value of the creditor's liquidation option. Such a payment ensures the entrepreneur continued control of project assets worth £100 at date 2 - and is therefore always attractive.

In the good state the entrepreneur has £60 in cash available when debt repayment is due. If the debt of £60 is paid in full the entrepreneur secures access to project assets worth £100 at date 2. The default option is not attractive since it leads to liquidation of project assets by the creditor. To see this, note that the creditor's return from liquidation in the good state (£80) exceeds any post-default payment the entrepreneur can offer. The entrepreneur will therefore prefer to pay the debt of £60 in full.

Expected repayments therefore amount to 1/2·£60 + 1/2·£16 = £38. A competitive date 0 finance market ensures that up to £38 will be made available to the entrepreneur for project finance.

It should be clear that where the creditor's *bad state* liquidation value is less than £20 liquidation will never occur in that state, whatever the face value of the debt. The creditor's ability to extract repayment from the entrepreneur is limited by this liquidation value. Bad state repayment (and hence initial cash availability) is therefore increasing in L_B for L_B < £20.

Case B (High L_B Creditor Borrowing):

Suppose now that the creditor's liquidation values are given by:

$L_B = 50$

$L_G = 80$

Feasible debt is then limited to £20, allowing maximum project finance of £20.

Full debt repayment is just feasible in the bad state of the world from date 1 project returns of £20. Such repayment ensures control of assets worth £100 at date 2. Failure to repay the debt results in liquidation - since the creditor's value for project assets (£50) exceeds the cash available to the entrepreneur (£20).

In the good state the entrepreneur has £60 in cash to make scheduled repayments. Payment of £20 guarantees the entrepreneur future project returns worth £100. In the event of default, the entrepreneur cannot retain control of project assets since the creditor's liquidation value L_G = £80 > £60. The debt of £20 is therefore paid in full.

The face value of the debt is limited by the entrepreneur's ability to meet scheduled repayments in the bad state. If the debt were set above £20, entrepreneurial control of the project would then inevitably be lost. This ceiling on debt restricts the cash that can be extracted from the entrepreneur in the good state. For L_B > £20, debt (and initial finance) is therefore limited to £20.

Note that the problems encountered with borrowing from a high L_B creditor can be ameliorated somewhat through the provision of an additional £30 cash reserve to the entrepreneur, beyond the finance used for project investment. She can then avoid liquidation in the bad state, whatever the debt face value, using the cash available at date 1 (£20 in project returns + £30 cash reserve). However, in the good state post-default repayments are limited to £80 - the creditor's liquidation value L_G. Not all date 1 cash (£60 project returns + £30 cash reserve) can therefore be extracted in the good state. Maximum project finance can be boosted in this way to F = 1/2·£80 + 1/2·£50 - £30 = £35.

Cases A and B together highlight the **non-monotonic** relationship between creditor liquidation incentives and maximum project finance.

Case C (Two Creditor Borrowing):

Finally consider entrepreneurial borrowing from both high and low L_B creditors. Optimal project financing then involves borrowing £4 from the high L_B creditor and £36 from the low L_B creditor - a total of £40. High L_B creditor debt is set at £4 and is senior to the low L_B debt which is set at £56.

In the bad state the entrepreneur has £20 cash available to repay debt with a total face value of £60. Clearly the entrepreneur cannot meet both debt repayments in full. Failure to repay the senior debt will result in certain liquidation, since that creditor's value for project assets (£50) exceeds the cash available to the creditor (£20). If the entrepreneur repays the senior creditor in full she is left with £16 from date 1 returns. This is just sufficient to avoid liquidation by the junior creditor. Two creditor borrowing therefore results in full project cash extraction in the bad state.

In the good state the entrepreneur has £60 in available cash, and can just meet scheduled debt repayments. Default in not attractive since a minimum of £80 must then be paid to one or both creditors to avoid liquidation. Scheduled debt repayment ensures that all date 1 project cash is extracted in the good state too.

The combination of debt contracts allows full extraction of date 1 project returns in both good and bad states. Project finance is therefore available up to a maximum of F = 1/2·£60 + 1/2·£20 = £40.

We have seen in the single creditor case that low L_B debt is open to successful renegotiation in the bad state of nature. This enables such debt to be set high, extracting maximum project cash in good states, while guaranteeing continued entrepreneurial control of the project in the bad state. The down side of low L_B debt is its inability to extract entrepreneurial cash in the bad state, precisely because of this amenability to renegotiation.

High L_B debt has complementary properties. High L_B creditors will be reluctant to renegotiate with the entrepreneur in the event of bad state default (since liquidation returns are high). This inflexibility enables maximum cash to be extracted in the bad state. To prevent unavoidable liquidation such power must be restrained. To do this high L_B debt must be set low to guarantee repayment feasibility in all states. This debt ceiling limits cash extraction in good states.

A combination of high and low L_B debt optimally aligns these contrasting attributes, enabling maximum cash extraction in both good and bad states.

1.3 The Model

At the heart of the model is an entrepreneur with an idea. For the project to proceed an essential, indivisible asset must be purchased at cost F. Since the entrepreneur has no money of her own she must obtain finance from the capital market if the project is to go ahead.

Once it is up-and-running, the project generates returns over two periods - date 1 and date 2.

Date 1 returns v_1 depend solely on the state of nature :

$$v_1 = \begin{cases} X_G & \textbf{probability } p_G \\ X_B & \textbf{probability } p_B \end{cases}$$

The magnitude of date 2 returns v_2 is determined by the level of entrepreneurial effort. This effort is undertaken at date 1/2. Generating date 2 returns Y involves a private effort cost, C(Y), incurred at date 1/2.

Assumption 1.1: $C(0) = 0, \quad C' > 0, \quad C'' > 0$

Let $Y^* = \text{argmax} \;\; Y - C(Y)$

In a first best world finance F would be available provided:

$$F \leq p_G X_G + p_B X_B + Y^* - C(Y^*)$$

Two additional assumptions regarding the magnitudes of date 1 and date 2 returns are made:

Assumption 1.2: $0 < X_B < X_G \leq Y^*$

Assumption 1.3: $Y^* - C(Y^*) > p_G X_G + p_B X_B$

Assumption 1.2 ensures that if optimal investment is undertaken by the entrepreneur, long term (date 2) project returns exceed short term (date 1) returns, in all states of the world. Assumption 1.3 states that provided optimal entrepreneurial effort has been undertaken at date 1/2 then long term project returns even net of effort costs exceed expected date 1 returns.

Only the entrepreneur can initiate the project and get it "up and running". Full entrepreneurial control in the early stages of the project is therefore essential. As a result date 1 project returns are specific to the entrepreneur, and accrue in the first instance to her. However, once entrepreneurial effort has been undertaken and the project is under way it can be run by a substitute manager. Some fraction of date 2 returns can therefore be extracted without continuing entrepreneurial involvement in the project - a detailed discussion will follow.

For project development to proceed external finance is required. A competitive market of identical financiers is assumed to exist at date 0. These financiers have unlimited access to capital and to 0 NPV investment projects. There is, however, a shortage of investment projects generating positive profit opportunities.

Although observable to all active parties, project returns are not verifiable to a court of law. As a result the terms of any contract between the entrepreneur and a financier cannot be made conditional on actual project returns. In particular, control of project assets cannot be allocated directly on the basis of realised date 1 returns. Neither can reward for entrepreneurial investment be made on the basis of realised returns. It is also assumed that the human capital investment made by the entrepreneur is not contractible. No contracts can therefore be written directly rewarding the entrepreneur for undertaking specified investment.

Since the returns from the project are unverifiable, and accrue to the project manager in the first place, financiers cannot be offered an equity share in the project. We will assume however that monetary transactions between the entrepreneur and the financier are verifiable, and hence can be written into the terms of a contract.

The financier will be prepared to supply essential start-up cash to the entrepreneur provided he can expect to break even on the loan. The financier must therefore be assured that either the entrepreneur will make voluntary re-payments, or that he will have sufficient access to and control of valuable project returns to cover his initial outlay.

Of particular relevance are simple debt contracts involving debt face value D to be repaid at the end of date 1. Failure to repay the specified amount constitutes a default on the agreement. Post-default, the following procedure is adopted: The entrepreneur can make a single take-it-or-leave-it offer to the creditor. If this is accepted, the entrepreneur makes the revised payment and keeps control of project assets. If it is not accepted, control of project assets passes to the creditor. The entrepreneur thus has all the bargaining power in any post-default negotiations with the creditor.

As a considerable simplification we will not allow debt contracts that involve randomised liquidation in the event of default.[6]

A financier can generate returns $L_i(Y)$ in state i (i = G,B) by managing the project himself at date 2 or by selling project assets, provided entrepreneurial investment Y has been undertaken at date 1/2.

Assumption 1.4: $L_i(Y) \leq Y$

To simplify notation, let $L_i = L_i(Y^*)$

The liquidation value L_i is a measure of the financier's ability to manage, and generate a return from project assets. This ability to generate date 2 value from project assets is never greater than that of the entrepreneur herself. Since the entrepreneur initiated the project it is assumed that no party can generate a

[6] In particular this rules out debt contracts where the extent of fractional liquidation is dependent on repayments made. Such contracts can be used to discourage voluntary liquidation, while ameliorating the consequences of liquidity default somewhat - see e.g. Bolton and Scharfstein (1993).

greater return from the project than the entrepreneur herself. When the entrepreneur is replaced as manager the loss of her human capital is likely to lead to at least some reduction in project value.

It will be informative to adopt the following additional structure for liquidation payoffs.

We will assume that the creditor has the right to sell any assets he controls (following default by the entrepreneur). Indeed we will further assume that at date 1 there is a competitive market of project managers willing to take on the project. Once the project is under way, it is assumed that these managers can generate as much value from the project as the original entrepreneur. However, they cannot initiate projects - entrepreneurial innovation skills are a scarce resource.

In good states of the world these managers have unlimited financial resources, but in bad states of the world their cash is tied up. This last assumption is crucial. The idea here is that the fortunes of industry "insiders" - the potential substitute managers - are correlated with those of the entrepreneur himself. When the entrepreneur's project returns are low (bad states of the world), the substitute managers are cash constrained too. By contrast, when date 1 project returns are high (good states of nature) industry insiders are in a strong financial position - with access to plentiful cash reserves - guaranteeing vigorous demand for project assets.

The market value M_i for project assets (assuming optimal entrepreneurial investment) is therefore given by:

$$M_i = \begin{cases} Y & \text{if } i = G \\ 0 & \text{if } i = B \end{cases}$$

To reiterate, the idea here is that potential substitutes for the entrepreneur i.e. agents with the requisite skills to take over and run the project are likely to come from related businesses. Shocks that adversely affect the entrepreneur are therefore likely to affect the replacement manager's prospects too. This structure is in the spirit of Shleifer and Vishny (1992).

Whenever project asset markets are weak i.e. in the bad state, the financier must rely on his own ability to manage the project to generate value from liquidated assets. The private value of project assets to the financier himself, assuming entrepreneurial investment at date 1/2, is given by λY ($\lambda \leq 1$).

Thus $L_i(Y) = \max[\lambda Y, M_i]$

As a further simplification we will assume that $Y^* = X_G$. This allows us to focus on events in the bad state of the world, and on a single key creditor characteristic - his private value for project assets (measured by λ).

In this case, the financier's liquidation values in bad and good states are therefore given by $L_B(Y) = \lambda Y$ and $L_G(Y) = Y$ respectively.

1.4 Single Creditor Analysis

We will begin our analysis by supposing that the date 0 capital market consists of a single class of identical potential financiers. In particular, all financiers share a common ability to manage project assets in the bad state - measured by parameter λ. We will focus on the following issue: What is the largest project finance requirement F that can be met, and what financial structure supports it?

Note that for any project finance to be forthcoming the creditor must be certain that effort will be undertaken by the entrepreneur at date 0. Without such effort the date 2 value of project assets is 0. If this is the case then there is no value to the entrepreneur in securing date 2 control of project assets. Consequently no incentive will exist to make any debt repayments at the end of date 1. Furthermore, if no effort is undertaken at date 0 (implying $Y = 0$) the date 2 liquidation value of project assets to the creditor is 0.

Lemma 1.1:

No project finance will be forthcoming at date 0 unless financiers can ensure effort will be undertaken by the entrepreneur at date 1/2.

Contractual incompleteness prevents direct reward for effort. In fact, such effort can only be motivated by giving the entrepreneur control of date 2 project returns. The entrepreneur will then benefit directly from increased effort through larger date 2 returns. The greater the probability that the entrepreneur will lose control of project assets at the end of date 1, the smaller her expected date 2 return on effort, and hence the weaker will be her effort incentives.

To secure control of project assets for date 2 the entrepreneur must either avoid default by paying the debt D in full, or else she must offer the creditor his state-dependent liquidation value L_i during post-default debt renegotiation.

In general it will be optimal to provide the entrepreneur with a cash reserve R at date 0 to facilitate debt repayment or renegotiation.[7] For now we will ignore this possibility, but will return to consider it in some detail below.

Lemma 1.2:

For a debt with face value D to be consistent with continued entrepreneurial control of project assets in state i, the following condition must be satisfied:

$$\min[D, L_i] \leq X_i$$

If the entrepreneur can avoid liquidation in state i, she always will do, as the date 2 gains always exceed the largest possible repayment needed to secure them ($L_i(Y) \leq Y$).

[7] Hart and Moore (1989) analyse the trade-off between high debt plus cash reserve ("pure transfer") and low, manageable debt ("pure debt") in some detail.

A creditor will only be prepared to provide the entrepreneur with project finance if he can expect to break even on the loan. Expected repayments must therefore cover the initial loan F i.e.

$$p_G \min[D, L_G] + p_B \min[D, L_B] \geq F$$

Competitive pressure in the date 0 capital market will ensure that this condition holds with equality.

Ensuring that the entrepreneur can secure date 2 control of the project in all states of the world does not of course guarantee that she will choose to do so. The benefits of such control are the date 2 returns generated by entrepreneurial effort. The costs of such returns are not just the direct costs of effort, but also the cost of the debt repayments needed to secure access to such returns. Effort will be undertaken and repayments made provided:

$$Y^* - C(Y^*) - p_G \min[D, L_G] - p_B \min[D, L_B] \geq 0$$

For liquidation to be avoidable in all states we know from Lemma 1.2 that:

$$p_G X_G + p_B X_B \geq p_G \min[D, L_G] + p_B \min[D, L_B]$$

Assumption 1.3 then ensures that first best effort will always be undertaken if date 2 control of project assets can be assured.

Alternatively, the optimal finance contract could involve the entrepreneur maintaining control of the project in state i only. In this case liquidation will occur in state j. The creditor's expected return on his initial lending is then given by:

$$p_i \min[D, L_i] + p_j L_j$$

This expected return must cover the initial loan F + R.

To encourage crucial entrepreneurial investment, for which the entrepreneur will reap a reward in state i only, debt must be limited to ensure that:

$$p_i Y_i - C(Y_i) - p_i \min[D, L_i] \geq 0$$

where $Y_i = \text{argmax } p_i Y - C(Y)$

If the following condition holds then liquidation in one state alone will never occur.

Condition 1A:

Let $Y_i = \text{argmax} \;\; p_i Y \; - \; C(Y)$. Then:

(i) $\quad p_G Y_G \; - \; C(Y_G) \; - \; p_G X_B \;\; \leq \; 0$

(ii) $\quad Y_B \; \leq \; X_G$

The first part of Condition 1A states that, if the entrepreneur is to be encouraged to invest when liquidation in the bad state is inevitable, the debt face value must be set so low that in fact default can be avoided in the bad state too. Similarly, the second part of the condition states that default in the good state alone will not occur. Where continued project control is feasible it will always be chosen by the entrepreneur.

Since effort must be attractive for finance to be forthcoming we can therefore restrict our attention to the case where date 2 entrepreneurial control of the project can be assured in all states of the world. The issue is then simply one of maximising the finance available at date 0:

$$\max_{D} \quad F = p_G \min[D,L_G] + p_B \min[D,L_B]$$

$$\text{s.t.} \quad \min\,[D,\, L_i] \leq X_i$$

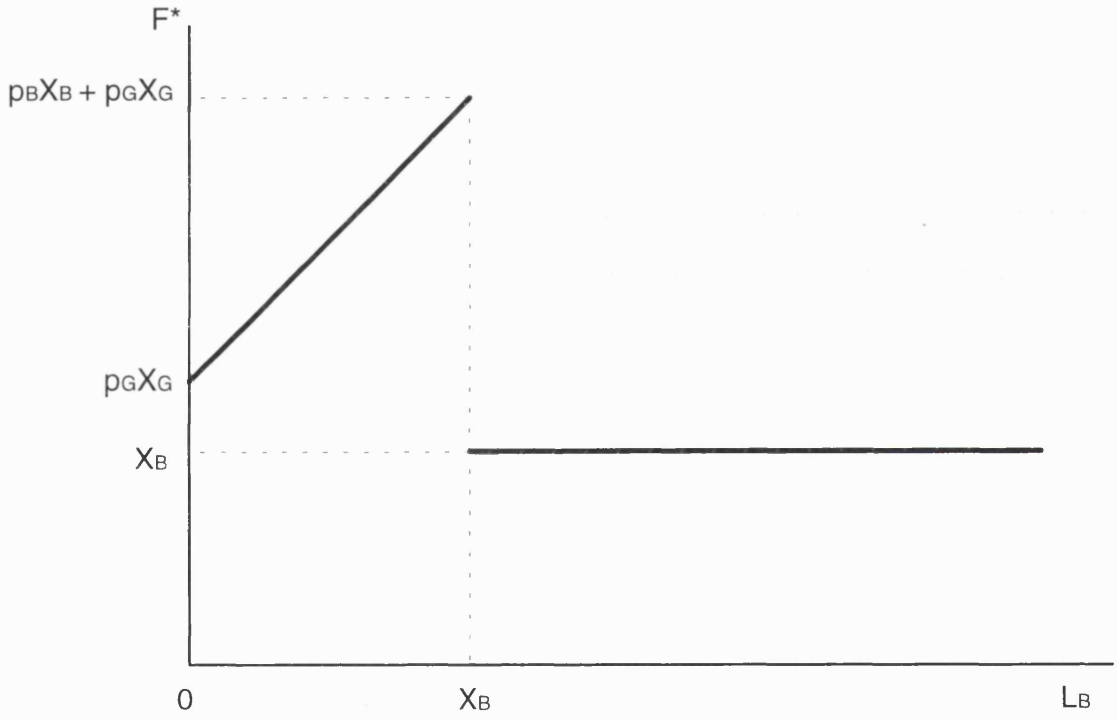
The solution to the above problem is as follows:

$L_B \leq X_B$	$D = X_G = Y$	$F^* = p_B L_B \; + \; p_G X_G$
$L_B > X_B$	$D = X_B$	$F^* = X_B$

The relationship between F^* and L_B is illustrated in Figure 1.1.

When $L_B < X_B$ the entrepreneur can and always will avoid liquidation in the bad state, whatever the face value of the debt. She can secure continued control of project assets by offering the creditor his liquidation value L_B post-default. In effect, a low value for L_B acts as a commitment on the part of the creditor to renegotiate the debt contract in bad states. This enables the creditor to set a high

Figure 1.1



debt face value D and extract maximum cash from the entrepreneur in the good state, when project returns are high. Note that in the good state a healthy asset resale market ensures $L_G = Y$. Of course, such renegotiation possibilities limit the cash that the creditor can extract in the bad state. However the entrepreneur has little cash to offer in this state in any case.

When $L_B < X_B$ a marginal increase in L_B will not affect the creditor's commitment to renegotiate debt to manageable levels in bad states of nature. However a higher L_B allows greater cash extraction in precisely these states, where the creditor relies on his liquidation threat to force repayments from the entrepreneur. The prospect of higher expected repayments ensures greater finance is available at date 0. F^* is therefore increasing in L_B for $L_B < X_B$.

Without entrepreneurial effort, debt repayments at date 1 will never occur. Yet without the prospect of repayment no creditor will be prepared to provide finance at date 0. To encourage optimal effort, the entrepreneur must be guaranteed date 2 project control in all states of the world. When $L_B > X_B$ this can only be achieved by setting the face value of the debt at an always repayable level $D = X_B$. If the debt is set any higher, default is inevitable in the bad state and liquidation will follow, since the creditor's liquidation value (L_B) exceeds the cash available to the entrepreneur (X_B). Of course, a low debt face value limits the cash that can be extracted from the entrepreneur in *good* states, when date 1 project returns are high. This restricts the finance available at date 0.

Cash Reserve Provision (R > 0)

Suppose now that the entrepreneur can provide a cash reserve R in addition to project finance F at date 0. It is easy to see that such a reserve may be optimal. A cash reserve R adds to the funds available to the entrepreneur at date 1 and thus relaxes any cash constraints faced at that time, when debt repayment is due. Of course this extra cash reserve is lent by the creditor in anticipation of future repayment. Expected repayments at date 1 must therefore rise, or finance F must be reduced, as R is increased.

To find the maximum finance available in this new setting, with cash reserve provision, we must solve the following optimisation problem:

$$\max_{R,D} \quad F = p_G \min[D, L_G] + p_B \min[D, L_B] - R$$

$$\text{s.t.} \quad \min[D, L_i] \le X_i + R$$

$$Y^* - C(Y^*) - (F + R) \ge 0$$

The relationship between maximum project finance F^* and the key lender characteristic L_B is illustrated in Figure 1.2.

Proposition 1.1

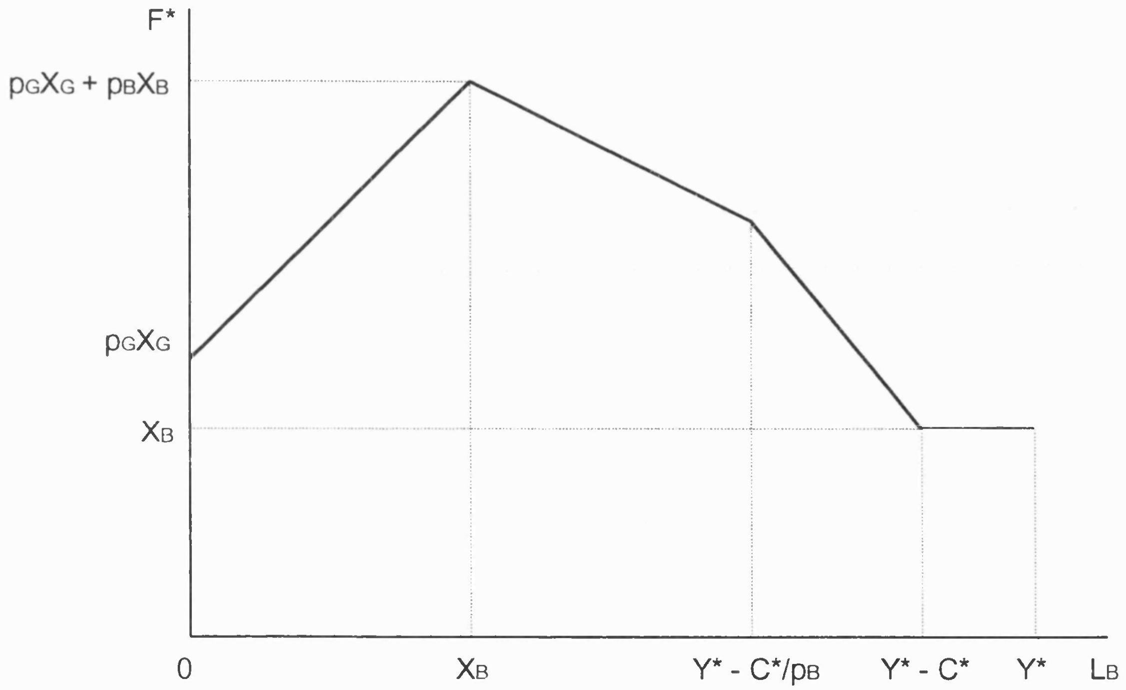
For creditors with low private value for project assets ($L_B < X_B$) maximum project finance F^* is strictly increasing in L_B, while for higher L_B values ($L_B \ge X_B$) it is non-increasing. Indeed for $X_B < L_B < X_G - C^*$ maximum finance available is strictly decreasing in L_B. The relationship between F^* and L_B is therefore non-monotonic.

For creditor's with $L_B < X_B$ a cash reserve is of no value in increasing project finance. The problem then is one of extracting cash from the entrepreneur in bad states of the world. Providing the entrepreneur with additional funds merely exacerbates this problem. Zero cash provision, as assumed above, is optimal.

With $L_B > X_B$ project returns alone are insufficient to avoid liquidation in the bad state unless debt is set low ($D = X_B$). For higher debt to be consistent with no post-default liquidation, the entrepreneur must have her project income supplemented. If the entrepreneur is given a cash reserve $R = L_B - X_B$ at date 0, she can avoid liquidation post-default in the bad state. Debt can then be set high ($D > X_G$), extracting maximum cash in the good state.

In the good state, project returns alone are sufficient to avoid liquidation at the end of date 1 (since $L_G = Y^* = X_G$). Every £1 of cash reserve can therefore be retained by the entrepreneur. The greater is L_B, and therefore R, the larger the loss to the creditor from this reserve provision in the good state. For given debt

Figure 1.2



repayments, the project finance available at date 0 must therefore fall.

If the creditor provides the entrepreneur with project finance F and cash reserve R, debt must be structured such that expected repayments cover this loan in equilibrium. This of course means that in order to benefit from date 2 project returns the entrepreneur must not only incur effort costs C(Y) but also make expected repayments F + R to get access to those returns. Investment will only be attractive if:

$$Y^{*} - C(Y^{*}) - (F + R) \geq 0$$

Note that if any effort is attractive it will be the first best level. As L_B increases the cash reserve R necessary to avoid liquidation in the bad state, with high debt face value increases. Eventually the constraint must bind, limiting total repayments. Further increases in R can only be accommodated by decreases in F and in debt face value D. Eventually a point is reached where $L_B = D$ and $F = X_B$. Further increases in L_B then have no effect on F.

In the Appendix we prove a version of Proposition 1.1, for the more general case where $Y^* \geq L_G$ and $Y^* \geq X_G$. The basic structure of our results carries over to this more general case. Two sources of difference, however, are worthy of comment.

First, note that the general case allows $L_G < X_B$. Setting debt face value at $D = X_B$ is not then restrictive. In the bad state, the creditor can never hope to extract more than $\min[L_B, X_B]$, while in the good state repayment will be limited to the liquidation value L_G ($< D = X_B$). Clearly, project finance is then increasing in L_B for $L_B < X_B$ and *non-decreasing* in L_B for $L_B \geq X_B$. The advantage of a commitment to renegotiate debt in the bad state (provided by low L_B) is the ability to set high debt in the good state. With $L_G < X_B$, low L_B commitment is redundant, and merely reduces potential repayment in the bad state.

If $L_G > L_B > X_B$, it is optimal to provide the entrepreneur with a cash reserve $R = L_B - X_B$, provided effort investment is sure to be undertaken. In the general case, where $Y^* > X_G$ is entirely possible, it may be that $L_G > X_G + R$, provided $R = L_B -$

X_B is not too large. In such circumstances, the creditor can extract all date 1 project returns from the entrepreneur *and* recover all cash reserves provided at date 0, while also ensuring date 2 entrepreneurial control of project assets. Maximum feasible project funding ($F^* = p_G X_G + p_B X_B$) is then available for a *range* of values of $L_B \geq X_B$.

1.5 Multiple Creditor Borrowing

As the analysis in the previous section has demonstrated, a creditor with a given λ (ability to manage project assets) and hence L_B has particular advantages and disadvantages in his ability to finance project development.

Low L_B creditors are flexible to renegotiation of debt contracts. Low renegotiated payments will be accepted in bad states of the world, however high the face value of debt. This flexibility allows such investors to set high debt face value, extracting maximum repayment in good states of the world (since a healthy asset resale market then exists). The relative weakness of low L_B creditors is their inability to extract project cash in the *bad* state, when asset resale markets are weak.

A creditor with very high L_B was seen to have contrasting strengths and weaknesses. A high L_B value enables the creditor to extract all available cash in the bad state. However, this strength may well be excessive. The creditor may have too great an incentive to wrest control of the project from the entrepreneur. To counter this tendency, the creditor must either set low debt face value (guaranteeing the entrepreneur's ability to repay), or he must provide the entrepreneur with sufficient reserves to cover shortfalls in the bad state. In either case the outcome is that high L_B creditors are relatively ineffective at extracting cash from the entrepreneur in the *good* state of the world, when the entrepreneur is cash rich.

As shown above these opposing effects are only perfectly resolved for a unique creditor with $L_B = X_B$. In general however such an ideal creditor will not exist. The question we now address is whether multiple creditors can be employed to effectively balance these forces more generally.

Suppose that there are two non-identical creditors - lender α and lender β. These creditors have personal values for the project assets of L_B^α and L_B^β respectively, where we will assume $L_B^\alpha > L_B^\beta$ without loss of generality. We will further suppose that lender j provides F_j of investment cash and an amount R_j of cash reserve to the entrepreneur at date 0. The face value of lender j's debt is D_j.

Crucially we will assume that in the bad state of the world the high L_B creditor at least is cash constrained. Without this assumption side contracting between creditors will destroy any multiple creditor effects (hence such an assumption lies at the heart of the literature on this topic). Since a high L_B creditor is likely to be an industry insider and hence affected by the same shocks as the entrepreneur, this assumption may not be unreasonable.

In the event of default on a single loan, procedures are as in the basic single creditor model. However, it is possible that the entrepreneur will now default on two loans simultaneously. In such circumstances the following procedure is adopted:

Following default the entrepreneur can put a single take it or leave it offer of revised repayments to both creditors. If the revised schedule is accepted by *both* creditors then the new payments are made and the entrepreneur maintains control of project assets. If the revision is rejected by either creditor then liquidation proceeds. Project assets pass first to the junior creditor. He must then either clear the senior creditor's debt or hand over the project assets to that creditor.[8]

[8] Note that this allocation of rights to creditors on liquidation is analogous to the revised bankruptcy procedures proposed by Aghion, Hart and Moore (see Hart (1995), Chapter 7 for a discussion). Under their procedure, control of liquidated project assets passes first to the senior creditor. However, the junior creditor has the right to purchase this control at a price equal to the value of the senior debt. The Aghion-Hart-Moore approach generalises to multiple creditor classes, with multiple creditors in each class.

Again we wish to find the set of debt contracts that maximise the funding available for project finance.

If both potential creditors have low L_B values i.e. $L_B^\beta < L_B^\alpha < X_B$ then it is easily seen that the maximum finance available for project investment is given by:.

$$F = p_G X_G + p_B L_B^\alpha$$

This is no more than that obtained from the high L_B creditor alone, setting debt face value $D_\alpha = X_G$. Essentially, repayment in the bad state is limited to at most the liquidation value of the senior creditor (a maximum of L_B^α).

It should be clear from the analysis in Section 1.4 that when $L_B < X_B$ project finance is maximised by ensuring that the creditor is as tough as possible in the event of default in bad states of the world. Introducing additional creditors with even lower L_B values cannot lead to greater cash extraction.

Suppose now that both potential creditor types have high L_B values i.e. $L_B^\alpha > L_B^\beta > X_B$. In this case the maximum project start-up that can be financed is given by:

$$F = p_G [X_G - (L_B^\beta - X_B)] + p_B X_B$$

This is no more than the finance available from the low L_B creditor alone. With $L_B > X_B$ the creditor is excessively tough, as the analysis in section 1.4 makes clear. An additional cost is then incurred in securing essential entrepreneurial control of project assets. Adding an additional creditor with even tougher liquidation incentives cannot improve the situation in our framework.

Finally we can consider the case where one of the creditors has a low L_B value, and one a high L_B value i.e. $L_B^\beta < X_B < L_B^\alpha$. In this case the optimal debt structure is given by:

$D_\beta = X_G - D_\alpha$ $\qquad R_\beta = 0$ $\qquad D_\alpha$ priority over D_β.

$D_\alpha = X_B - L_B^\beta$ $\qquad R_\alpha = 0$

Such a debt structure supports maximum project finance F given by:

$$F = p_G X_G + p_B X_B$$

where $F_\alpha = X_B - L_B^\beta$ and $F_\beta = p_G (X_G - X_B) + L_B^\beta$. Note that no greater F is possibly feasible, given the incomplete contracts assumptions of the model.

The total face value of combined debt $D_\alpha + D_\beta = X_G$. In the good state neither contract will be renegotiated. A healthy market for project assets ensures that liquidation values are high, and consequently default is unattractive. Both debt agreements will be honoured in full by the entrepreneur, allowing all date 1 project returns (X_G) to be extracted.

In the bad state of the world the entrepreneur cannot avoid default on at least one of the loan agreements, since the cash available to her is limited ($X_B < X_G = D_\alpha + D_\beta$). In fact the entrepreneur will pay the senior debt, $D_\alpha = X_B - L_B^\beta$, in full. Default on the junior debt is then inevitable. Liquidation can, however, be avoided if the entrepreneur offers the junior creditor all her remaining cash (L_B^β). This is just sufficient to ensure successful renegotiation of the junior debt. Note that default on the senior debt is never attractive since the entrepreneur then has insufficient cash to avoid liquidation ($X_B < L_B^\alpha$). Once again, date 1 project returns (X_B) can be extracted in full.

It should be clear that the above structure dominates one where the low L_B financier is senior creditor. In that case default on both loan agreements is optimal in the bad state. Liquidation can then be avoided by offering the senior creditor only a cash payout of L_B^β ($< X_B$). To see this note that if liquidation were to proceed the cash constrained junior creditor could not clear the senior debt. Liquidated assets would therefore pass to the senior creditor yielding him a return L_B^β. The junior creditor would get nothing.

It can never be optimal to raise the face value of the senior debt D_α. Liquidation in the bad state could then not be avoided, since $D_\alpha + L_B^\beta > X_B$ and $L_B^\alpha > X_B$, unless the entrepreneur were provided with an additional cash reserve $R > 0$ or the junior debt D_β were reduced. Since cash reserves are not needed for liquidation-avoiding repayment in the good state, they can be retained by the entrepreneur. Consequently, the creditor makes a loss in expected terms on such reserves. Project finance must therefore be reduced if creditors are to break even. Alternatively, if the junior debt is reduced it must be to a level such that $D_\alpha + D_\beta \leq X_B$, to avoid liquidation in the bad state. Clearly repayment in the good state too is then restricted to X_B, again reducing project finance. Reducing D_α can never raise project finance since then repayment in the bad state = $D_\alpha + L_B^\beta < X_B$. Increasing D_β has no effect since post-default renegotiation of the junior debt will leave equilibrium repayments unchanged.

The working of this contract is simple. The senior debt is set low (so that it can be repaid in all states of nature) but is never renegotiated, ensuring maximum cash extraction in the bad state. The junior debt owed to the low L_B creditor is set high (to extract maximum repayment in the good state, where L_B is irrelevant) but is renegotiable in the bad state (since junior creditor liquidation value is low), facilitating continued entrepreneurial project control.[9]

The results of this section are summarised in the following proposition:

Proposition 1.2

(i) Multiple (two) creditor borrowing yields greater funding for project finance than single creditor debt if and only if one creditor has a relatively low asset liquidation value while the other has relatively high asset liquidation value - more precisely if $L_B^\beta < X_B < L_B^\alpha$.

(ii) Under the terms of the optimal contracts, senior debt is owed to the high L_B creditor, is set low and is never renegotiated. Junior debt is owed to the low L_B creditor and is renegotiated in the bad state.

[9] Note that, unlike the analysis in Bolton and Scharfstein (1993), it is the contrasting characteristics of creditors rather than their number that drives the optimality of multiple creditor borrowing.

1.6 Trading Partner Borrowing

Throughout the preceding analysis we have assumed that all returns from the entrepreneurial project accrue to the agent in control of project assets. Of course, one or more creditors may also gain a share of those returns - but only indirectly through entrepreneurial debt repayment. In general, however, the project might be expected to generate direct returns for a number of other parties - employees, suppliers, customers, etc. In this section we will focus on the role of a generalised group of project trading partners. In particular we wish to address the following question: Should the trading partner and creditor roles be combined?

We will assume that an overall trading partner profit of T is generated in the event of full entrepreneurial control of the project. For simplicity we will assume that this trade profit is realised at date 2 only, and is directly proportional to date 2 core project returns. An individual trading partner's profit is proportional to his share of project trade.

We will begin by assuming that all trading partners share a common ability to manage project assets. In particular each can personally generate date 2 profits L_B ($\leq Y$) from core project assets. When a trading partner takes control of the core project we will assume his limited ability to manage project assets inflicts a proportional effect on trading partner profits i.e. they are reduced by a factor L_B/Y.

A competitive, cash rich pool of potential trading partners is available at date 0. Each is prepared to offer the entrepreneur a cash advance to secure future trading partner profits. A trading partner will also lend the entrepreneur additional cash provided he can expect to break even on this loan. To abstract from the multiple creditor issues raised in the previous section, we will assume that the entrepreneur borrows from a single trading partner, involved in a fraction σ of project trade.

Combining the trading partner and creditor roles has no impact in the good state. In the event of default, project assets can be sold to cash rich substitute managers for their full date 2 value. These managers are fully able to manage project assets, given initial investment by the entrepreneur. In particular, the value of project trade will remain unaffected by such control switches. To prevent post-default liquidation, the entrepreneur must offer the creditor the full value of date 2 project returns.

Events in the bad state are rather different. Suppose the entrepreneur defaults on loan repayments at date 1. If the trading creditor then takes control of project assets, his overall return is:

$$L_B + \sigma T\,[L_B / Y]$$

By managing project assets himself the creditor can generate core project returns L_B at date 2. However, his limited management ability also has a knock-on effect on the value of trade profits.

Alternatively, the creditor can accept a cash offer x from the entrepreneur made in return for her continued control of the project. If this offer is accepted overall creditor revenue is given by $x + \sigma T$ - the immediate cash payment plus the full value of date 2 trade profits.

A cash offer x will secure post-default entrepreneurial control of the project if:

$$x + \sigma T \;\geq\; L_B + \sigma T\,[L_B/Y]$$

i.e. if $x \geq \max[0, L_B - \sigma T\,[1 - L_B/Y]]$ (-we restrict x to being non-negative).

It is immediately apparent that controlling a share σ of project trade effectively reduces the creditor's liquidation value in the bad state from L_B to:

$$\max[0, L_B - \sigma T[1 - L_B/Y]]$$

In deciding whether to liquidate project assets, the trading creditor explicitly considers the effects on trade profit of entrepreneurial separation from project assets, whereas a non-trading creditor would not. In taking control of project assets, and replacing the entrepreneur with a less effective manager (- the creditor himself!), the trading creditor reduces the value of his own trade profit. This trade effect discourages liquidation.[10]

Figure 1.3 illustrates the relationship between F* and L_B in the trading creditor case, for the extreme case of complete creditor - trading partner separation (σ = 0) and unified creditor - trading partner roles (σ = 0). The situation is as in Figure 1.2, but now a given value L_B corresponds to an effective liquidation value of max[0, L_B - σT [1 - L_B/Y]].

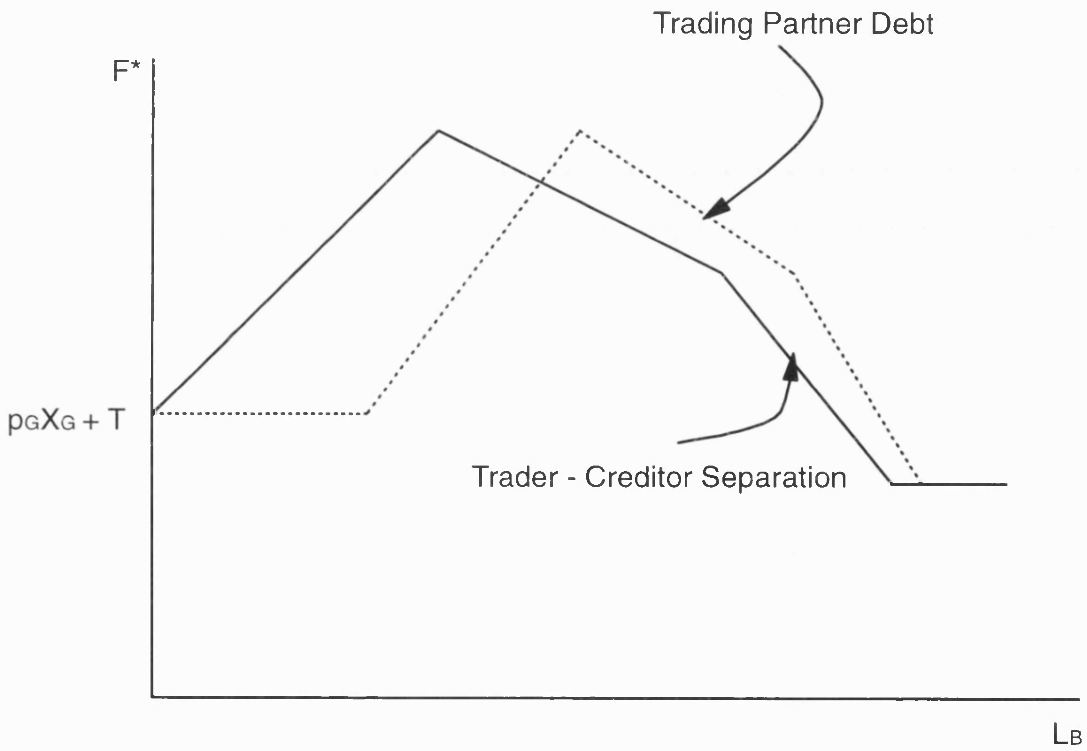
Where the creditor has low L_B value his ability to extract cash from the entrepreneur in the bad state is already limited by his low liquidation value. Introducing a further discouragement to liquidation further reduces this ability. As a result the funding available for project finance is reduced still further.

For creditor's with large L_B values a share of project trade increases the funds available for project finance. Such creditors' liquidation tendencies are too great in bad states of nature. In order to ensure continued entrepreneurial control of project assets in bad states of the world, and hence encourage entrepreneurial effort, such creditor's must limit their ability to liquidate project assets. This is achieved by setting low face value for debt, or else by providing the entrepreneur with a cash reserve which can be used to avoid liquidation. In either case such measures limit the extraction of cash from the project in good states of nature. Combining loan provision with a share of project trade in such cases, by reducing liquidation incentives, alleviates these problems. In particular since the trade effect is specific to the bad state it does not impair cash extraction in the good state.

The greater the value of project trade to the combined trading partner - creditor the weaker his liquidation incentive in the event of default in bad states of the world. The magnitude of this effect will vary as the creditor undertakes more or

[10] A somewhat related point is made in Aghion and Tirole (1994), where the links between the trading relationship and financing incentives are also explored. There, a buyer who holds equity in the supplier firm will bargain less aggressively over supply terms, since it regains a fraction of any lost cash via its equity share.

Figure 1.3


Trading Partner Debt
F*
pGXG + T
Trader - Creditor Separation
LB

less of the trading partner business.

The effective liquidation value $\hat{L}_B$ is given by:

$$\hat{L}_B = L_B - \sigma T [1 - L_B/Y]$$

To limit values of $\hat{L}_B$, a higher private value for project assets (higher liquidation incentives) must be offset by a greater share of trading partner business (lower liquidation incentives). The shape of the resulting iso-$\hat{L}$ curves in (L_B, σ) is illustrated in Figure 1.4.

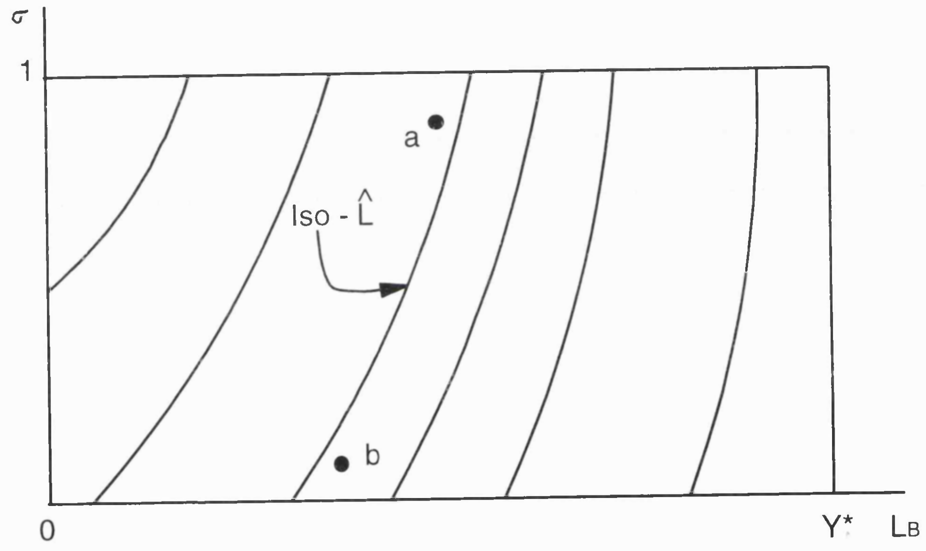
As the creditor's private value for project assets, L_B, increases an offsetting increase in trade share σ is required to maintain constant liquidation incentives. The greater the creditor's share of trading partner business the more he loses (via reduced trade profit) if the entrepreneur is separated from project assets. However, as L_B increases the extent of these losses diminishes since the creditor becomes increasingly able to manage core project assets himself. The offsetting share increments must therefore increase with L_B to effectively limit liquidation incentives.

As outlined in Proposition 1.1 the relationship between bad state liquidation value L_B and maximum project finance F^* is non-monotonic. For values of $L_B < X_B$, F^* is increasing in L_B. In such circumstances L_B should be set as large as possible. Clearly this involves a complete separation of trading partner and creditor roles (re. Figure 1.2). Conversely when $L_B > X_B$ project finance is maximised by reducing the creditor's effective project liquidation value. When L_B is very large this involves the creditor taking the largest possible share of project trade (σ = 1). For creditors with intermediate L_B values [$X_B < L_B < Y^*(X_B + T)/(Y^* + T)$] the optimal value of $\hat{L}_B = X_B$ can be achieved. The optimal share of trade value is then given by:

$$\sigma = \frac{Y^* (L_B - X_B)}{T (Y^* - L_B)}$$

The (negative) effect of trade value on liquidation incentives then perfectly offsets

Figure 1.4



the direct effect of high private value L_B for project assets. The relationship between L_B and optimal trade share σ is illustrated in Figure 1.5.

The results of the above discussion are summarised in the following proposition:

Proposition 1.3

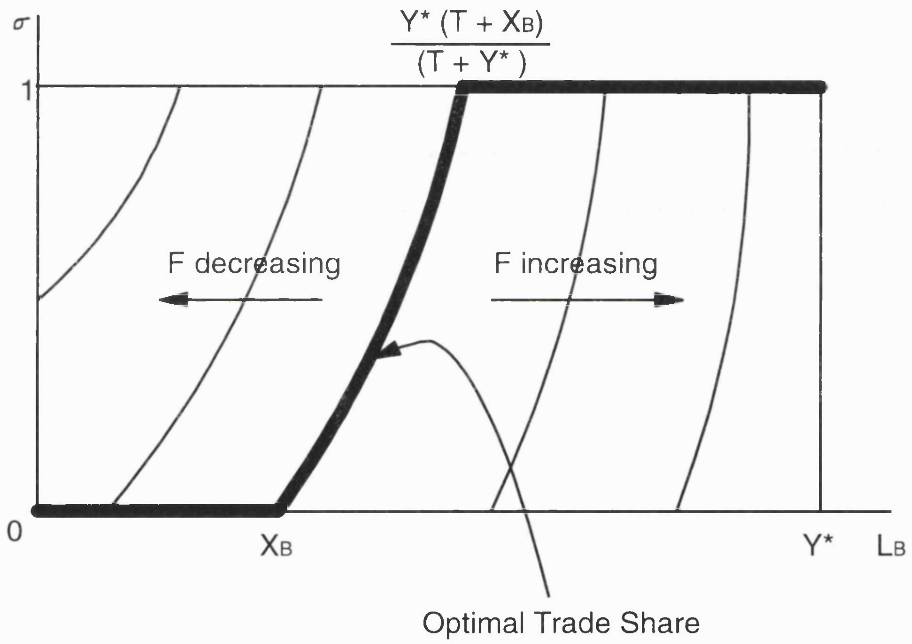
For creditors with given low value for liquidated project assets it is optimal to separate the trading partner and creditor roles. Conversely, for creditors with high values for liquidated project assets a single trading partner should provide project finance.

In other words when creditor value for project assets is high, the creditor should own the trading partner business.

In our analysis of trading partner borrowing we have assumed that the private value L_B of project assets is identical for all creditors, whether trading partner or not. However a trading partner's experience interacting with the entrepreneur may enable him to manage core project assets more effectively - generating a higher L_B value. Even with a higher L_B the impact of trade on the effective liquidation value is nevertheless ambiguous. The trade gain from continued entrepreneurial project management may completely offset the increased incentive to liquidate generated by a higher L_B, as points a and b in Figure 1.4 illustrate.

As an extension of this discussion suppose the trading creditor's L_B is in fact a continuous, increasing function of his project involvement. It seems reasonable to suppose that as the creditor's share of trading partner business (and hence the extent of his interaction with the entrepreneur) increases, so will his ability to generate value from the project. If trade interaction with the entrepreneur has a sufficiently positive effect on creditor L_B then increasing the creditor's share of project trade will actually *harden* liquidation incentives. The effect of increased interaction on L_B then more than offsets the increases in the adverse trade effect. In such circumstances the relationship between L_B and optimal trade share is radically altered.

Figure 1.5


σ
$\frac{Y^* (T + X_B)}{(T + Y^*)}$
1
F decreasing
F increasing
0
X_B
Y^*
L_B
Optimal Trade Share

Suppose that the relationship between trade share and L_B is given by:

$$L_B = G(\sigma , \lambda)$$

where L_B is an increasing function of creditor trade share σ. λ reflects the value a non-trading creditor can generate from core project assets (see Section 1.3). The above discussion can then be summarised in the following proposition:

Proposition 1.4

Define λ^* such that $G(1, \lambda^*) = X_B$.

If the rate of increase of creditor L_B with increases in market share (σ) is sufficiently great then it is optimal for creditor's with $\lambda \leq \lambda^*$ to take a maximum market share ($\sigma = 1$) and for creditor's with $\lambda Y^* > X_B$ to take a minimal market share ($\sigma = 0$).

In particular this holds if:

$$\frac{dG}{d\sigma} \geq \frac{T(Y^* - L_B)}{(\sigma T + Y^*)}$$

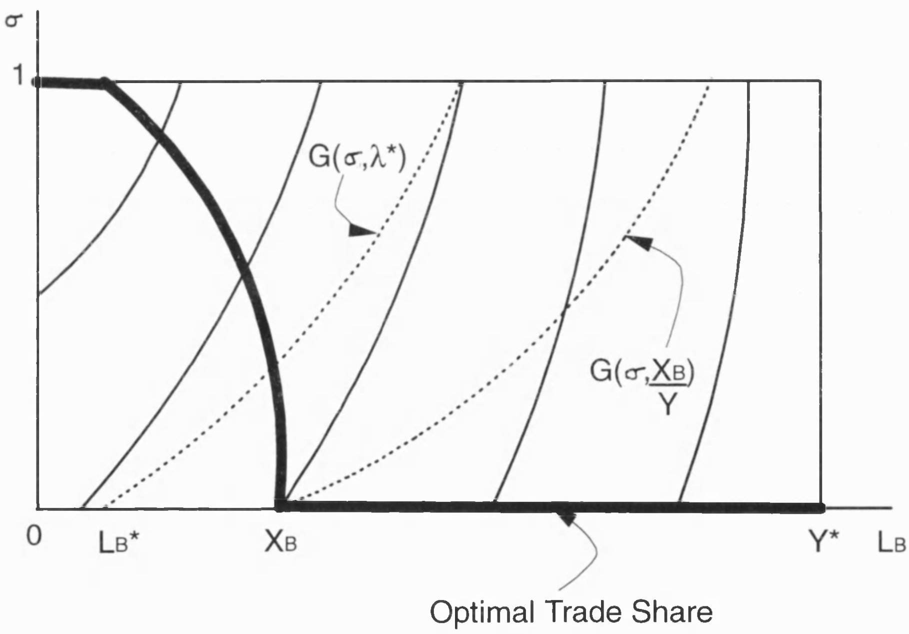
The condition in Proposition 1.4 ensures that at all points the slope of the function G exceeds that of the iso-L contour. This will hold, for instance, if:

$$\frac{dG}{d\sigma} \geq \frac{(Y^* - L_B)}{\sigma}$$

This is a natural formulation since it implies that the effect of additional market share on ability to manage project assets decreases as the creditor becomes proficient (L_B close to Y^*) and as trade share tends towards 1.

The optimal asset value/trade share relationship in this case is sketched in Figure 1.6. The potential effect of trade interaction in increasing creditor value for project assets is seen by contrasting Figures 1.5 and 1.6.

Figure 1.6



1.7 Conclusions

In this chapter we have considered the debt financing of an entrepreneurial project, where the innovator has a critical asset-specific investment contribution to make but no wealth. To maximise project funding, debt structure should minimise both voluntary, strategic default and the liquidation of project assets following unavoidable liquidity default. These goals are linked since minimising unavoidable liquidation of project assets encourages long-term asset-specific investment by the entrepreneur. By raising the returns to future project control, such investment in turn encourages greater debt repayment by the entrepreneur.

Creditor liquidation incentives are critical, and depend on two factors: the state of asset resale markets, and the creditor's private value for project assets - L. In goods states of nature, when the entrepreneur is cash rich, asset markets are likely to be healthy, enabling the creditor to extract substantial repayment from the entrepreneur. However, in bad states of the world, asset resale markets are likely to be weak, and the creditor must rely on his own value for project assets (L) to prevent strategic default. The characteristics of the creditor, i.e. *from whom* the entrepreneur borrows, are therefore important. Lending by low L creditors is limited by their ability to extract repayments in bad states of nature. Conversely, borrowing from high L creditors is limited by their excessive liquidation incentives. A non-monotonic relationship between liquidation value and maximum project finance is derived.

Borrowing from multiple creditors, with diverse characteristics (L values), is found to be optimal in general. Senior, high L debt is set low, but is never renegotiated - allowing maximum cash extraction in the bad state. Junior low L debt is set high, to extract maximum cash in good states when the entrepreneur is cash rich, but is readily renegotiated in bad states to manageable levels.

Finally,the role of borrowing from trading partners is considered. The effects of entrepreneur-project separation on trade profit then dilutes liquidation incentives. A trade share is therefore beneficial when creditor L is high, but lowers credit from low L financiers. However, countering this, interaction with project assets may also raise creditor value for project assets (via experience effects, etc.), increasing liquidation incentives.

Appendix 1

Lemma A1:

In considering the provision of a cash reserve R, we can restrict attention to

$R \in \{0, \max[0, L_B - X_B]\}$

Proof:

(i) Entrepreneurial cash constraints at date 0 $\Rightarrow R \geq 0$.

(ii) Suppose $0 < R < L_B - X_B$.

Maximum debt consistent with date 2 entrepreneurial control in the bad state:

$D = X_B + R$.

Cash retained by entrepreneur in good state $= (X_G + R) - (X_B + R) = X_G - X_B$.

Maximum finance, $F^* = p_G X_G + p_B X_B - p_G(X_G - X_B) = X_B$.

Suppose instead that $R = 0$ and $D = X_B$.

Then $F^* = X_B$.

(iii) Suppose $R = R^* > L_B - X_B$.

Bad State: Cash retained $= X_B + R^* - \min[D^*, L_B]$.

$= \max[X_B+R^*-D^*, X_B+R^*-L_B]$

Good State: Cash retained $= X_G + R^* - \min[D^*, L_G]$

$= \max[X_G+R^*-D^*, X_G+R^*-L_G]$.

Now suppose instead that $R = R^+ = L_B - X_B < R^*$, and $D^+ = D^* - (R^* - R^+)$

Bad State: Cash retained $= X_B + R^+ - \min[D^+, L_B]$

$= \max[X_B+R^*-D^*, X_B+R^+-L_B]$

Good State: Cash retained $= X_G + R^+ - \min[D^+, L_G]$

$= X_G + R^+ - \min[D^* - (R^*-R^+), L_G]$

$=\max[X_G+R^*-D^*, X_G+R^+-L_G]$

Clearly, cash retained with $R = R^* \geqslant$ cash retained with $R = R^+$.

Therefore $F^*(R = R^*) \leqslant F(R = R^+)$.

Proposition A1

Let $\delta = L_B - X_B$ and $\beta = (L_G - L_B)/(X_G - X_B)$.

(i) If for L_B such that $\delta = 0$ (i.e. $L_B = X_B$) $\beta > 0$, the relationship between maximum project finance F^* and the creditor's bad state liquidation value L_B is non-monotonic. In particular $dF^*/dL_B > 0$ for $L_B < X_B$, and $dF^*/dL_B < 0$ for at least some $L_B = \bar{L}$, where $X_B < \bar{L} < Y^*$.

(ii) F^* is non-decreasing in the good state liquidation value L_G i.e. $dF^*/dL_G \geq 0$.

Proof:

Let $F^{**} = p_B X_B + p_G X_G$

(i)Suppose $\delta < 0$ i.e. $L_B < X_B$.

Clearly $R = 0$.

Bad state: $\delta < 0 \Rightarrow \min[D, L_B] < X_B \Rightarrow$ repayment feasible irrespective of D.

Good state: repayment feasibility $\Rightarrow \min[D, L_G] \leq X_G + R = X_G$.

For maximum F^*, set $D = X_G$.

$$
\begin{aligned}
\text{Then } F^* &= F^{**} - p_B (X_B - L_B) - p_G \max[0, X_G - L_G] \\
&= p_B L_B + p_G \min[X_G, L_G]
\end{aligned}
$$

Clearly $dF^*/dL_B > 0$,$dF^*/dL_G \geq 0$.

(ii) Suppose now that $\delta \geq 0$ i.e. $L_B \geq X_B$, and $Y^* - C(Y^*) > F + R$

Bad state repayment = $\min[D, L_B] \leq X_B + R$

Good state repayment = $\min[D, L_G] \leq X_G + R$.

Case I: $R = 0$ and $D = X_B$

$F_I^* = F^{**} - p_G(X_G - X_B) = X_B$

Case II: $R = L_B - X_B$ and $D = X_G + (L_B - X_B)$

$F_{II}^* = F^{**} - p_G \max[0, X_G + (L_B - X_B) - L_G]$

Now $L_G - L_B = \beta\ (X_G - X_B)$

$\beta < 0 \Rightarrow F_I^* > F_{II}^* \ \Rightarrow dF^*/dL_B = dF^*/dL_G = 0.$
$\beta \geq 0 \Rightarrow F_I^* < F_{II}^*$

Further, $\beta \geq 1 \Rightarrow (L_G - L_B) \geq (X_G - X_B) \Rightarrow \max[0, X_G + (L_B - L_G) - X_B] = 0$
Then $F_{II}^* = X_B \Rightarrow dF^*/dL_B = dF^*/dL_G = 0.$

$0 < \beta < 1 \Rightarrow \max[0, X_G + (L_B - L_G) - X_B] = (X_G - X_B) + (L_B - L_G)$
Then $F_{II}^* = X_B + p_G(L_G - L_B) \Rightarrow dF^*/dL_B < 0$ and $dF^*/dL_G > 0.$

(iii)Now suppose $Y^* - C(Y^*) = F_{III}^* + R$
$\Rightarrow R = L_B - X_B \Rightarrow F_{III}^* = Y^* - C(Y)^* - (L_B - X_B) \Rightarrow dF^*/dL_B < 0, dF^*/dL_G = 0.$

From (i), (ii) and (iii):
$dF^*/dL_G \geq 0,$
$\delta < 0 \Rightarrow dF^*/dL_B > 0,$
$\beta > 0$ when $\delta = 0 \Rightarrow \exists\ \delta,\beta$ s.t. $\delta > 0,\ 1 > \beta > 0 \Rightarrow dF^*/dL_B < 0.$
QED.

Under the assumptions of Proposition 1: $Y^* = X_G = L_G$.
It follows that $\beta \geq 0$
Therefore $F_I^* < F_{II}^* = p_BX_B+p_GX_G+p_G(X_B-L_B) = X_B+p_G(Y^*-L_B)$

If $\delta < 0$ i.e. $L_B < X_B$, then $F^* = p_BL_B + p_GX_G$
If $\delta \geq 0$ then $F^* = \max[F_I^*, \min[F_{II}^*, F_{III}^*]]$
Now $F_I^* = X_B$, $F_{II}^* = X_B + p_G(Y^*-L_B)$, $F_{III}^* = Y^*-C(Y^*) - (L_B-X_B)$

$$\min[F_{II}^*, F_{III}^*] = \begin{cases} Y^* - C(Y^*) - (L_B - X_B) & \text{if } L_B \geq Y^* - \dfrac{C(Y^*)}{p_B} \\ X_B + p_G(Y^* - L_B) & \text{if } L_B \leq Y^* - \dfrac{C(Y^*)}{p_B} \end{cases}$$

$\max[F_I^*, \min[F_{II}^*, F_{III}^*]] = F_I^*$ if $F_I^* > F_{III}^*$, since $F_I^* < F_{II}^*$.

$F_I^* > F_{III}^*$, if $X_B > Y^* - C(Y^*) - (L_B - X_B)$ i.e. if $L_B > Y^* - C(Y^*)$

Therefore,

$$F^* = \begin{cases} p_B L_B + p_G X_G & L_B \leq X_B \\ p_B X_B + p_G X_G + p_G(X_B - L_B) & X_B < L_B \leq Y^* - \dfrac{C(Y^*)}{p_B} \\ Y^* - C(Y^*) - (L_B - X_B) & Y^* - \dfrac{C(Y^*)}{p_B} < L_B \leq Y^* - C(Y^*) \\ X_B & Y^* - C(Y^*) < L_B \end{cases}$$

References

Aghion, P. and P. Bolton (1992): "An Incomplete Contract Approach to Financial Contracting", Review of Economic Studies, 59:473-494.

Aghion, P. and J. Tirole (1994): "On the Management of Innovation", Quarterly Journal of Economics, 109:1185-1209.

Berglof, E. and E. Von Thadden (1994): "Short-Term versus Long-Term Interests: Capital Structure with Multiple Investors", Quarterly Journal of Economics, 109:1055-1084.

Bolton, P. and D. Scharfstein (1990): "A Theory of Predation based on Agency Problems in Financial Contracting", American Economic Review, 80:93-106.

Bolton, P. and D. Scharfstein (1993): "Optimal Debt Structure with Multiple Creditors", LSE Financial Markets Group Working Paper 161.

Dewatripont, M. and J. Tirole (1994): "A Theory of Debt and Equity: Diversity of Securities and Management-Shareholder Congruence", Quarterly Journal of Economics, 109:1027-1054.

Diamond, D (1991): "Monitoring and Reputation: The Choice Between Bank Loans and Directly Placed Debt", Journal of Political Economy, 99:689-721.

Gertner, R. , D. Scharfstein and J. Stein (1994): "Internal versus External Capital Markets", Quarterly Journal of Economics, 109:1211-1230.

Grossman, S. and O. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", Journal of Political Economy,94: 691-719.

Hart, O. (1995): "Firms, Contracts, and Financial Structure", Clarendon Press, Oxford.

Hart, O. and J. Moore (1989): "Default and Renegotiation: A Dynamic Model of Debt", LSE Working Paper.

Hart, O. and J. Moore (1990): "A Theory of Corporate Financial Structure Based on the Seniority of Claims", LSE Working Paper.

Rajan, R. (1992): "Insiders and Outsiders: The Choice Between Relationship and Arm's-Length Debt", Journal of Finance, 47:1367-1400.

Shleifer, A. and R. Vishny (1992): "Liquidation Values and Debt Capacity: A Market Equilibrium Approach", Journal of Finance, 47:1343-1366.

von Thadden, E.-L. (1995): "Long-Term Contracts, Short-Term Investment and Monitoring", Review of Economic Studies, 65:557-575.

Chapter 2

Integration and Investment Specificity

2.1 Introduction

A key characteristic of any investment is its specificity. Specific investments will be more effective in their optimal application, but less widely applicable than general investment. Where hold-up problems exist an inefficient specificity choice may be made. If an agent makes a specific investment, value is high but outside options are weak - the agent is effectively locked in. A general investment reduces value but strengthens outside options and hence may be chosen to escape lock-in.

Clearly if it is possible to contract upon the nature and specificity of investment ex ante then these hold-up costs can be overcome. The agent can be guaranteed sufficient reward for making the optimal, contractually specified, investment choice. However in a complex world it is not unlikely that investments cannot be adequately described ex ante and hence contingent contracts cannot be written (or at least writing such contracts may be prohibitively expensive). Verification of any contract by a third (arbitrating) party may be difficult, even if a contract can be formulated. In such a context residual rights of control - who controls what in areas where any contract is silent - become important. Following Grossman and Hart (1986) we can think of ownership as allocating these residual control rights. If an agent owns an asset it seems reasonable to suppose that they make the decisions concerning that asset, unless control has been ceded to another party through the provisions of a contract.

Ownership, in allocating control rights over an asset, may alleviate the threat of hold-up. If a worker owns the machines he or she works with then the dangers of lock-in following asset-specific human capital investment are eliminated. Conferring ownership rights will encourage efficient (i.e. specific) investment by that worker. However giving ownership rights to one agent in effect denies them to others. Non-owning workers will develop less tailored, more generally applicable skills. Consequently, as one asset-worker relationship is strengthened others are weakened. As stressed by Grossman and Hart (1986), the costs and benefits of a particular allocation of ownership rights go hand in hand.

Of course, the determination of what is a generally applicable skill may depend on the inter-related decisions of several agents. If a whole industry of independent firms utilises common asset technology then asset-specific investment will still be *generally applicable*, since there is no danger of lock-in to any one firm. However if the technology is unique to one firm, then the very same investment is idiosyncratic. In particular, where a single owner controls the industry the investment will be owner-specific. Clearly, the industry-wide allocation of ownership rights and the resulting technology choices of firms will impact on the costs of any particular ownership decision. In our model we attempt to capture this endogeneity - emphasising the links between integration decisions, market structure and the specificity of investment.

We believe that a location model is an appropriate framework within which to study these issues. An agent or firm's location represents its choice of investment. Where assets and workers are located apart transport costs are incurred with trade, reflecting the costs of inappropriately tailored investment.

The general thrust of our results is as follows. When the costs of a poor match between a worker's human capital and a physical asset are great, that worker should own the asset. Similarly if weak ties between physical assets are very costly those assets should be jointly owned and controlled. If they are not then an inappropriate choice of asset specificity may lead to excessive mis-match costs. If, on the other hand, highly tailored worker skills are more valuable than strong asset ties then upstream workers should control upstream assets and downstream workers should control downstream assets. Where it is essential to coordinate the skills of upstream workers with both upstream and downstream assets then those workers should be given control of all assets (both upstream and downstream) i.e. full forward integration results. Conversely when poorly matched upstream skills are relatively unimportant the downstream worker should control both upstream and downstream assets.

The costs of mis-match in the model are generated by attempts to alleviate lock-in to a given firm through the choice of more generally applicable though less effective investment. The magnitude of these costs will be influenced by the value of alternative opportunities i.e. the strength of the market. If several firms choose to locate close by then a particular investment has wider applicability. These opportunities in turn are sensitive to the general extent of integration in the industry. In particular, non-integrated downstream firms will adopt versatile general technologies, thus generating stronger demand for the services of downstream workers with specific skills. As a result more specialised human capital investment is encouraged.

An independent downstream sector may therefore ameliorate the costs of individual forward integration decisions, whereby downstream workers are denied ownership (control) rights. In such a context the coordination benefits of *some* integration may be sustained in the industry even when the costs of poorly tailored downstream worker skills are relatively high, encouraging a fragmented ownership structure. Asymmetries in the model set-up exclude a similar result for backward integration.

In an extension to our basic model we investigate the manner in which this last property influences the extent of horizontal integration. If a firm limits its integrating activity a greater number of independent firms are operational in the industry. These compete for the services of non-owning workers, expanding demand and encouraging more specific worker investment. A firm may therefore forgo superficially attractive integration possibilities, in order to elicit improved worker performance. We explore how this last feature depends on the contracts that can be written, and the possibility of ownership renegotiation. In particular, if ownership rights can be traded after investments have been made, initially independent firms may not remain so when actual transactions occur. Ex post integration is attractive to the extent that it allows the (partial) monopolization of the market for workers' services. Of course this perfectly foreseen possibility will influence the human capital investment decisions made by those workers. The ability to commit to the original ownership structure may then be important.

It was Ronald Coase, in his classic 1937 Economica paper, that first emphasised a transaction cost minimising approach to the "markets versus hierarchies" question. He argued that transactions would be organised within the firm when the costs of doing so were less than those of market transacting. In particular, he stressed the value of authority in a complex, evolving world where comprehensive contracts are an impossibility and constant re-contracting is costly.

This theme was taken up in the work of Oliver Williamson (see e.g. Williamson (1985)). He has stressed the relevance of private authority and ownership when contractual incompleteness, opportunistic behaviour and asset specificity concerns coincide.

Asset specificity is a key component in Williamson's writing. He identifies four types of specificity - human asset and physical asset specificity, site specificity, and dedicated assets - as important factors in determining the structure of transactions. Our model incorporates the first three types in a common framework. Where assets are specific to a given transaction there are dangers of ex post hold-up because of lock-in. We are then more likely to see integrated structures and internal transactions. Where asset specificity is low market transactions are more likely. Williamson fails however to explain how transfer of ownership affects the opportunistic behaviour of the new owner-turned-manager.

These issues have been addressed in a bilateral context in the seminal paper by Grossman and Hart (1986). Control of assets reduces the dangers from lock-in and hence improves ex ante investment incentives. Transferring ownership (and therefore control) of assets from agent A to agent B thus improves B's incentives but weakens A's incentives to invest. A given transfer of ownership rights (e.g. an integration decision) therefore generates both costs and benefits. Broadly their results are similar to some of ours. If control affects investment incentives, and upstream investment is very important then upstream workers should be given control. If, on the other hand, downstream investment is important then downstream workers should be given control.

It is useful to consider this analysis in a multi-lateral setting. In such a context the specificity of investments can be considered. A multi-lateral framework introduces a role for strategic interactions and allows outside options possibilities to be investigated. In addition the externality effects of any integration decisions, and the possible divergence between private and 'social' objectives can be considered. Such a framework has been formulated in a recent paper by Bolton and Whinston (1990), but in rather a different context to ours. They investigate issues of supply and foreclosure in a setting with restricted supply of input. The broad structure of their model is nevertheless similar to ours and the emphasis on outside options, and the investment inefficiencies that may be generated in their pursuit, is a common one.

A rather different approach to a multi-agent, multi-asset environment with incomplete contracts is taken by Hart and Moore (1990). They adopt a cooperative approach (based on the Shapley value) to the optimal allocation of asset control rights. Again the costs and benefits of ownership are determined by the effects on ex-ante investment in human capital. A number of our results echo their conclusions. Where there are strong ties between an agent's investment and a particular asset then that agent should own the asset (other things being equal), and where there are strong ties between physical assets they should be owned together (though the benefits of strong physical ties in their model occur indirectly via the effects on human capital investment).

In all of the above papers the transfer of ownership does not affect the control of ex ante investment decisions. The powers of ownership come in to play once investment has occurred and trade takes place. In contrast, our paper emphasises that owners of assets choose the physical investments that are made in those assets. Thus ex ante investment decisions will be affected directly by the allocation of ownership rights.

Furthermore, none of the above papers considers the choice of asset specificity in any detail. We believe that this is an important factor influencing ownership structures and deserves particular consideration. In this paper we develop an explicit model of specificity decisions along the lines of Bickenbach and Williams (1991).

An early explicit treatment of asset specificity is provided by Riordan and Williamson (1985). However their analysis focuses on the general differences between market and internal organizational forms. They do not consider the endogenous determination of specificity and its costs through the explicit allocation of ownership rights. Furthermore Riordan and Williamson examine a bilateral setting hence there is no room for externalities in their model.

There are also similarities between our paper and Holmstrom and Tirole's analysis of transfer pricing and organizational form (Holmstrom and Tirole (1991)). As in our work they highlight the relationship between ownership structure, investment coordination and managerial search for outside options in an incomplete contract environment. They also model the choice of investment specificity explicitly. Choosing a general market orientation improves outside options but may reduce the value of trade. However the analysis in their paper concentrates on the design of contractual incentives for non-owning managers and emphasises the internal organisation of the firm. Again, changes in ownership structure affect investment only indirectly, by altering the ex post terms of trade. In our model, by contrast, changes are also induced by the re-allocation of control over investment decisions. Furthermore, though the outside options available to managers are influenced by their own investment orientation decisions (and the rules on market trading imposed by the owners of the firm) in Holmstrom and Tirole's model, they are not affected by the actions of other managers and owners in the industry. A central theme of the current paper however is the relationship between overall industry organisation and the outside options available to individual managers. We investigate the impact of industry ownership structure on the costs of 'market orientation' and the resulting externality effects.

Finally, two closely related papers by Farrell and Gallini and Shepard emphasise the role of second sourcing in alleviating lock-in where investment is idiosyncratic and contracts are incomplete. A firm may license its proprietary technology to a second producer, ensuring some competition for consumers and hence encouraging technology-specific investment. As in our model a firm may thus prefer to limit its monopoly power through the choice of industry structure. The ability to credibly commit to such a policy is crucial. These papers do not however consider these effects in the context of a firms' integration decisions.

2.2 The Model

The model involves two independent markets for final goods, each consisting of a single consumer with unit demand. Each market is served by one of two downstream firms (D_1, D_2) i.e. D_i is a monopolist in market i. [11]

A downstream firm transforms input supplied by one of two upstream firms (U_1, U_2) into final output using a 1-1 technology. There are no input capacity constraints, hence there is no competition between downstream firms for supplies.

For simplicity we fix the locations of the upstream firms at the ends of a unit line (U_1 at 0, U_2 at 1). The downstream firms are free to locate (simultaneously) at any point along the unit line (D_1 chooses location c_1 and D_2 chooses location c_2). We can think of location as representing technology choice, for instance. A location close to one of the upstream suppliers indicates a highly specific investment while a location equi-distant from both suppliers corresponds to a more general technology.

The value of the basic product to the final consumer is v. However, this core value can be increased by employing the services of skilled workers. Associated with the downstream firms are two workers (w_1, w_2) who can each increase the value of a product by an additional v. Both workers can work for the same firm, each increasing the value of the product by v. The value of a worker's contribution does not depend on the identity of the employer or on the total number of workers employed. In the model we represent human capital investment as a location on the unit line.

Once downstream firms have chosen their locations each worker can choose to locate at any point on the unit line (w_1 locates at b_1 and w_2 at b_2). All workers locate simultaneously at time T=2. We can relax this assumption and allow workers to choose their location at the same time as that of any assets they own

[11] Downstream competition for consumer trade is therefore not a factor influencing integration decisions.

(see Footnote 17).

If a downstream worker locates a distance β from a downstream firm then a transport cost $r\beta$ is incurred by the firm in using the worker's skills. This can be thought of as the cost of transforming the skill, or alternatively as the reduced value of using an imperfectly tailored skill. Similarly, when a downstream firm locates a distance γ from an upstream firm then there is a cost $t\gamma$ to transforming the input. Again this cost reduces the basic value of the product. We assume that $v > r, t$ - ensuring that downstream firms can always source input from either supplier and that both downstream firms will always seek the services of a worker. The availability of outside options is thus guaranteed, whatever the location of the downstream asset or worker.[12] The relative values of the parameters r and t will play a critical role in our analysis.

We assume Bertrand competition between upstream firms to supply downstream firms, and between downstream firms for the services of workers. These trades involve spot contracts. Therefore, we might expect an upstream worker to work on an input which is then sold to a downstream firm, where a downstream worker performs additional value-enhancing tasks.

Prices will be influenced by location choice and the magnitude of transport costs. For instance, if a downstream worker locates distance 1/6 from D_1 and 1/2 from D_2 then D_1 is prepared to pay up to $v - r/6$ for the worker's services (i.e. value added net of transport costs) while D_2 will offer up to $v - r/2$. Bertrand competition ensures that the worker will in fact receive $v - r/2$, and work for D_1. In our example, the worker could (and would) increase the wage by locating distance 1/3 from both D_1 and D_2. Where a worker has several equally attractive employment offers, we assume worker i works for firm i. Similarly, if indifferent, downstream firm i will buy input from upstream firm i.

[12] The results are unaltered if we assume $v>r/2$, $v>t/2$ though calculation is messier. It is sufficient that when a firm or worker locates at the centre of the unit line trade can occur with either of two firms located at opposite ends of the line.

In this simplified model each upstream asset has an owner, m_i.[13] Downstream asset i can either be owned by downstream worker w_i (the non-integrated case) or by m_i (forward integration). Horizontal integration of either upstream or downstream assets is not permitted. This last restriction is needed to ensure that any non-owning worker or non-integrated downstream firm has a choice of trading partner (and thus an "outside option" in any negotiation). If both upstream firms were jointly controlled, for instance, then an independently owned downstream firm would face a monopolised input market. In total there are 4 possible overall ownership structures: Forward Integration, Non Integration and Partial Integration (- 2 symmetric structures). The owner of an asset decides on its location.

We assume that contracts cannot be written or verified conditional on the choice of location (investment) of either firms or workers. In the context of our location model this may seem extreme. However, we have in mind a situation where describing complex investments precisely is impossible. Profit sharing schemes are also deemed infeasible. The only contracts permitted are those conferring ownership (and hence residual control) rights at T=0. All information is identically available to owners and workers, though not to a court.

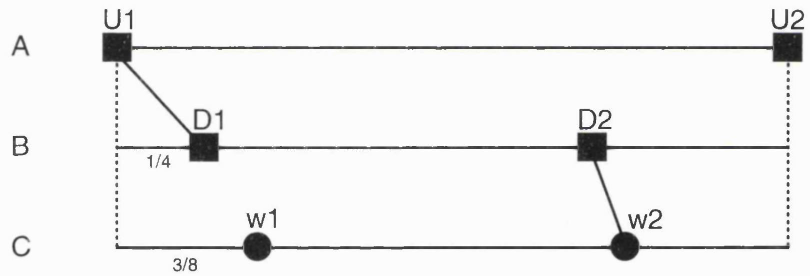
Ownership of D_i is allocated at T=0 to maximise the joint profits of m_i and w_i. We refer to the resulting industry structure as the Equilibrium Ownership Structure. Throughout we will contrast this ownership structure with that which maximises the combined profits of *all* upstream and downstream agents. We refer to this configuration as the Industry Optimal Ownership Structure. More on this below.

Figure 2.1 illustrates the form the model may take. Ownership ties are represented by the bold connecting lines. Thus on the right of the diagram downstream worker 2 (w_2) is shown to own downstream asset 2, while on the left of the diagram upstream owner 1(m_1) owns both upstream and downstream assets. Downstream worker 1 (w_1) is a non-owning worker. Overall the industry is partially integrated.

---

[13] For now this upstream owner takes no actions, but is assumed to be essential. In Section 6 we will consider m_i's role explicitly.

Figure
2.1



The unit line has been blown up into 3 parallel lines for ease of exposition. The fixed locations of the upstream firms are shown on line A, the locations of downstream firms on line B and the location choices of downstream workers on line C. For instance if D_1 locates a 1/4 unit from 0 and w_1 locates 3/8 of a unit from 0 then the transport costs incurred when w_1 works with D_1 are r/8, while the cost of D_1 buying input from U_1 are t/4.

2.3 The First Best Outcome

As a benchmark it will be useful to consider the first best outcome. The source of inefficiency in the model is location mismatch between downstream workers and firms, and between upstream and downstream firms. A social planner could thus achieve first best by the appropriate choice of downstream worker and asset locations. Upstream assets are of course fixed at the ends of the unit line.

Each downstream asset should be located next to an upstream asset, guaranteeing an input supply free of transport costs. Supplier identity is irrelevant. Each worker, in turn, should be located next to a downstream asset.

Result 2.1 [First Best]

To attain the first best outcome:

(a) Each downstream firm should locate next to an upstream firm i.e.

$$c_i = 0 \text{ or } 1 \qquad i = 1,2$$

(b) Each downstream worker should locate next to a downstream firm i.e.

$$b_i = c_j \qquad i = 1,2\ ;\ j = 1,2$$

Clearly there are several industry configurations that achieve first best. It is always inefficient for firms or workers to locate in the interior of the unit line.

It may be worthwhile noting, at this preliminary stage, that if upstream firms were allowed to choose location freely then the first best would exist as *an* equilibrium of the model. In this equilibrium all firms and workers share a common location. Clearly, transport issues are no longer relevant. Furthermore, no worker or asset-holder would have a strict incentive to deviate - by locating apart from all other assets and workers. When specificity (location) considerations cease to be relevant, competition guarantees efficiency (but more on this in the Conclusions).

2.4 Preliminaries

In this section we provide some basic results concerning the prices paid for inputs and the services of workers, and hence the profits accruing to the different parties. Prices are driven by location choice. Agents therefore *locate* their skills, and any assets they own, to maximise profits. The location choice results are at the heart of the paper, since they determine the costs and benefits of any ownership allocation.

There is Bertrand competition for the supply of inputs and the services of workers. Downstream firms buy input from the upstream firm that offers the lowest price, while workers work for the firm that offers the highest wage. We assume that internal transfer prices are set at competitive levels. [14]

Lemma 2.1 [Prices]

(a) If a downstream firm locates at c then it receives input at a price

$$p = t \max[c, 1-c]$$

If $c > 1/2$ the firm is supplied by U_2 while if $c < 1/2$ it is supplied by U_1.

(b) Suppose D_1 and D_2 locate at c_1 and c_2 respectively. If a downstream worker locates at b then he/she receives a price

$$p = v - r \max[|b-c_1|, |b-c_2|]$$

If $|b-c_1|>|b-c_2|$ the worker works for D_2 while if $|b-c_1|<|b-c_2|$ he/she works for D_1.

Proof: See Appendix 2.2.

When a worker locates closer to one employer and further from the alternative, outside options are weakened. As a result that employer can offer a lower winning price to secure the worker's services. In Klein et al's classic terminology

---

[14] Transfers between an agent as buyer and the same agent as supplier, or between the agent as employer and as employee do not affect decisions. However, this approach allows us to adopt a common notation for prices and profits, whatever the ownership structure.

"the appropriable specialised portion of the quasi-rent is that portion in excess of its value to the second highest-valuing user" (Klein, Crawford and Alchian (1978) p. 298). Specialised skills are less applicable, and hence command lower prices, in alternative use. Similarly, if a downstream firm locates its assets close to one of the upstream suppliers, its alternative supply source is distant and relatively inappropriate. As a result, the closer supplier can demand a higher price for its product. It is a feature of the model that the trading price is always determined by the second best offer.

From these price results we can establish the profits earned by each party. Of course, a worker may well earn a profit not only as a worker but also as an asset owner too. Since internal transfer prices are set at the competitive level, the allocation of ownership rights affects profits solely through its influence on location decisions. The owners of downstream assets (D_1, D_2) choose their locations at T=1, followed at T=2 by worker location choice. A firm or worker will choose a location that maximises its profits.

Lemma 2.2 [Location Choice]

(a) An owning downstream worker locates next to his/her downstream asset [i.e. $b_i = c_j$ if w_i owns D_j], while a non-owning downstream worker locates mid-way between downstream assets.

(b) Integrated upstream and downstream assets are located together. Non-integrated downstream firms will locate mid-way between upstream firms. Thus integrated D_1 locates at 0 and integrated D_2 at 1. Non-integrated D_1 and D_2 locate at 1/2.

Proof: See Appendix 2.2.

When upstream and downstream assets are owned together lock-in worries discouraging adjacent asset location disappear. Indeed, such locations will then be chosen, to minimise transport costs and hence maximise profits. An owner will select the technology that ensures the best possible match between input attributes and production needs. Similarly, if a worker owns the asset he/she works with, then the threat of ex post hold-up disappears. Owning workers will locate next to their assets to minimise transport costs. Such workers will acquire value maximising firm-specific skills. Ownership affects profits through its impact on location choice, and hence on transport costs.

For example, if upstream owner 1 (m_1) owns D_1 and U_1 both physical assets will be located next to each other (at location 0). However, the non-owning downstream worker (w_1) locates mid-way between D_1 and D_2. On the other hand if U_1 and D_1 are separately owned then w_1 locates next to D_1 (to minimise costs) but w_1-D_1 combined locate at 1/2 - between upstream firms - to minimise lock-in to either input supplier.

2.5 Ownership choice

There are three possible industry ownership configurations in our basic model. In this section we will first consider the outcome of equilibrium location choice, for each structure in turn. We will then analyse the allocation of ownership rights undertaken at T=0. Finally, we will compare this with the Industry Optimal ownership configuration.

Full Integration

Each pair of upstream and downstream assets is jointly owned and controlled. Integrated assets are located together (at the ends of the unit line) to minimise the transport costs of input supply. Non-owning downstream workers locate at 1/2, equi-distant from both integrated firms - as illustrated in Figure 2.2. This (general human capital investment) minimises worker lock-in to either potential employer. However, as a result, a transport cost r/2 is incurred in adapting workers' general skills for a particular use. Bertrand competition for workers' services ensures it is they who bear this cost. The profits of the four parties are as follows:

$$\Pi_{m1} = v \qquad \Pi_{w1} = v - r/2$$
$$\Pi_{m2} = v \qquad \Pi_{w2} = v - r/2$$

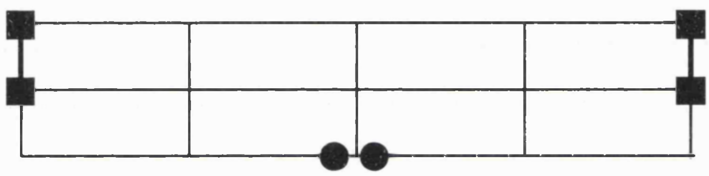
Non-Integration

Ownership of assets is now separated. Downstream assets are owned by downstream workers (w_1, w_2) and upstream assets are owned by upstream

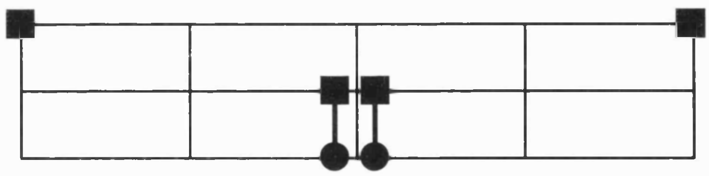
Figure 2.2

Equilibrium Location Patterns

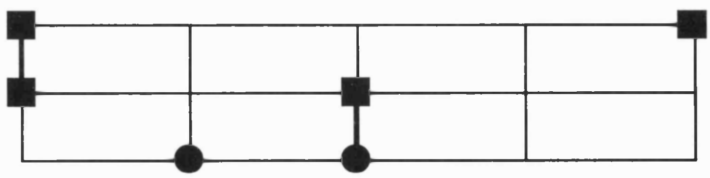
Full Integration



Non-Integration



Partial Integration



owners (m_1, m_2). Each downstream worker will locate next to his/her downstream asset, avoiding transport costs. Ownership eliminates lock-in worries and adjacent location maximises the value of the worker's human capital. However, to minimise lock-in to either upstream input supplier, each downstream worker-asset combination locates in the centre of the line, at 1/2.[15] The "appropriable element of the asset quasi-rent" is then minimised too. A transport cost t/2 is incurred transforming either supplier's input to the specifications of each downstream firm (i.e. a distance of 1/2, at unit cost t). The profits of the four parties are now given by:

$$\Pi_{m1} = 0 \qquad \Pi_{w1} = 2v - t/2$$
$$\Pi_{m2} = 0 \qquad \Pi_{w2} = 2v - t/2$$

Partial Integration (U_1-D_1 integrated)

This is the most interesting ownership structure. One pair of upstream and downstream assets are integrated, while the other pair are owned separately. Assume U_1 and D_1 are integrated. Equilibrium location patterns are again illustrated in Figure 2.2. Integrated D_1 locates at 0, next to its upstream division U_1. Non-integrated downstream firm D_2 locates at 1/2 (together with its worker-owner). The potential employers of non-owning worker w_1 are therefore located at 0 and 1/2. Consequently, this worker will choose to locate at 1/4, mid-way between D_1 and D_2. A cost of only r/4 is then incurred in transforming that worker's human capital. Profits for the four parties in this case are:

$$\Pi_{m1} = v \qquad \Pi_{w1} = v - r/4$$
$$\Pi_{m2} = 0 \qquad \Pi_{w2} = 2v - t/2$$

Having obtained equilibrium location choices and profits for each ownership configuration, we can now consider which configuration will prevail in equilibrium. The allocation of ownership rights over asset D_i is chosen to maximise the joint

[15] Each downstream firm could be separately owned by an independent outside party in the non-integrated case. To avoid lock-in to either input supplier both downstream firms locate at 1/2. This in turn creates a perfectly competitive market for non-owning downstream workers' skills at 1/2. With competition, the allocation of ownership rights is unimportant. Workers are happy to locate next to firms, eliminating worker-asset mismatch. This outcome exactly replicates that generated by downstream worker control. The key here is the separation of upstream and downstream asset ownership.

profits of m_i and w_i. These joint profits, for each industry structure, are given below:

Integration	$\Pi_{m1\text{-}w1} = 2v - r/2$	$\Pi_{m2\text{-}w2} = 2v - r/2$
Non-Integration	$\Pi_{m1\text{-}w1} = 2v - t/2$	$\Pi_{m2\text{-}w2} = 2v - t/2$
Partial Integration	$\Pi_{m1\text{-}w1} = 2v - r/4$	$\Pi_{m2\text{-}w2} = 2v - t/2$

Joint profit maximisation clearly involves minimising the transport costs incurred by the relevant upstream-downstream pair.

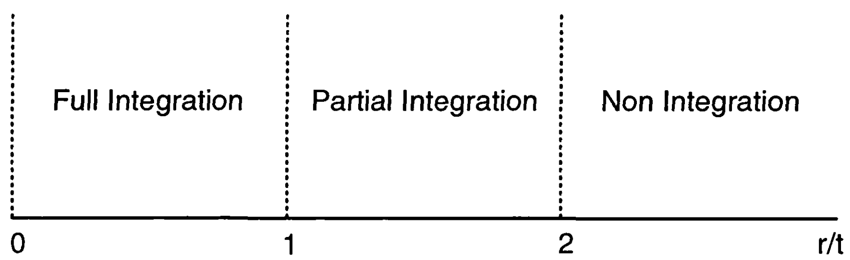
A general trade-off is apparent. Integration ensures strong ties (short distances) between assets, but weak ties (larger distances) between assets and workers. Conversely, non-integrated structures encourage strong ties between workers and assets, but correspondingly weaker ties between assets. When the (transport) costs of using badly tailored input are greater than the costs of non-specific worker skills then the physical assets should be owned together. On the other hand, if the value of specific worker skills is relatively high then downstream assets should be worker owned. Costs are driven by two factors: unit transport cost values, r and t, and location patterns.

Turning to the first of these effects, when unit worker transport costs (r) are high relative to unit input transport costs (t), non-integrated structures are attractive. On the other hand, low values of r/t encourage integrated outcomes. Figure 2.3 shows the relationship between equilibrium ownership structure and r/t. For $r/t < 1$ an integrated industry structure emerges in equilibrium, while for $r/t > 2$ equilibrium industry structure is non-integrated.

Total transport costs are also affected by location decisions i.e. by the distances between assets and workers. Critically, these location decisions may depend on overall industry ownership structure. To see this, compare m_1 and w_1's profits in a fully integrated industry with those under partial integration (where U_2 and D_2 are assumed non-integrated). The cost of forward integration to the m_1-w_1 pair is the transport cost incurred when the non-owning w_1 makes its individual profit maximising location decision. This downstream worker will always locate mid-way between D_1 and D_2. The transport cost incurred therefore falls from r/2 to r/4 as the industry as a whole becomes more fragmented, and the distance separating

Figure 2.3

Equlibrium Ownership Structures



downstream firms contracts from 1 to 1/2.

The more idiosyncratic are firms' skill requirements the more reluctant non-owning workers will be to invest in firm-specific skills. Integration, in encouraging idiosyncratic options, raises the costs of using such workers' skills. Non-integrated downstream firms will adopt less specialised technology (to avoid lock-in to particular suppliers), encouraging more firm-specific worker investment.

It is important to note that, since the downstream markets are independent, there is no strategic incentive here for vertical integration in order to discourage specific investment i.e. we do not see strategic foreclosure effects.

The results are formally collected below.

Proposition 2.2 [Equilibrium Ownership Structure]
The equilibrium allocation of industry ownership rights depends critically on the magnitude of unit worker transport costs (r) relative to input transport costs (t). In particular:
(i) low values of r/t (< 1) generate a fully integrated structure in equilibrium,
(ii) high values of r/t (> 2) imply a non-integrated structure, and
(ii) for intermediate values (1 < r/t < 2) the industry is partially integrated i.e. integrated and non-integrated firms coexist.

The relationship between equilibrium ownership structure and r/t is illustrated in Figure 2.3.

Equilibrium ownership structures that do not maximise overall industry profits are inefficient. In assessing this we must consider the combined profits of m_1, m_2, w_1 and w_2, for each ownership configuration:

Π(Full Integration)) = 4v - r

Π(Non-Integration) = 4v - t

Π(Partial Integration) = 4v - t/2 - r/4

For a range of r/t values the industry optimal (efficient) ownership structure is not observed.

Proposition 2.3 [Industry Optimal Ownership Structure]
There is excessive integration in the industry. For $2/3 < r/t < 1$ a fully integrated industry is observed in equilibrium though partial integration is efficient.

Industry optimal and equilibrium ownership structures can be compared in Figures 2.3 and 2.4.

In assessing the benefits of merger, non-integrated upstream and downstream firms trade-off the benefits of closer asset ties (and a cost saving of t/2) against the added cost, r/2, of more general investment by the downstream worker. However, a merger also has consequences for the industry as a whole. Integration shifts downstream asset location from 1/2 to the end of the unit line. This increases the distance between downstream firms, weakening the options facing a worker with specific skills. Consequently, less firm-specific investment is undertaken by these workers. The human capital mismatch costs for an existing integrated firm in the industry are therefore increased, from r/4 to r/2. Integration of assets U_j and D_j thus imposes a negative externality on integrated firm i.

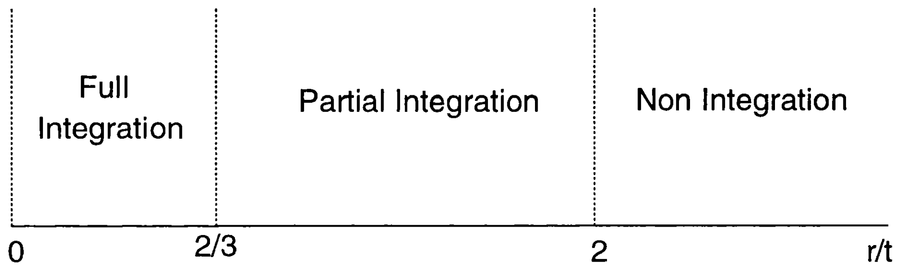
Of course, in such circumstances the integrated firm could offer the firms considering merger a fee conditional on their not integrating. If ownership structures were observable to a court, such a contract would be enforceable. In this case the Industry Optimal Ownership Structure could be sustained - as predicted by the Coase Theorem. However where ownership rights can be verified in a court by the owner, but are not openly observable, such contracts may not be effective.[16]

We will now consider a more general version of the model, that allows fully endogenously determined ownership.

[16] The integrating firms could then accept payment from the integrated party in exchange for not integrating, and then proceed to integrate secretly. Of course, a secret re-allocation of control rights may still encounter problems. In order to assert ownership rights in the event of a dispute, ownership must be proved to a court. If this violated previous agreements, then damages might be claimed. Any secret asset transfer agreements, therefore, might also involve agreements over responsibilities for damage payments.

Figure 2.4

Industry Optimal Ownership Structures



2.6 The General Case

In the basic model upstream assets are always owned by inactive entrepreneurs. We did not permit downstream workers to control both upstream and downstream assets. Yet this would appear to remove all the inefficiencies of non-specific investment. Since upstream entrepreneurs have no investment (location) decisions, removing their control rights would not have adverse efficiency implications. The benefits of integration could be achieved, at no apparent cost, by giving downstream workers control of upstream and downstream assets.

However, more realistically, upstream owners too have skills that contribute to overall value. Transfers of ownership that improve downstream workers' incentives to make specific investments may therefore cause a deterioration in upstream workers' incentives. As Grossman and Hart (1986) stresses, integration will then generate costs as well as benefits. To examine this issue in more detail we must extend our simple model somewhat.

We will introduce a pair of upstream workers (m_1 and m_2) that can enhance the value of input by an additional amount v. These workers must choose locations (a_1 and a_2 respectively) on the unit line. If an upstream worker locates distance α from a firm then the firm incurs costs $s\alpha$ using the worker's skills ($v > s$). Upstream firms compete a la Bertrand for workers. Once again it should be emphasised that both upstream workers can work for the same firm (but when indifferent m_i will work for U_i). Upstream workers choose their location at T=2 - *after* the location of physical assets. [17]

All assets in the model are now owned by workers. We restrict ownership of U_i and D_i to either upstream worker i or downstream worker i. Both U_i and D_i can be owned by m_i (forward integration, or upstream control), both can be owned by w_i (backward integration), or m_i can own U_i and w_i can own D_i (non-integration).

---

[17] We can relax this assumption and allow workers to choose location simultaneously with any assets they own, provided worker location choice is unobservable until the end of T=2. If sequential worker location decisions were observable then the possibility of ownership trade between location choices would arise. It is important that non-owning workers locate after downstream assets.

We do not allow upstream control of downstream assets combined with downstream control of upstream assets,[18] nor do we allow any horizontal integration for the reasons outlined in the simple case above.

The key mechanism driving location choice is precisely that highlighted in the basic model. Therefore, we will not describe the equilibrium location patterns in detail here. However, it is worth repeating the fundamental location principles:
(a) Non-owning workers locate mid-way between potential employer firms, while worker-owners locate next to their assets.
(b) Vertically integrated assets are located next to each other, while independent downstream firms locate mid way between upstream assets.

These results, from Section 2.4, are proved for the general case in Appendix 2.2.

We now have two types of worker investment to consider. Our optimal ownership structure will therefore depend not only on the cost of transforming downstream worker skills relative to input transformation costs (r/t), but also on the cost of transforming upstream worker skills relative to the cost of transforming input (s/t).

Proposition 2.4 [Equilibrium Ownership Structures]
The equilibrium ownership structure is determined by the relative unit cost of transforming upstream human skills (s/t) and the relative cost of transforming downstream human skills (r/t):
(i) For low values of r/t and/or s/t full integration is observed in equilibrium,
(ii) With high values of r/t and s/t both sets of assets are non-integrated.
(iii) Partial integration is the equilibrium industry ownership structure for a range of intermediate parameter values ($s/t > 1$, $1 < r/t < 2$).

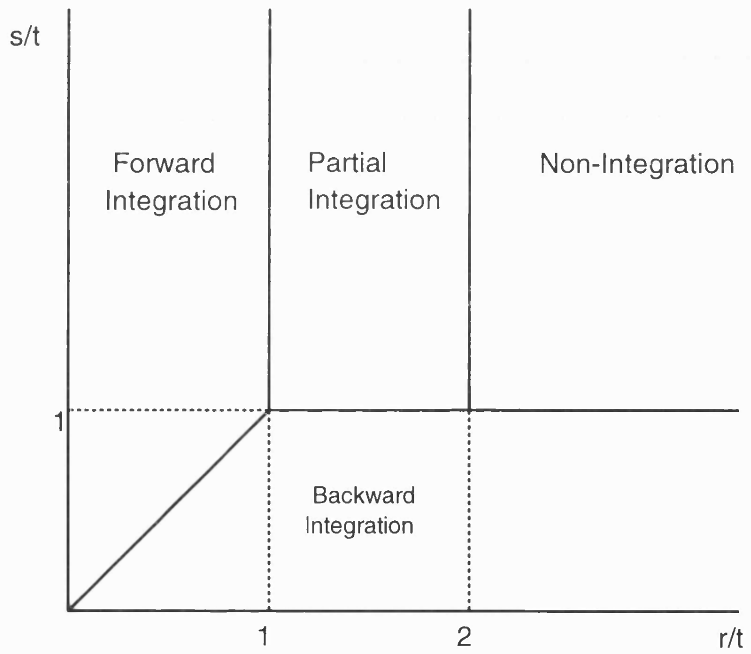
Proof: See Appendix 2.2.

---

[18] Suppose we allow w_i to own U_i, and m_i to own D_i. In the model as specified this "inverted" ownership allocation is never desirable. However **if** w_i could commit to locate next to U_i then D_i may locate next to U_i too. The location structure (and combined profit) is then identical to that obtained with backward integration. A mutual hostage-type scenario develops (cf. the value of reciprocal exposure in Williamson (1983)). D_i is prepared to locate next to U_i (though this exposes it to lock-in) because it in turn can hold-up w_i (given the latter's location commitment). This issue is explored further in Appendix 2.1.

Figure 2.5

Equilibrium Ownership Structure



BLPES FETCHING REQUEST SLIP

(one item per slip)

001831

REQUESTER'S NAME (Block Capitals) D. LYNN	DATE

TYPE OF USER

LSE STAFF/ STUDENT	EXTERNAL BORROWER ACADEMIC PERMIT	VACATION/STUDY PERMIT ORANGE DAY TICKET	FEE PAYERS

TYPE OF MATERIAL

PAMPHLET COLLECTION	COURSE COLLECTION SET TEXT	THESIS	U.N.	PARLIAMENTARY PAPER

OTHER (please specify)

CLASSIFICATION/ SHELF MARK F 7334

AUTHOR/TITLE/VOLUME/YEAR
(Please provide as much detail as possible)

WILLIAMS, PHILLIP
INCOMPLETE CONTRACTS, CONTROL
RIGHTS, ...

I AGREE TO RETURN THIS ITEM BEFORE THE SERVICE COUNTER CLOSES AND NOT TO REMOVE IT FROM THE LIBRARY

SIGNATURE:

The relationship between equilibrium ownership patterns and the key parameter ratios, s/t and r/t, is shown as a phase diagram in Figure 2.5. The obvious asymmetry is driven by the fixed location of upstream firms.

It is immediately apparent, in this more general model, that backward integration does not universally dominate all other ownership structures. With backward integration, upstream workers are non-owning workers. As a result they choose general, non-specific location at 1/2 - generating transport costs s/2. When the cost of transforming non-specific upstream human capital is relatively high (s/t high), upstream worker ownership of supplier assets is preferable. Upstream worker transport costs are then eliminated.

Note that for high s/t, the relationship between ownership patterns and r/t is that seen in the basic model i.e. Figure 2.3 corresponds to a horizontal cut through Figure 2.5, with s/t set high. Once again, though now in a more general setting, partial integration emerges as an equilibrium for a range of parameter values.

In general, when the costs of poorly tailored human capital are high relative to the costs of non-specific input (r/t and s/t high), strong worker-asset ties are valuable. Fragmented ownership structures are then attractive. Conversely, when r/t or s/t is low it is relatively important to secure strong physical asset ties. Integrated structures will then dominate, with upstream and downstream assets being owned together.

We can again consider whether the structures observed in equilibrium maximise the overall profits of all four workers (given that workers always choose location to maximise their individual profits).

Proposition 2.5 [Industry Optimal Ownership Structure]
No ownership structure achieves first best. The ownership equilibrium exhibits excessive tendencies towards full (forward and backward) integration. For a range of parameter values a partially integrated structure maximises overall profits, but full integration is observed in equilibrium.

Proof: See Appendix 2.2.

STAFF USE ONLY

MATERIAL NOT LOCATED

1. INCORRECT CLASSIFICATION/INSUFFICIENT INFORMATION
   (Please check the catalogue again)

2. MATERIAL IS ON THE OPEN SHELVES AT

   . . . . (. . . . . FLOOR)

3. NOT FOUND ON 1st FETCH . . . . . . .INITIALS . . . . HOUR
   (If UN/US refer to reference desk)

4. NOT FOUND ON 2nd FETCH . . . . . . .INITIALS . . . . HOUR

ITEM MISSING/NOT AVAILABLE

SHELF CHECKS . . .1 . . . . 2 . . . .

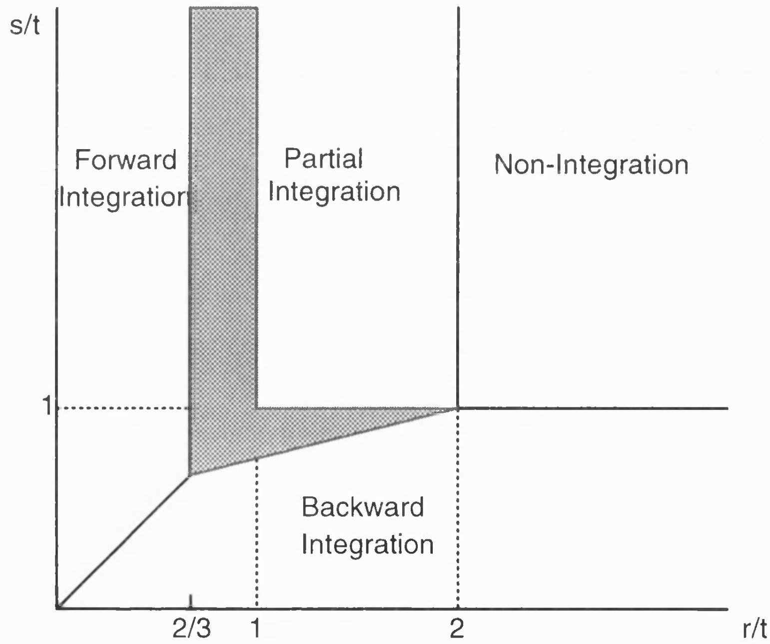
ITEM NOT RETURNED/IN USE

SHELF CHECKS . . .1 . . . . 2 . . . .

ITEM BARCODE:

Figure 2.6

Industry Optimal Ownership Structure



The shaded region in Figure 2.6 highlights parameter combinations where the equilibrium industry ownership structure is inefficient. In this region there is excessive integration.

For high values of s/t and low values of r/t in this region there is too much forward integration. The source of inefficiency in this case is precisely that highlighted in the basic model. The integrating firms do not take account of the adverse impact of their decision, in particular the effect of their downstream firm location choice, on the employment options of non-owning downstream workers. Consequently, increased human capital transport costs are incurred by other integrated firms.

For relatively low values of s/t and intermediate r/t values the cause of inefficiency is excessive backward integration. This case is a little more complex. When $s/t < 1$ non-integrated firms are attracted to the possibility of backward merger (and the subsequent adoption of idiosyncratic downstream technology). However in considering this transition (and the subsequent shift in downstream asset location) the integrating parties again fail to take account of the impact on non-owning downstream workers employment options. This adverse effect of backward merger may render forward integrated structures, where downstream workers are non-owners, unattractive to other parties. Since $s/t < 1$ and integration is preferable to non-integration, a fully backward integrated industry structure then results.

2.7 Limits to Horizontal Integration

In the simple model presented in Section 2.2 and extended in the last section horizontal integration was not permitted. However, our analysis has highlighted benefits to fragmentation. First, of course, more workers gain control of the assets they work with, encouraging asset-specific human capital investment. Secondly, fragmentation creates greater employment opportunities for non-owning non-owning workers. The mismatch costs of inappropriate worker investment are then reduced. These arguments suggest self-imposed limits to horizontal integration, even where this is permitted.

To examine the issue in more detail we add a third downstream firm (D_3) and an associated worker (w_3) to our industry. This firm sells one unit of output in a third, independent market. Values and costs are identical to those of the other downstream firms. At T=1 the owner of asset D_3 chooses its location on the unit line, while worker w_3 takes his/her location decision at T=2. We allow the new downstream firm to be owned either by downstream worker w_3 or downstream worker w_1 or upstream worker m_1. The ownership structure is chosen to maximise the joint profits of all three workers. To simplify our analysis, we will assume upstream assets are always owned and controlled by upstream workers. As before, this amounts to assuming that s/t is large.

There are four candidate ownership structures for assets D_1 and D_3:
Forward Integration - m_1 owns and controls D_1 and D_3.
Partial Forward Integration - m_1 owns D_1; w_3 owns D_3.
Horizontal Integration - w_1 owns D_1 and D_3.
Non-Integration - w_1 owns D_1, w_3 owns D_3.

D_2 may be upstream (Forward Integration) or independently (Non-Integration) owned. It should be clear therefore that 8 industry ownership configurations are possible.

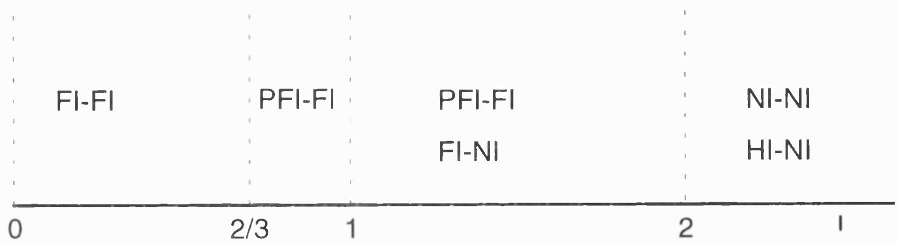
Given the independent final markets assumption, prices (Lemma 2.1) and location choice (Lemma 2.2) follow the patterns for the basic model. In particular:

(a) An owning downstream worker locates next to his/her downstream asset while a non-owning downstream worker locates between downstream firms.
(b) Integrated upstream and downstream firms locate next to each other. Non-integrated downstream firms locate between upstream firms i.e. at 1/2.

Again we are interested in the relationship between the allocation of ownership rights for the downstream assets, and the key parameter ratio - r/t. This is illustrated in Figure 2.7.

Figure 2.7

Equilibrium Ownership Structure



Result 2.8 [Equilibrium Ownership Structures]

(i) For a range of parameter values horizontal integration is limited in equilibrium.

(ii) Where non-integration is an equilibrium ownership structure, horizontal integration is also an equilibrium structure.

Proof: See Appendix 2.2

Figure 2.7 clearly shows that as r/t increases the industry becomes increasingly fragmented. As the costs of transforming non-specific human capital increase it becomes more and more important to ensure strong ties between downstream workers and assets. Dispersed ownership is therefore encouraged.

For 2/3 < r/t < 1 industry equilibrium structure involves upstream (integrated) ownership of downstream assets D_1 and D_2. However D_3 is downstream worker owned. An element of horizontal integration is absent, though weak ties between assets are relatively costly (r/t < 1). Indeed, the non-integrated downstream asset will be located at 1/2, generating additional input transport costs t/2 that exceed the direct benefits of closer worker-asset ties (- a cost saving r/2). However fragmentation has beneficial knock-on effects for the remaining, vertically integrated firm. Outside options for non-owning workers are improved, leading to a reduction in transport costs of r/4. This indirect benefit outweighs the direct costs of fragmentation. Of course, other vertically integrated firms benefit from this option too. Consequently some vertical integration is observed in equilibrium even when r/t > 1.

The above structure is also **an** equilibrium for 1 < r/t < 2. However, a configuration with fully (i.e. vertically and horizontally) integrated and non-integrated firms coexisting is also an equilibrium. Again some fragmentation in the industry as a whole, by altering the location decisions of downstream firms improves the employment options for non-owning downstream workers. As a consequence, the costs of worker human capital mis-match are ameliorated, and the costs of vertical integration to others reduced.

Not surprisingly, when unit worker transport costs are high (r/t > 2) we observe non-integrated structures in equilibrium. All three worker-owned downstream firms will locate together at 1/2, avoiding lock-in to either input supplier. Such non-specific investment imposes transport costs t/2 for input supply, but these are outweighed by close ties between downstream worker-owners and assets.

Note that lateral merger (D_1 and D_3 owned by w_1) is also an equilibrium for the very same parameter values. The three vertically separated downstream firms always locate at 1/2, irrespective of any horizontal merger. A perfectly competitive market for the services of downstream workers is therefore ensured, even if merger of two of the firms takes place (- two independent firms are sufficient for Bertrand competition). As such the non-owning worker (w_3) can locate at 1/2, adjacent to its employer yet avoiding any lock-in. Horizontal integration then imposes no additional transport costs.

Intuitively, once independent downstream firms locate together, idiosyncrasy is no longer an issue. The efficiency implications of ownership then disappear. In the absence of asset specificity concerns, standard competitive exchange is an effective "contracting process" (Williamson (1985), p.31-32).[19]

In this section we have shown that a self-imposed restriction on horizontal integration may be optimal. The basic mechanism at work in our model imposes natural, endogenously determined limits on the degree of horizontal as well as vertical integration. Of course, a commitment to independence is critical in deriving benefits from fragmentation. This issue is explored further in Section 8.

2.8 Renegotiation

In the basic model and extensions considered above we crucially assumed that ownership rights were allocated once and for all at T=0. Suppose instead that the allocation of ownership rights is renegotiable *after* location decisions have been

[19] We would expect this result to change if the model were adapted to allow for downstream differentiated product competition, for instance. The desire to secure a 'market niche' would then discourage downstream firms from agglomerating at the centre of the line.

made.[20] Since such investments are then sunk, renegotiation cannot directly lead to changes in location choice. However, anticipating that renegotiation may affect the future terms of trade, agents' location decisions will be sensitive to this possibility. To explore this issue we will focus on an ownership structure where D_1 and D_2 are vertically integrated, but D_3 is owned by downstream worker w_3 (see also Figure 2.8) i.e. the horizontal dimension to integration is limited.

Downstream asset D_3 is located next to its worker-owner w_3, to minimise mismatch. Both are located at 1/2 to minimise lock-in to either supplier. Non-owning downstream worker w_1 locates at 1/4 since this minimises lock-in to either the independent downstream firm (D_3) at 1/2, or the integrated downstream firm (D_1) at 0. Suppose however (as an illustration) that we allow trade of the ownership rights for asset D_3 after location choice, from w_3 to m_1 (the upstream owner of D_1). Further suppose take-it-or-leave-it offers from the buyer are considered by the owner. This will leave w_3's location decisions unchanged, since the aim is to maximise independent firm value (and guarantee the highest price from take-over).

Once investment has been undertaken, it is efficient for the downstream firms to integrate. Since location decisions are then sunk, integration imposes no further investment (location) costs. However, joint control of two downstream firms effectively reduces a non-owning downstream worker's employment options. This enables the asset owner to extract a greater share of the worker's surplus.[21] Of course the non-owning worker will foresee this threat and will locate between D_3 and D_2, at 3/4 (see Figure 2.8). [22]

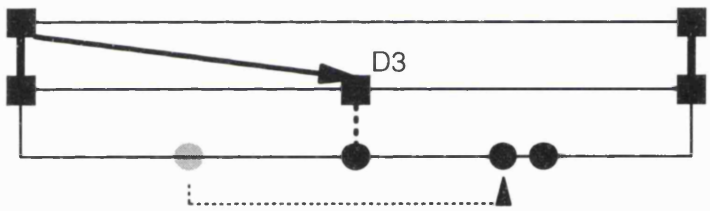
---

[20] Note that the other trades in the model are spot transactions and hence renegotiation proof.

[21] It is necessary to transfer ownership rights to achieve this since other collusion contracts are unenforceable.

[22] We cannot allow renegotiation to involve D_2 too, since w_1's outside options would then completely disappear. Such an eventuality cannot be permitted in the current model. It seems at least plausible that such complete ownership renegotiation might be proscribed by antitrust authorities.

Figure 2.8



As a result of this location change, both w_1 and w_3 will now work with asset D_3. In the particular context of our model this change in location decision and employer does not affect efficiency. In both cases non-owning workers locate distance 1/4 from their employers, and no extra losses are incurred when both workers are linked to the same firm. However, it is easy to envisage situations where mis-match costs would be increased. To take one example, one could imagine a scenario where D_3 must locate in the region [0,1/3] - in order to serve a particular market for instance. Whilst this would alter the details of our earlier analysis it is clear that the possibility of renegotiation increases mismatch costs (w_1 now locates at 2/3 rather than at 1/6).

Where ownership is observable the possibility of contracts restricting ownership changes, as in Section 2.4, may alleviate these problems. Agents could then commit to particular ownership structures before investments were made. In effect this brings us back to the situation studied in the sections above. Of course, if ownership is observable then the possibility of conditional contracts will allow the Industry Optimal Ownership Structure to be supported (see above), unless antitrust regulations bar such contracts. In practice, horizontal mergers-for-monopoly are prime candidates for antitrust investigation.

2.9 Conclusions

In this paper we model ownership decisions in a world where the specificity of investments can be chosen. Specific investments maximise value but the investor is vulnerable to hold-up if he/she becomes locked in to a particular relationship. Therefore, unless an agent is assured of full control, he/she is likely to choose a more general investment. General investment adds to costs but broadens alternatives.

When mis-match between a worker and an asset is relatively costly, that worker should generally be given control of the asset. Similarly when it important to form strong ties between assets those assets should be owned and controlled together.

The lock-in caused by a given investment depends on its value in alternative applications. For instance, if many firms use a highly specialised input then investment in the input technology could still be widely applicable. Much depends on the structure of the market for input. We show that it may be optimal to maintain a degree of fragmentation in asset ownership even when close ties between assets are relatively important. Independent firms provide outside opportunities for non-owning workers and so reduce the impact of worker-asset mismatch. Natural bounds are therefore placed on the degree of integration. In addition, since independent firms do not take account of their role in providing outside opportunities for the workers of other firms, externality effects may arise.

A key extension to the model would be the introduction of final market competition. The presence of independent firms is then more problematic - improved outside options for workers must be balanced with increased competition effects. Similar problems emerge for firms too. The need to find a market niche may prevent a general location choice

Appendix 2.1

Here we will consider the case for an "inverted" ownership structure, outlined in Footnote 7 i.e. downstream worker ownership of the upstream asset (w_i owns U_i) and upstream control of the downstream asset (m_i owns D_i). For simplicity we will focus on i=1.

Wherever there is downstream control of U_1, the upstream worker m_1 will always locate at 1/2, mid-way between U_1 and U_2, irrespective of other ownership and location decisions. The value of the worker's human capital to either firm is then reduced by a transport cost s/2. Clearly downstream control of U_1 will never be optimal when s is high.

As a benchmark, note that backward integration (downstream control of U_1 and D_1) yields a combined w_1-m_1 profit of:

$$\Pi_{m1\text{-}w1} = 3v - s/2$$

Both downstream worker and asset locate at 0, while the upstream worker locates at 1/2. We have already shown that with inverted ownership the upstream worker will always locate at 1/2, incurring transport cost s/2. Clearly, inverted ownership can never be strictly preferred to backward integration, since the transport costs incurred are always at least as large. It follows, a fortiori, that the inverted structure can emerge only where backward integration dominates all other ownership structures i.e. when $r/t > s/t$ and $s/t < 1$ (see Figure 2.4). Furthermore, to be attractive at all, the inverted structure must induce both w_1 and D_1 to locate at 0.

Throughout, we have emphasised the interrelationship between industry-wide location choice and the ownership and location decisions of individual agents. In the current context the location of downstream firm D_2 may influence downstream worker w_1's own decision. For parameter values that can sustain the inverted structure i.e. $r/t > s/t$, $s/t < 1$, D_2 will always locate at 1 in equilibrium. Again, backward integration is the benchmark scenario.

If w_1 is unable to commit to location before D_1 is located, the results of our basic model apply. In particular, non-integrated D_1 will always locate at 1/2 thus reducing value by a transport cost t/2. The inverse structure will then never be chosen by w_1 and m_1 at T=0, since backward integration yields lower costs irrespective of w_1's location decision.

Now suppose that w_1 **can** commit to a location decision. The inverted ownership structure will be attractive only if w_1 will commit to location 0 and given this commitment, D_1 will locate at 0 too. We will proceed in two stages. First we consider whether D_1 will locate next to w_1, and second whether w_1 will locate at 0.

Suppose w_1 commits to location α (where $\alpha < 1/2$ w.l.o.g.), and let D_1 locate at β. There are two opposing forces at work influencing D_1's decision. To maximise supplier options D_1 would like to locate at 1/2, between U_1 and U_2. If D_1 locates closer to α it is also closer to U_1, and is then exposed to some supplier lock-in. However, in also locating closer to the downstream worker, D_1 is able to extract greater profit from the employment relationship with w_1.

In choosing β, where $\alpha \leq \beta \leq 1/2$, D_1 will trade off input supply costs $[t(1-\beta)]$ against its share of employee value $[r(1-\beta)]$. It's profit is then given by:

$\Pi_{D1} = v - t(1-\beta) + r(1-\beta) = v + r - t + (t - r)\beta$

Clearly, if r/t > 1 then $\beta = \alpha$, while if r/t < 1 then $\beta = 1/2$. Thus D_1 will locate next to w_1 at α (given the latter's location commitment) provided r/t > 1.

Assuming D_1 will locate at α, will w_1 choose to commit to $\alpha=0$? In locating at 0 w_1 increases its lock-in to D_1, since the cost of working for D_2 (at 1) is maximised. However, if D_1 locates at α then choosing $\alpha=0$ also maximises D_1's lock-in to the w_1 owned supplier, U_1. Reciprocal lock-in is therefore maximised. If $\beta = \alpha$ i.e. D_1 locates next to w_1, these two components of w_1's profits are given by:

$\Pi_{w1-U1} = [v - r(1-\alpha)] + [t(1-2\alpha)] = v + t + r + (r-2t)\alpha$

Clearly if r/t < 2 w_1 will choose $\alpha = 0$, provided D_1 will also locate at 0 (which it will do if r/t > 1).

Therefore if 1 < r/t < 2 w_1 will commit to location 0 (if it commits at all), and D_1 will also choose to locate at 0. This yields overall profits of 3v - s/2, hence the inverse structure could be observed in equilibrium. Finally, we must now show that w_1 will indeed choose to commit to location (taking ownership as given).

Suppose that w_1 does not commit. As explained above D_1 then locates at 1/2, and w_1's profits are given by:

$$\Pi_{w1\text{-}U1} = v - r/4$$

If D_1 locates at 1/2 w_1's optimal location is at 3/4, 1/4 unit from both D_1 and D_2.

Commitment to location 0 is therefore worthwhile to w_1 if $v + t - r > v - r/4$ i.e. if $r/t < 4/3$.

Tying all this together, it is possible that the inverted ownership structure (downstream control of upstream assets and upstream control of downstream assets) will exist in equilibrium if:

(a) w_1 can commit to location before D_1 locates.

(b) Upstream worker transport costs are relatively low i.e. $s/t < 1$.

(c) Relative downstream transport costs lie in the range $1 < r/t < 4/3$.

Appendix 2.2

Proof of Result 2.1 [Prices]:

(a) If the downstream firm locates at c then the winning bid is given by

$$p = t \max[c, 1-c].$$

The cost to U_1 of supplying input is tc, therefore U_1 will demand a price of at least tc before supplying.

The cost to U_2 of supplying input is t(1-c), therefore U_2 will demand a price of at least t(1-c) before supplying.

Suppose w.l.o.g. that c < 1-c. then U_1 offers to supply for $p^* = t(1-c) > tc$ (less an infinitesimal amount if trading with D_2) and is accepted.

If U_2 offers $p < p^*$ then $\Pi_{U2} = p - t(1-c) < 0$.

If U_1 offers $p < p^*$ then still accepted but $\Pi_{U1}(p) - \Pi_{U1}(p^*) < 0$.

If U_1 offers $p > p^*$ then U_2 can offer p^+ where $p > p^+ > p^*$ and be accepted. Then $\Pi_{U2} = p^+ - t(1-c) > 0$, $\Pi_{U1}(p) = 0 < \Pi_{U1}(p^*)$.

If U_2 offers $p^* = t(1-c)$ and downstream firm is D_1 then U_1 still sells unit at $p^* = t(1-c)$ - due to tie-breaking assumption.

If U_2 offers $p^* = t(1-c)$ and downstream firm is D_2 then U_2 captures sale.

Then $\Pi_{U1}(p^*) = 0$. By shading price infinitesimally U_1 captures sale and $\Pi_{U1}>0$.

If c > 1-c then U_2 offers to supply for p = tc (plus a very small amount) and is accepted.

(b) Suppose D_1 locates at c_1, D_2 at c_2 and w_i at b. Assume w.l.o.g. that $c_2>c_1$.

The value of the worker's skills to $D_1 = v - r|b-c_1|$

The value of the worker's skills to $D_2 = v - r|b-c_2|$

Therefore D_1 is prepared to pay up to $v - r|b-c_1|$ while D_2 is prepared to pay up to $v - r|b-c_2|$ for the worker's services.

The winning price is $p = v - r \max[|b-c_1|,|b-c_2|]$.

(The firm with the highest valuation just outbids the low valuation firm)

Suppose w.l.o.g. that $|b-c_1|<|b-c_2|$.

Then D_1 offers $p^* = v - r|b-c_2|$ (plus an infinitesimal amount to w_2) and wins.
If D_2 offers $p > p^*$ then $\Pi_{D2} = v - r|b-c_2| - p < 0$.
If D_1 offers $p > p^*$ then still wins and $\Pi_{D1}(p)-\Pi_{D1}(p^*) < 0$.
If D_1 offers $p < p^*$ then D_2 can offer p_2 such that $p < p2 < v\text{-}r[b\text{-}c]$ and win.
Then $\Pi_{D2}(p_2) = v - r|b-c_2| - p_2 > 0$, $\Pi_{D1}(p)=0 < \Pi_{D1}(p^*)$.

Again if worker is w_1 then D_1 captures services for p^* (tie-breaking assumption).
For w_2 D_1 must top D_2's best price (p^*) slightly.

If $|b-c_1|>|b-c_2|$ then D_2 offers $p = v - r|b-c_1|$ (plus a very small amount to w_1), and is accepted.

Prices in the general case:
Prices for downstream worker services and input are as in Result 1. If an upstream worker locates at a then he/she receives a price

$$p = v - s\max[a,1-a]$$

If $a>1/2$ the worker works for U_2 whereas if $a<1/2$ he/she works for U_1.
For $a=1/2$ worker i works for U_i (tie-breaker assumption).

The proof of this last result follows the pattern for the downstream workers in (b) above.

Profits for upstream workers are therefore given by:
$\Pi_{m1} = v - s\max[a_1,1-a_1]$
$\Pi_{m2} = v - s\max[a_2,1-a_2]$

Proof of Result 2.2 [Location Choice]:

(a) From Result 1 we know that the downstream worker receives a price for his/her services $p = v - r\max[|b-c_1|,|b-c_2|]$.

The non-owning worker chooses location, b, to maximise profits $\Pi_w = p$.

Suppose $|b-c_1| > |b-c_2|$, and assume w.l.o.g. that $c_2 > c_1$. Then either $b-c_1 > b-c_2$, or $b-c_1 > c_2-b$.

In both cases $\Pi_w = p = v - r(b-c_1)$. The price, and hence the profit, the worker receives can therefore be increased by reducing b i.e. locating closer to D_1:

$\Pi_w(b-e) - \Pi_w(b) > 0$.

Similarly if $|b-c_1| < |b-c_2|$ then the worker can increase profit by increasing b i.e. locating closer to D_2.

Profit Π_w is therefore maximised when $b-c_1 = c_2-b$ i.e. $b = (c_2+c_1)/2$.

If the downstream worker 1 (w_1) owns downstream asset 1 then location (b_1) is chosen to maximise

$$\begin{aligned}\Pi_{w1\text{-}D1} = 2v \quad &+ r\max[0,|b_1-c_2|-|b_1-c_1|] + r\max[0,|b_2-c_2|-|b_2-c_1|] \\ &- r\max[|b_1-c_1|,|b_1-c_2|] - t\max[c_1,1-c_1] \\ = 2v + r\max[0,|b_2-c_2|&-|b_2-c_1|] - t\max[c_1,1-c_1] - r|b_1-c_1|\end{aligned}$$

Clearly this is maximised setting $b_1=c_1$.

Similar reasoning applies to w_2.

(b) The proof of (b) follows the structure of the proof of (a). We know from (a) that non-owning downstream workers locate between downstream firms (and by assumption worker i works for firm i when indifferent). We also know that if downstream asset i is downstream owned then w_i locates next to the asset. Hence if D_1 (for instance) is downstream owned, profits are given by:

$\Pi_{D1} = v - t\max[c_1,1-c_1] + r\max[0,|b_1-c_2|]$

Clearly profit maximising involves setting $c_1=1/2$.

If D_1 is owned with U_1 then given the reaction of workers, the profits of the integrated structure are given by:

$$\Pi_{U1\text{-}D1} = v - t\max[c_1, 1-c_1] + t\max[0, 1-2c_1]$$

$$= v - tc_1$$

Clearly profit maximisation involves setting $c_1=0$

Similar reasoning applies to D_2.

Location in the general case:

The location decisions for downstream firms and workers are the same as those in Result 2. An owning upstream worker (m_i) locates next to the upstream asset U_i. A non-owning upstream worker locates between upstream assets.

The proof of this last result follows the structure of the proof of (a) for Result 2. Non-owning upstream worker m chooses location a to maximise

$$\Pi_m = p = v - s\max[a, 1-a].$$

Clearly this involves setting a=1/2

If m_1 owns U_1 (for instance) it chooses location (a_1) to maximise

$$\begin{aligned}\Pi_{m1\text{-}U1} = \quad & v + t\max[0, 1-2c_1] + t\max[0, 1-2c_2] + s\max[0, 1-2a_2] \\ & + s\max[0, 1-2a_1] - s\max[a_1, 1-a_1] \\ = \quad & v + t\max[0, 1-2c_1] + t\max[0, 1-2c_2] + s\max[0, 1-2a_2] - sa_1\end{aligned}$$

Clearly this involves setting $a_1=0$.

Similar reasoning applies to m_2.

Proof of Result 2.6 [Equilibrium Ownership Structure]:

Both sets of upstream and downstream workers choose the ownership structure of assets simultaneously at T=0. The following matrix gives the joint profits of each set of workers for each ownership combination (derived using the price and location choice results above) **net** of the fixed value component - 3v - which is independent of structure. Clearly the differences in profits are driven by transport cost differences.

		2: FI	2: NI	2: BI
1	FI	-r/2, -r/2	-r/4, -t/2	-r/2, -s/2
1	NI	-t/2, -r/4	-t/2, -t/2	-t/2, -s/2
1	BI	-s/2, -r/2	-s/2, -t/2	-s/2, -s/2

Now if r>s and s/t<1 then downstream control is the only equilibrium strategy for each set of workers. We will therefore observe the Full Backward Integration overall ownership structure.

If r<s and r/t<1 then upstream control is a dominant strategy for both sets of workers, and we will observe Full Forward Integration in equilibrium.

If s/t>1 and r/t>2 then non integration is a dominant strategy for each set of workers and we will see Non-Integration as the equilibrium structure.

If s/t>1 and 1<r/t<2 then the possible equilibria involve both Forward Integrated and Non-Integrated structures i.e. Partial Integration.

This covers all possible values of r/t and s/t.

Proof of Result 2.7 [Industry Optimal Ownership Structure]:

Below we list the total profits of all four workers **net** of the fixed value - 3v - for all possible overall ownership structures (excluding symmetric equivalents):

Π(FI-FI) = -r

Π(NI-NI) = -t

Π(BI-BI) = -s

Π(FI-NI) = -t/2 - r/4

Π(FI-BI) = -s/2 - r/2

Π(NI-BI) = -t/2 - s/2

We can immediately rule out structures FI-BI and NI-BI which are (weakly) optimal for only an infinitesimal area of our parameter space

FI-FI is optimal if $r<t$, $r<s$ and $r<t/2+r/4$ i.e. $r/t<1$, $r/t < s/t$ and $r/t < 2/3$.

NI-NI is optimal if $t<r$, $t<s$ and $t<t/2+r/4$ i.e. $r/t > 1$, $s/t > 1$ and $r/t > 2$.

BI-BI is optimal if $s<t$, $s<r$ and $s<t/2+r/4$ i.e. $s/t < 1$, $s/t < r/t$, $s/t < r/4t + 1/2$.

FI-NI is optimal if $t/2+r/4<r$, $t/2+r/4<s$, $t/2+r/4<t$ i.e. $r/t > 2/3$, $r/t < 2$,

and $1/2 + r/4t < s/t$.

Proof of Result 2.8 [Equilibrium Ownership Structures]:

The matrix below gives the joint profits net of fixed values for (i) m_1,w_1 and w_3 and (ii) m_2 and w_2, for all possible industry ownership structures:

	FI	NI
FI	$-r$, $-r/2$	$-r/2$, $-t/2$
PFI	$-r/4-t/2$, $-r/4$	$-r/4-t/2$, $-t/2$
NI	$-t$, $-r/2$	$-t$, $-t/2$
HI	$-r/4-t$, $-r/2$	$-t$, $-t/2$

There are 8 ownership configurations. The conditions for each to be an equilibrium are given below:

NI-NI $r/4>t/2$, $r/2>t$, $r/4+t/2>t$ i.e. $\mathbf{r/t > 2}$

FI-NI $r>t$, $r/2<t$, $r/2<r/4+t/2$ i.e. $\mathbf{r/t > 1, r/t < 2}$

HI-NI $r/4>t/2$, $t<r/2$, $t<r/4+t/2$ i.e. $\mathbf{r/t > 2}$

FI-FI $r<t$, r/t, $r<r/4+t/2$ i.e. $\mathbf{r/t < 1, r/t < 2/3}$

PI-FI $r/4<t/2$, $r/4+t/2<r$, $r/4+t/2<t$ i.e. $\mathbf{r/t < 2, r/t > 2/3}$

HI-FI $r/4+t < t$ - no!

PI-NI $t/2<r/4$, $r/4+t/2<t$, $r/4+t/2<r/2$ i.e. $\mathbf{r/t > 2, r/t < 2}$ - no!

NI-FI $r/4<t/2$, $t<r$, $t<r/4+t/2$ i.e. $\mathbf{r/t < 2, r/t > 1, r/t > 2}$ - no!

References

Bickenbach, F. and I. Williams (1991): "Investment Specificity, Vertical Integration and Market Foreclosure". Mimeo.

Bolton, P. and M. Whinston (1990): "Incomplete Contracts, Vertical Integration and Supply Assurance". Harvard University Discussion Paper, No.1493.

Coase, R.(1937): "The Nature of the Firm". Economica, 4: 386-405.

Farrell, J. and N.Gallini (1988):"Second Sourcing as Commitment: Monopoly Incentives to Attract Competition". Quarterly Journal of Economics, 103: 673-694

Grossman, S. and O. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration". Journal of Political Economy, 94: 691-719.

Hart, O. and J. Moore (1990): "Property Rights and the Nature of the Firm". Journal of Political Economy, 98: 1119-1158.

Holmstrom, B. and J. Tirole (1991): "Transfer Pricing and Organizational Form". Journal of Law, Economics and Organization, 7: 201-228.

Klein, B., R. Crawford and A. Alchian (1978): "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", Journal of Law and Economics, 21: 297-326.

Lyons, B. (1991): "Specialised Technology, Economies of Scale and the Make-or-Buy Decision: Evidence from UK Engineering". University of East Anglia Economics Research Centre Discussion Paper No. 9126.

Riordan, M. and O. Williamson (1985): "Asset Specificity and Economic Organization". International Journal of Industrial Organization, 3: 365-378.

Shepard, A. (1987): "Licensing to Enhance the Demand for New Products". RAND Journal of Economics, 18: 360-68.

Williamson, Oliver E.(1975): "Markets and Hierarchies: Analysis and Antitrust Implications". Free Press, New York.

Williamson, Oliver E. (1983): "Credible Commitments: Using Hostages to Support Exchange". American Economic Review, 73: 519-540.

Williamson, Oliver E. (1985): "The Economic Institutions of Capitalism". Free Press, New York.

Chapter 3:

Investment Specificity, Vertical Integration and Market Foreclosure [23]

[23] This chapter is based on joint work with Frank Bickenbach

3.1 Introduction

Do vertical mergers enhance efficiency or promote market power? A vigorous debate has raged between proponents of the rival efficiency and anti-competitive views. Much of this attention has focused on the possibility of integration induced market foreclosure. A strident attack on the anti-competitive case has been provided by Robert Bork. He argues strongly that "*the law against vertical merger is merely a law against the creation of efficiency*" [Bork (1978), p.234]. He is dismissive of the view that vertical merger will lead to distortion in the supply and purchase prices faced by non-integrated rivals. Integration is seen as an efficient response to transaction cost concerns. Oliver Williamson, too, takes this view of "*the main purpose served* (by vertical integration)*: economising on transaction costs*" [Williamson (1985), p.85-86]. In response to this challenge, a number of recent papers have attempted to set the competition perspective on a firmer theoretical foundation. Two distinct lines of argument have emerged.

The first, more traditional, approach emphasises the effects of integration on post-merger price setting. Papers in this vein include Salinger (1988), Ordover, Saloner and Salop (1990) and Hart and Tirole (1990). They focus on the impact a vertical merger has on supply pricing - both for the integrating buyer and for non-integrated rivals. In particular, the upstream division of an integrated firm will internalise the impact on the downstream division of any supply agreements undertaken. Essentially, in incomplete contract settings where commitment is difficult if not impossible to enforce, integration substitutes for contractual solutions in implementing a joint profit maximising pricing strategy.

A second, more radical approach, was initiated in an important paper by Patrick Bolton and Michael Whinston [Bolton and Whinston (1993)]. Their argument strikes at the transaction cost basis that underlies much of the benign view of vertical merger. Building on the incomplete contract approach to integration formulated in Grossman and Hart (1986), they show that the very investment incentive effects that encourage efficient vertical integration in the bilateral context can lead to inefficiency in a multilateral environment. When a supplier integrates with one of its buyers, over-investment in the internal relationship is encouraged. This distorts supply patterns, and may lead to inefficient foreclosure of non-integrated rivals. Significantly, such foreclosure is not driven by changes in post-investment pricing behaviour.

In this paper we will follow the Bolton-Whinston approach, emphasising the importance of investment *specificity* in a multilateral context. Within this framework the interaction between efficiency and anti-competitive motives for integration will be explored. In particular, we extend the basic Bolton-Whinston framework to examine the effects of vertical integration on horizontal downstream competition, where we believe our specificity approach is particularly insightful.

We consider a simple setting where a single retailer can be supplied by either one or two differentiated manufacturers. The retailer must make three sequential decisions: (i) whether to integrate with one of the suppliers, (ii) whether its technology should be largely specific to a single supplier or flexible, and (iii) whether to source product from one or both manufacturers. In turn, a number of key factors drive the outcome of this decision process: (a) the (exogenous) cost of merger, (b) the relative supply costs for specific and general technology, (c) the relative values of single versus multi-product retailing, and (d) the toughness of inter-retailer competition. We will focus on the interaction between these factors and the retailer's decisions.

A location framework is a natural setting for our analysis. With the suppliers fixed at opposite ends of a unit line, the retailer's once-and-for-all location decision represents its investment choice. A position at either end of the line constitutes investment that is highly specialised towards a particular manufacturer, while a mid-point location indicates general purpose investment. Locating at one end of the line minimises the cost of transacting with the adjacent manufacturer, but maximises the transport costs of exercising the alternative supply option. A mid-point location ensures the widest range of supply options.

In an incomplete contract environment, where enforceable contracts on future supply terms cannot be written, an independent retailer may be reluctant to locate at either end of the line. Once such location commitment is made, the retailer may effectively be locked-in to a particular supply relationship. Indeed, whenever single-product retailing is attractive, the independent retailer will always locate at the mid-point of the unit line, maximising supply options. Yet, in precisely these circumstances, an end-point location will be efficient, minimising transport costs. Vertical merger may then be attractive, overcoming lock-in concerns and encouraging adjacent retailer-supplier location. Of course, even then, integration will only take place if the resulting transport cost benefits outweigh the exogenous merger costs. However, in this case, when integration is individually attractive it will always increase overall welfare.

Costly integration will only occur if it leads to a *different* retailer location decision. This may, but need not, affect the retailer's product sourcing decision i.e. leading to single product rather than multi-product retailing. Where this change does occur, the supply market for the non-integrated manufacturer is foreclosed. Since the integrating parties take no account of such third party foreclosure effects in deciding to merge, a negative externality results. Clearly, inefficient foreclosure may therefore occur.

However, foreclosure need not always be inefficient. A move from double sourcing at the mid-point of the unit line to integrated single-product retailing at the end-point involves a saving on transport costs. This efficiency gain counteracts the loss of multi-product surplus. Any anti-competitive conclusion deduced from foreclosure therefore depends delicately on the resolution of the trade-off between these opposing forces.

In the context of our model, when multi-product retailing is optimal wherever the retailer locates, the lock-in threat will not materialise. Consequently, the retailer is indifferent to location in these circumstances. Integration may still take place, if it results in a different retailer location decision. Efficiency implications then depend solely on the structure of transport costs. Privately attractive but inefficient integration may occur.

Our results also hold for a simple re-interpretation of the model, where a monopoly manufacturer must decide to supply one or two retailers. The potential anti-competitive effects of integration can then be viewed as impacting on consumers. In this alternative context, upstream capacity constraints are important for our results to hold. The analogue of single product purchasing is then single retailer supply, thereby creating a retailing monopoly. Again, vertical integration may bring about such an exclusive dealing arrangement, where standard contractual mechanisms fail.

In a key development to our basic model we introduce a second downstream retailer. This allows us to consider final market competition and its endogenous relationship with vertical integration. We will demonstrate that competition considerations may drive the integration process, which will in turn affect the competitive environment (via location choice). Such interactions can generate partially integrated equilibria, where integrated and non-integrated firms co-exist, and chains of integration, where the integration decisions of individual firms are driven by industry-wide merger activity.

The new force driving vertical integration is a desire to alleviate the pressures of downstream competition. One obvious strategic response is product differentiation (see e.g. Porter (1980) for a detailed discussion). In the context of our model, a natural means of achieving this is for the retailers to locate at opposite ends of the unit line. Making idiosyncratic, supplier specific investments enables retailers to enhance their "uniqueness".

Of course, non-integrated, independent retailers will be reluctant to invest in supplier-specific technology. Instead, they are likely to locate together, at the centre of the unit line - avoiding lock-in to either manufacturer, but intensifying downstream competition. Vertical integration of retailer and supplier, by eliminating lock-in concerns, provides a mechanism for encouraging competition-reducing differentiation. The potential anti-competitive effects of vertical integration are then clear-cut. Retailers can locate at the ends of the line without fear of hold-up. Furthermore, close location ties between manufacturing and retailing stages are generated. Such vertical linkages are often cited as a source of advantage to firms in competitive environments. Against these benefits will be set the inevitable costs of integration.

In his seminal analysis of the emergence of *big business* in the United States, Alfred Chandler has emphasised the importance of the vertical dimension in firm organisation (see e.g. Chandler (1959) and (1977)). He stresses the growing value of integrated production and distribution as specialised, idiosyncratic marketing services were required by manufacturers - for instance, with sewing machines (Singer) and advanced agricultural machinery (McCormick). Chandler highlights the scale effect of market expansion, through urban growth and the emergence of the railway network, as the driving force in encouraging integrated coordination. However, with an improved transport infrastructure, the growth in competitive pressure was a vital factor too. Indeed, the value to producers of product differentiation, effected through vertical integration into marketing, was increasingly recognised as a fruitful response to tougher competition in consumer goods industries (Chandler (1959), p.11-12).

In our extended model, we find that competition considerations may result in integration where it would otherwise not take place. Partial integration is the industry equilibrium response to moderate levels of competition. Integration by a subset of retailers is then sufficient to ease competitive pressures. Other firms can benefit, without incurring the additional costs of merger. However, as competition intensifies, a differentiation strategy may only succeed if *all* retailers participate, implicitly coordinating their actions. Individual integration decisions will then only be taken in response to (or in anticipation of) wider merger activity. Such chains of integration generate a fully vertically integrated industry.

Our basic model framework is set out in the next section. This is followed by a numerical example that highlights some of the main issues raised in the paper. Our formal analysis is presented in Section 3.3. In Section 3.4 we briefly outline two simple variations to the basic model. The basic framework is extended in Section 3.5 to consider the impact of final market competition.

3.2 The Basic Model

The basic model involves three interacting firms. A downstream retailer D can stock the products of two potential suppliers, U1 and U2. The upstream firms produce differentiated products, represented by their (fixed) locations at the ends of a unit line. U1 is located at 0, U2 at 1. D however can choose its location, l. For simplicity we allow it to locate at 0 or 1/2. Model symmetry allows us to ignore the possibility of location at 1. If D chooses l=0 its assets are tailored to U1 supply while a mid-point location (l=1/2) indicates investment in general technology.

In purchasing product from upstream firms, D must decide whether to source from a single firm, or both firms. In making its decision the retailer will be influenced by the values of stocking a unique product line (m) relative to that of stocking product from both upstream firms (2d, or d per product line). We will assume that m > d. Note, however, that either multi-product or single product retailing may generate greater overall value.

Product must be transported from the upstream supplier to the retailer. The cost of transporting product from U1 is denoted by t1 while the transport cost for U2's product is t2. The magnitudes of these transport costs will depend on D's location. We will assume here that the total transport cost is independent of the volume stocked from a given supplier (but see Section 5(i)). This cost can therefore be thought of as the set-up cost of stocking a particular product line.

The cost of transporting supplies over half the unit line is given by $\underline{t}$, while the transportation cost from end to end is $\bar{t}$, where $\bar{t} > \underline{t}$. No transport cost is incurred if retailer and supplier locate together. For example, if D locates at 0 no costs are incurred in purchasing product from U1, but purchasing product from U2 involves a transport cost $\bar{t}$. We will assume that $d > \bar{t}$, ensuring that both upstream firms will always be active in the supply market, irrespective of downstream firm location. Any additional downstream costs are embedded in the product values, m and d.

For simplicity, we will suppose that D buys product sequentially, in two batches. D receives sealed bids from both upstream firms to supply each batch. It must then make simple accept/reject decisions. Given a first batch sourcing decision, the retailer must then decide whether to stock a single product line, by purchasing the second batch from the same supplier, or to opt for double sourcing by purchasing the second product batch from a new supplier. Bidding for the second batch takes place after the first supply contract has been awarded.

Ours is a world of incomplete contracts in which supply prices and allocations, as well as location choice, cannot be enforceably specified at T = 0. The idea here is that supply requirements are uncertain at T=0 and location (quality) can never be verified by an outside arbiter. Profit sharing agreements between independent firms cannot be implemented either.[24] The only effective T=0 contracts are those conferring asset ownership rights. Note that, in the absence of enforceable price or profit agreements, exclusive dealing contracts are never attractive to the retailer and will be ignored in our analysis.

We will consider the effects of vertical integration between the downstream firm and one of the upstream firms. Model symmetry allows us to restrict attention to the possibility of D-U1 integration only. Horizontal integration is not permitted, on anti-trust grounds. The integrated firm is assumed to maximise the joint profits of both its upstream and downstream divisions. A fixed cost E is incurred in undertaking a vertical merger. This is intended to broadly capture the legal, bureaucratic and incentive costs of integration. It is relatively simple to generate such a cost explicitly (e.g. as outlined in the model of Chapter 1), but for simplicity we do not do so here.

At T=1, uncertainty about supply needs is resolved and D decides on a location (to maximise profits). Supply are then purchased (T=2). After transporting supplies the final output is sold to consumers (T=3) and revenues are realised.

[24] For a discussion of the profit sharing assumption, with and without integration, see Hart and Tirole (1990).

3.3 Analysis

As a benchmark for our analysis we will first consider efficiency. We assume that the monopoly downstream firm is able to extract all consumer surplus. In considering efficiency we can therefore restrict attention to the combined profits of the three firms. The efficiency of two decisions must be considered - the product sourcing decision and the downstream firm's location choice.

At each location there is a supply choice. Product can be sourced from U1 alone, yielding value m-t1. Alternatively, product can be sourced from both upstream firms, generating an overall profit of 2d-t1-t2. Note that our assumptions (innocuously) rule out single sourcing from U2 since $t1 \leq t2$.

For a given location, efficient sourcing implies combined profits of Π^*, where:

$$\Pi^* = \max[m-t1, 2d-t1-t2]$$

The efficient choice of D's location, either at 0 or 1/2, must also be considered. We first derive the efficient supply allocation for a given location, and then determine the efficient location.[25]

Proposition 3.1 [Efficiency]

(a) For a given location, efficiency implies:

U1 supplies both batches if $t2 > 2d-m$

U1 and U2 supply one batch each if $t2 < 2d-m$

(b) The efficient location choice is given by:

$l=0$ if $2(d - \underline{t}) < \max[m, 2d-\bar{t}]$

$l=1/2$ if $2(d - \underline{t}) > \max[m, 2d-\bar{t}]$

Proof:

(a) Our assumption that $t1 \leq t2$ guarantees that it always optimal to source at least one unit from U1. If $t2 > 2d - m$ the value of sourcing both units from U1 (m-t1) exceeds that of double sourcing (2d-t1-t2). Clearly if $t2 < 2d - m$ the reverse

[25] To ease notation we will drop the location argument, l.

is true.

(b) If l=0 the transport costs for U1 supply are 0, and for U2 supply are $\bar{t}$. Thus efficient profits are max[m, 2d-$\bar{t}$]. Similarly when l=1/2 both Us have transport costs $\underline{t}$ so efficient profits are max[m-$\underline{t}$, 2(d-$\underline{t}$)]. The efficient location choice is then determined by which of these expressions is the larger. Since m > m-$\underline{t}$ locating at 1/2 is only efficient if 2(d-$\underline{t}$) > max[m, 2d-$\bar{t}$]. QED.

Stocking both products as opposed to one involves an additional transport cost (t2). Double sourcing is optimal only if the added value of multi-product retailing exceeds this cost. If 2d-m < 0, then stocking a single product is a fortiori optimal.

It is important to recognise that, because of changes in transport costs, the efficient sourcing decision may (but need not) depend on location choice. If 2d-m > $\bar{t}$ both products will always be stocked, irrespective of D's location. Transport cost considerations alone then determine the optimal location - two costs of $\underline{t}$, for l=1/2, being compared with one cost of $\bar{t}$ for l=0. If 2$\underline{t}$ > $\bar{t}$ an end-of-line location is optimal, even if product is sourced from both upstream firms. Clearly, when single sourcing is optimal D should always locate at l=0. Stated differently, location at 1/2 is efficient only if multi-product sourcing is efficient.

Finally, note that if $\underline{t}$ < 2d-m < $\bar{t}$ a change in D's location choice from l=1/2 to l=0 will result in an efficient switch from double to single sourcing. Choosing l=0 will then be efficient, if the transport cost savings (2$\underline{t}$) exceed the loss of the incremental value of multi-product retailing (2d-m).

Integration incurs additional costs E and no benefits, *for a given location choice*. From a first best perspective, integration is therefore never desirable. However, where D's self-interest drives its location decision, integration alone may bring about efficient location choice - as we will see below.

The integration and location decisions made in equilibrium will depend on the overall magnitude of profits and their division among firms. Competition between suppliers clearly plays a critical role. The following result summarises the outcome of supply competition, for given location.

Lemma 3.1 [Supplier Competition]

(a) Sourcing decisions are always efficient given location.

(b) U1 always supplies the first unit, for a price 2d-m-t1

(c) The second unit is supplied by:

U1 if $2d-m-t2 \leq 0$

U2 if $2d-m-t2 > 0$

The price paid is $|2d-m-t2|$

(d) In the non-integrated case profits are given by:

$\Pi_D = m-d + \min[m-d, d-t2]$

$\Pi_{U1} = \max[2d-m-t1, t2-t1]$

$\Pi_{U2} = \max[2d-m-t2, 0]$

In the integrated case profits are given by:

$\Pi_{U1\text{-}D} = m-t1-E$

$\Pi_{U2} = \max[2d-m-t2, 0]$

Proof: See Appendix.

The basis for the above results is the fact that the upstream firm with the higher value supply can always undercut its rival, and will do so. The supply prices are determined by the value of D's best alternative i.e. by the second highest value supply option. If product is sourced from both Us, then D's outside option in negotiating with either supplier is single sourcing. The differential value added is then m-(2d-t2). In our basic model this is independent of transport costs. The supplier captures the residual value.

Where product is sourced from a single supplier (always U1 in our model) then D's outside option, and hence its total share of trade value, is m-t2. This is the value of exclusive sourcing from U2. The result is driven by the perfect foresight of suppliers during the supply process. In dealing with U1, D's threat point is U2 supply. Such trade would yield D a payoff of d-t2 for a single unit, and m-t2 for two units. Clearly if one unit were to be bought from U2 then U1 could win the competition for the supply of the second unit by offering D a share m-d of trade, again yielding D an overall profit of m-t2.

Given that an identical supply competition is used when integration occurs, supply price and allocation are independent of this decision.[26] Thus, for a particular location, integration does not alter trade returns (though integration costs E are incurred). The effects of integration must therefore arise from induced changes in D's location choice. It is to this issue that we now turn.

Lemma 3.2 [Location]

(a) Under non-integration D chooses location l=1/2 if 2d-m < $\bar{t}$ and is indifferent about location otherwise.

(b) The integrated firm always locates at l=0.

Proof:

(a)Non-integration: D chooses location to maximise Π_D = m-d + min[m-d, d-t2]. Clearly locating at l=0, thus maximising d-t2 can never be strictly profitable. Indeed if m-d > d-$\underline{t}$ > d-$\bar{t}$ or d-$\underline{t}$ > m-d > d-$\bar{t}$ then D locates at l=1/2 in maximising min[m-d, d-t2]. However if m-d ≤ d-$\bar{t}$ < d-$\underline{t}$ then min[m-d, d-t2] = m-d, and location is irrelevant.

(b) Integration: D's location is now chosen to maximise $\Pi_{D\text{-}U1}$ = m-t1-E. Clearly this entails minimising t1 i.e. choosing l=0 since t(0) = 0. QED.

In the non-integrated case D chooses its location to maximise downstream profits only. When product is sourced from both upstream firms, D's share of the profit on each supply contract is determined by the incremental outside option value of single sourcing - m-d. Since this value is independent of transport costs, location choice is irrelevant in this case. Where D opts for single sourcing (- always from U1 given our assumptions) its share of profits is m-t2 i.e. the value of the U2 sourcing alternative. Clearly downstream profit maximisation then involves

---

[26] We therefore concur with Bork when he states that "the real cost of any transfer from the manufacturing unit to the retailing unit includes the return that could have been made on a sale to an outsider". In our model the retailer will source internally only when this generates greater value than external supply.

minimising the costs of U2 supply. D2 will therefore choose $l=1/2$.[27] Though $l=1$ is not permitted in our model, symmetry and the adverse consequences of lock-in to U2 would render this option unattractive to D in any event.

In the integrated case, the effects of D's location choice on U1's profits are internalised. D will locate at 0. Lock-in to U1 is no longer a concern, simply involving a transfer between divisions of the integrated enterprise. Internal transport costs are then minimised. Note that since additional transport costs are not incurred on incremental product volume, the location decision is never driven by attempts to raise D's share the value of trade with non-integrated U2 (but see Section 5(i)).

Where the non-integrated D locates at 1/2, integration always brings about a change in location choice. However, in the case where the non-integrated D is indifferent about location no such prediction can be made.

Proposition 3.2 [Integration]

(a) U1 and D will integrate if this changes D's location choice and $E < \underline{t}$.

(b) If $E \geq \underline{t}$, D and U1 will never integrate.

Proof:

If integration does not result in a change in location, then the sole effect is to reduce the combined D-U1 profit by the merger cost, E.

If integration changes D's location choice (from $l=1/2$ to $l=0$) then D-U1's gains are given by:

$\Pi_{U\text{-}D1}(l=0) - \Pi_{U1}(l=1/2) - \Pi_{D1}(l=1/2) = m\text{-}E - m\text{-}\underline{t} = \underline{t} - E$

Then D and U1 have a strict incentive to integrate iff $E < \underline{t}$. QED.

---

[27] Note that, in our formulation, the non-integrated D is concerned only with its outside option. It considers the efficiency consequences of its actions for the second best supply value, not the value of realised supply. This is, of course, an extreme formulation. In general bargaining between the buyer and supplier will result in a sharing of the efficiency gains. Some of the externality effect observed in our model would then be internalised.

Integration is attractive only if it will lead to a change in D's location decision, from l=1/2 to l=0. The cost of internal supply is then reduced by $\underline{t}$. For integration to be worthwhile this cost saving must exceed the merger cost, E. Clearly if no change in location, and therefore no transport cost gain, is forthcoming, then integration will never proceed.

In making the integration decision, U1 and D consider their own future joint profits only. Integration may therefore impose a negative externality on the non-integrated upstream firm U2. There are two mechanisms by which this can occur.

First, consider the case where U2 always supplies one unit, irrespective of D's location. D's share of the gains from such trade, driven by its outside option, is then always m-d. The upstream firm collects the residual profit, 2d-m-t2, and therefore bears the full increase in transport costs ($\bar{t}$ - $\underline{t}$) if D changes its location decision from l = 1/2 to l = 0.

The transport costs induced by the integrated D's location decision can, in addition, affect supply patterns. In particular, D may utilise U2 as a source of supply when located at 1/2, but rely exclusively on U1 when l=0 is chosen. In this case, if integration induces a change in D's location decision, it will result in foreclosure of U2's sales.

From a first best perspective, integration is never efficient since it involves a merger cost E. However, recognising that integration may be necessary to induce efficient location choice, merger may improve *equilibrium* welfare. Clearly, when the non-integrated retailer makes an efficient location choice, integration can never be welfare-improving.

Proposition 3.3 [Welfare and Integration]
(i) If 2$\underline{t}$-E < 2d-m < $\bar{t}$ then any vertical integration that occurs will reduce welfare.
(ii) If 2$\underline{t}$-E < 2d-m *and* the retailer locates efficiently when indifferent, then any integration that takes place will reduce overall surplus

Proof:

(i) The integrated retailer always locates at 0. From Proposition 1, integration only takes place if $E < \underline{t}$ and the non-integrated retailer locates at 1/2.

$2\underline{t} - E < 2d\text{-}m \Rightarrow 2\underline{t} - (2d\text{-}m) < E < \underline{t} \Rightarrow m\text{-}d < d\text{-}\underline{t}$.

Consequently the non-integrated firm double sources at 1/2.

Surplus from integration, and subsequent location at 0 = $\max[m, 2d\text{-}\bar{t}] - E$.

Surplus from non-integration and location at 1/2 = $2(d\text{-}\underline{t})$.

Now $2d\text{-}m < \bar{t}$ implies $m > 2d\text{-}\bar{t}$ therefore integration inefficient if:

$$2(d\text{-}\underline{t}) > m - E$$

i.e. if $$2d\text{-}m > 2\underline{t} - E.$$

(ii) Where double sourcing occurs irrespective of location then D is indifferent about location. If it then locates efficiently, integration (at cost E) must reduce surplus.

From (i) we know that, if $2\underline{t} - E < 2d\text{-}m$ then double sourcing is optimal at 1/2 and a switch to single sourcing at 0 is welfare reducing. QED.

When single sourcing is optimal at 0, the non-integrated retailer will locate at 1/2. Where this takes place, integration, in shifting location from 1/2 to 0, can only be inefficient if it affects U2's profits (i.e. it has externality effects). This will only be the case if double sourcing occurs at 1/2. A switch from double sourcing at 1/2 to single sourcing at 0 generates a clear transport cost saving ($2\underline{t}$). However, a merger cost E is incurred, and the benefits of multi-product retailing (2d-m) are lost. Merger is inefficient if the costs outweigh the benefits.

Suppose that single sourcing is everywhere optimal. A non-integrated retailer will then always locate at 1/2, though location at 0 is efficient. Nevertheless, vertical integration that brings about a change in downstream location decision would not always increase welfare. To be specific, if $E > \underline{t}$ merger costs exceed transport cost savings. Of course, where single sourcing is optimal irrespective of location, inefficient integration never takes place, since all welfare effects are internalised by D and U1.

If double sourcing is optimal at every location the non-integrated retailer will be indifferent about location. The efficient location minimises overall transport costs. If $2\underline{t} > \bar{t}$ then location at 0 is efficient, while location at 1/2 is optimal if $\bar{t} > 2\underline{t}$. Of course, if the indifferent retailer always chooses the efficient location then integration can generate no benefit, and involves an added cost E.

Our analysis has suggested that integration may well cause a switch from double to single sourcing. The market for U2's product is then effectively foreclosed. We will now address the question of whether such foreclosure is inefficient.

Proposition 3.4 [Foreclosure]

Conditional on integration:

(a) Foreclosure occurs if $\underline{t} < 2d\text{-}m < \bar{t}$.

(b) This foreclosure is inefficient if $2\underline{t} < 2d\text{-}m$.

Proof:

(a) Under non-integration D is supplied by both U's if $m\text{-}d < d\text{-}\underline{t}$.

When integrated D locates at l=0 and is supplied by U1 only if $m\text{-}d > d\text{-}\bar{t}$.

Combining these conditions proves (a).

(b) Foreclosure yields overall profits m. Since the integrated firm locates at l=0 there are no transport costs for D1. Therefore from the efficiency results above, this is inefficient if $m < 2(d\text{-}\underline{t})$. Rearranging gives (b). QED.

Integration changes D's location decision from l=1/2 to l=0. Foreclosure results if supply patterns are sensitive to location choice. This will be the case if double sourcing occurs at l=1/2 i.e. $m\text{-}d < d\text{-}\underline{t}$, while single sourcing is efficient for l=0 i.e. $m\text{-}d > d\text{-}\bar{t}$. The benefits of multi-product retailing must therefore be sensitive to transport costs.

For a given location choice, the downstream firm's sourcing decision is always efficient. In considering the efficiency of foreclosure we therefore focus on the following question: Given that a change in location choice from l=1/2 to l=0 will result in foreclosure of U2, does such a move maximise overall profits? Single sourcing at 0 involves no transport costs, while double sourcing at 1/2 incurs transport costs of $2\underline{t}$. For foreclosure to be efficient, these cost savings must exceed the lost value of multi-product retailing - 2d-m.

Comparing Propositions 3.2 and 3.3 it is worth noting that the efficiency of integration does not hinge on foreclosure effects. Suppose double sourcing is optimal at all locations and the non-integrated retailer locates at 1/2. Integration then brings about a change in location, but no foreclosure. Depending on the structure of transport costs, this may or may not bring about efficient location. Note also that foreclosure-inducing integration can be efficient.

3.4 Variations and Extensions to the Basic Model

i) *Unit transport costs*

An obvious variation on our basic model is to impose a transport cost *per batch* supplied by upstream firms. This contrasts with the basic setup, where a single transport cost is incurred *per product line*.

D's profits in this case can be obtained by simple adaption of the basic model results. Where double sourcing occurs, D's outside option is the single sourcing alternative. Of course this option now involves an additional cost, since a transport cost is incurred on both batches. D is thus able to extract m-d-t2 from supplier U1 and m-d-t1 from supplier U2 (cf. basic model results). Downstream profits are therefore given by 2(m-d)-(t1+t2). Where D opts for single sourcing (from U1) the outside option of sourcing from U2 now has value m-2t2.

Now, when D finds double sourcing attractive at all locations, it is in general no longer indifferent about that location. To see this, note that D's profit is given by 2(m-d)-(t1+t2). The retailer therefore aims to minimise t1+t2. Where $\bar{t} < 2\underline{t}$ it is clear that D will prefer to choose l=0. Though such a location increases D's lock-in to U1 (by reducing the value of the U2 single sourcing option by $\bar{t}$ - $\underline{t}$), this is offset by an increase of $\underline{t}$ in D's outside option when dealing with U2.

Conversely, when $\bar{t} > 2\underline{t}$, D prefers to locate at 1/2. Only when $\bar{t} = 2\underline{t}$ will D now be indifferent to location choice.

Where a location choice of l=0 results from single sourcing, the non-integrated D will no longer always locate at l=1/2. In particular, this may not be so when single sourcing is optimal at l=0, but double sourcing is optimal at l=1/2. In the basic model, when such circumstances prevail, location at l=1/2 dominates. The key to this is to note that where single sourcing prevails a mid-point location is clearly optimal. However, when double sourcing occurs the value of the single sourcing option is independent of location. On introducing a per batch transport cost, the single sourcing option for the second batch is now transport cost dependent. Location at l=1/2 is no longer a dominant action for D.

When a per batch transport cost is incurred, this additional cost is always passed on to the downstream firm. Profits for the U-D1 combination are therefore given by:

$$\Pi_{D\text{-}U1} = m\text{-}2t1$$

Where integration affects the location decision, it will therefore be optimal in a wider range of circumstances i.e. integration will now occur when a change in location is induced and $E < 2\underline{t}$.

Note that this profit expression holds irrespective of which supplier D trades with. Where integrated D trades with U2 there is an additional incentive to locate at l=0. The reduction in the transport costs of internal supply (by U1) is of value not only when internal sourcing actually occurs, but also when supplies are sourced externally from U2. In such cases, by raising the value of D's (internal) outside option, locating at l=0 increases D's share of external trade profits.[28] Note that this effect is absent in our basic model above since there the value of a second unit of supply from a given supplier is not affected by additional transport costs, and hence by location choice.

[28] This effect is the driving force behind some of the results in Bolton and Whinston (1990). See, for example, their Proposition 4.1.

ii) *Upstream monopoly*

A simple re-interpretation of the basic model allows us to consider a scenario where a monopoly upstream firm supplies competing downstream firms. The results derived above hold for this case too. However it is then vital that the single supplier is capacity constrained. A downstream firm, by securing all upstream output, can then be sure that it monopolises the downstream market. Of course, to achieve this a downstream firm must outbid its rival for all supply capacity. Thus a downstream firm may (inefficiently) purchase all upstream capacity to secure downstream monopoly, even though a single unit is sufficient for its production needs.

In the absence of such capacity constraints (and given the impossibility of exclusive dealing contracts) the upstream supplier cannot guarantee either downstream firm a monopoly position. Indeed the supplier would always have an incentive, having supplied one retailer on monopoly terms, to proceed to supply the other firm. Foreseeing this outcome, competition for supplies disappears, and with it so does supplier profit.

This effect is precisely that highlighted by Hart and Tirole (1990) in their first model (Ex Post Monopolization). A non-integrated upstream firm, with unlimited capacity can never commit to supplying a single downstream firm. Monopoly profits are fully dissipated. Hart and Tirole show that, in this context, integration will eliminate upstream incentives to oversupply the downstream market. Where downstream monopoly is attractive, the integrated upstream supplier has every incentive to confer that monopoly power on its own downstream subsidiary.

In our model, by contrast, it is the downstream firm that is the monopolist. Since it is the residual claimant it will not purchase from both suppliers if single sourcing is optimal. The Hart-Tirole motive for integration is therefore absent in our basic model. Their effect would re-appear, however, if the retailer sold services to upstream manufacturers, but final sale revenues were earned directly by the upstream firms.

3.5 Multiple Retailers

An important development of the basic model involves the introduction of an additional retailer. Only then can the possibility that both vertically integrated and non-integrated firms coexist in equilibrium be considered. We will show that such an industry structure can emerge, even though all downstream and all upstream firms are initially identical.

Secondly, we will be able to consider the endogenous relationship between downstream competition and vertical integration. A key concern in the anti-trust literature is whether vertical merger, that leaves horizontal concentration at both upstream and downstream levels unchanged, can adversely effect competition. In the context of a stylised model of such competition, we will show that downstream competition can indeed be weakened by integration. We will also explore the potential existence of chains of integration.

The two upstream firms, U1 and U2, are again located at the ends of the unit line. However, now two downstream firms, D1 and D2, can each locate at either end of the line or mid-way between the endpoints. We will permit vertical integration between U1 and D1, and between U2 and D2. U2 and D2's decision to integrate follows that by U1 and D1. Again a cost E is incurred in undertaking a merger. In evaluating the merits of integration, upstream-downstream pairs aim to maximise their combined profits.

Where a downstream firm is indifferent as to its supply source, we will assume that Di purchases from Ui. To simplify matters considerably we will also assume that there is no value to either downstream firm in multi-product supply i.e. $d=0$. In the absence of competition, each downstream firm will therefore extract surplus m from consumers.

Our stylised model of competition will take the following form. We will suppose that if downstream firms locate strictly closer than Δ apart, then competition reduces each firm's revenues by θm.[29] Allowing Δ to vary will enable us to analyse the impact of increasing downstream competition on integration decisions.

$\Delta = 0$

In this case there is no competition between downstream firms, irrespective of their location decisions. In such circumstances non-integrated downstream firms always locate at 1/2 (see Figure 3.1(i)). In doing so, they maximise their outside options and hence minimise lock-in to either upstream supplier. An integrated downstream firm will locate next to its upstream partner, at the appropriate end of the line (Figure 3.1(iii)). Integration removes lock-in worries and firms will then seek to eliminate inefficient transport costs. In the absence of competition, joint retailer-supplier profits are independent of industry-wide integration and location decisions. Partial Integration will therefore never be observed.

Profits for U-D pairs ($\Delta = 0$):

Non-Integration	$\Pi_{U\text{-}D} = m - \underline{t}$
Full Integration	$\Pi_{U\text{-}D} = m - E$

Clearly integration is attractive if and only if the transport cost savings outweigh the fixed cost of integration i.e. $E < \underline{t}$.

---

[29] Note, in particular, that this competition effect is assumed independent of the levels of retailers' transport costs. This is natural in our setting, since transport costs are fixed not (per unit) variable costs. Our analysis of the competition effects of integration will therefore focus on location effects. In general, integration would also lead to changes in variable costs that will impact on competition. The costly merger process could then be seen (partly) as an investment in variable cost reduction. Numerous authors (e.g. Bonanno and Vickers (1988), Fershtman and Judd (1987)) have shown that the strategic effects of integration on rivals' behaviour may outweigh the direct cost benefits, rendering vertical integration unattractive. Vertical separation may then be the optimal response to tougher competition.

$0 < \Delta \leq 1/2$

With $\Delta > 0$ the possibility of downstream competition emerges. In this case, if downstream firms locate together then competition reduces their revenues by θm. On the other hand, if downstream firms locate at least a 1/2 unit apart, competition is avoided.

In the non-integrated case there are now two opposing forces at work. A desire to avoid lock-in to either upstream supplier drives both downstream firms towards location at 1/2. However, countering this, the prospect of competition encourages those downstream firms to locate apart. Below, we derive the condition for the lock-in effect to dominate.

Lemma 3.3 [Non-Integrated Location Choice]

If $\theta m < \bar{t} - \underline{t}$ then non-integrated downstream firms locate at 1/2.

Proof:

Locating at 1/2 is a dominant strategy for each non-integrated downstream firm.

Suppose first that Dj locates at 1.

Di's profits as a function of location are then given by:

$(1-\theta)m - \bar{t}$	at 1
$m - \bar{t}$	at 0
$m - \underline{t}$	at 1/2

Clearly locating at 1/2 maximises downstream profit.

Suppose now that Dj locates at 1/2

Di's profits for each location are then:

$m - \bar{t}$	at 1
$m - \bar{t}$	at 0
$(1-\theta)m - \underline{t}$	at 1/2

Profit is again maximised at 1/2, if $(1-\theta)m - \underline{t} > m - \bar{t}$ i.e. if $\theta m < \bar{t} - \underline{t}$. QED.

Locating at one end of the unit line is bad for independent retailer profit, because it reduces the value of the second best sourcing option. Compared with a mid-point location, the transport cost incurred with this option increases from $\underline{t}$ to $\bar{t}$. Offsetting this is the potential value of creating distance between retailers. We will assume that the lock-in effect always dominates.

Condition 3A: $\theta m < \bar{t} - \underline{t}$.

Note that an integrated downstream firm will always locate next to its upstream subsidiary. Not only does this location decision minimise transport costs but in addition, by locating away from 1/2, the integrated retailer will avoid competition with any non-integrated downstream firm. Given Condition 3A, vertical integration will therefore always lead to a separating of retailers.[30] With competition, the incentives to integrate are increased.

Profits for U-D pairs ($0 < \Delta \leq 1/2$):

Non-Integration	$\Pi_{U\text{-}D} = (1-\theta)m-\underline{t}$
Partial Integration (Dj-Uj integrated)	$\Pi_{Uj\text{-}Dj} = m-E$
	$\Pi_{Ui\text{-}Di} = m-\underline{t}$
Full Integration	$\Pi_{U\text{-}D} = m-E$

When $E > \theta m + \underline{t}$ both upstream-downstream pairs will remain unintegrated, since the costs of integration then outweigh any competition and transport cost benefits. Conversely when $E < \underline{t}$ both firms will integrate. However for intermediate E values, where $\underline{t} < E < \theta m + \underline{t}$, we will observe a partially integrated industry structure (see Figure 3.1(ii)). Note that both firms would prefer that their rival integrated (and located at an end-point) alone. In this way they would benefit from reduced competition without bearing the necessary costs of integration. Since U1 and D1 have a first mover advantage they will remain

[30] Dixit (1983) develops a model where the structure of vertical relationships may result in greater spatial separation of retailers. However, in his model, this increased separation of downstream firms is driven by reduced entry into the retail industry. Furthermore, Dixit focuses on a traditional analysis of the value of contractual vertical restraints (franchise fees, royalties, etc.), in comparison with full integration, to a *monopoly* supplier. Transaction cost aspects of integration are ignored.

unintegrated, forcing U2 and D2 to incur the costs of merger.

From an overall producer perspective, individual incentives to integrate are too weak. As stated above, for $E > \theta m + \underline{t}$ non-integration is the equilibrium industry structure. The costs of integration then exceed the competition and transport cost benefits of integration *for the individual firm*. However, in making their integration decision, a given manufacturer-retailer pair fail to internalise the competition benefits of merger to the rival retailer. Where $2\theta m + \underline{t} > E > \theta m + \underline{t}$, partial integration is optimal from an overall firm viewpoint, but will not be sustained in equilibrium.

The key observation from this section is the emergence of an asymmetric partially integrated industry structure for moderate competition, despite initially identical firms.

$1/2 < \Delta \leq 1$

Competitive pressures in the downstream market are now strong, and profit reducing interaction occurs unless Ds locate at opposite ends of the unit line.

A fear of lock-in still attracts non-integrated downstream firms towards the mid-point of the unit line. Countering this, the competition effect encourages these firms to locate at the end-points. Condition 3A again ensures that the lock-in threat dominates any competition effect encouraging firms to locate at the endpoints. Independent downstream firms will therefore always choose to locate at 1/2.

Integrated downstream firms locate next to their upstream partners. Again, such location avoids inefficient transport costs, and may eliminate downstream competition. The role of this competition effect may be crucial, and depends delicately on integration patterns in the industry as a whole.

To be precise, if firms Uj and Dj do not integrate, no competition benefits will follow from integration of Ui and Di. Even if Di integrates with Ui and locates at

0, competition between downstream firms will not be eliminated, since non-integrated Dj continues to locate at 1/2 (and $\Delta > 1/2$!). However, if Uj and Dj are integrated, and Dj locates at 1, then integration of Ui and Di (in shifting Di's location from l=1/2 to l=0) brings about the elimination of downstream competition. The delicacy of this competition effect prevents a partially integrated industry structure from emerging, as the following profit results make clear.

Profits for U-D pairs ($1/2 < \Delta \leq 1$):

Non-Integration:	$\Pi_{U\text{-}D} = (1-\theta)m-\underline{t}$
Partial Integration (by Dj-Uj):	$\Pi_{Uj\text{-}Dj} = (1-\theta)m-E$
	$\Pi_{Ui\text{-}Di} = (1-\theta)m-\underline{t}$
Full integration:	$\Pi_{U\text{-}D} = m-E$

It should be clear that with tough competition the partial integration structure will never emerge in equilibrium. Integration by one upstream-downstream pair alone cannot reduce competition. For $E < \theta m+\underline{t}$ both U-D pairs will integrate, while for $E > \theta m+\underline{t}$ both firms will remain unintegrated.

Note that for intermediate E, where $\underline{t} < E < \theta m + \underline{t}$, U1 and D1 only integrate in anticipation of merger by U2 and D2. Likewise, U2 and D2 will only integrate in response to D1-U1 merger. Chains of integration are therefore observed.

These competition results are summarised below.

Proposition 3.4 [Competition and Integration]

For $\underline{t} < E < \theta m + \underline{t}$, the competitive environment plays a critical role in determining industry structure:

(i) In the absence of competitive forces ($\Delta = 0$), both downstream firms will be non-integrated.

(ii) Moderate competitive pressure ($0 < \Delta \leq 1/2$) generates partially integrated outcomes i.e. vertically integrated and non-integrated firms co-exist.

(iii) A fully vertically integrated industry emerges when downstream competitive pressures are intense ($1/2 < \Delta < 1$). Furthermore, a chain of integration is observed.

In our analysis of the basic model (with monopoly D) it was assumed that the downstream firm was able to extract all consumer surplus generated by product sales. This allowed us to focus on (upstream and downstream) producer surplus alone in assessing efficiency. However, once competition between downstream firms is permitted, explicit consideration of consumers is illuminating. In particular, it seems reasonable to suppose that at least a fraction of the downstream profit dissipated through competition is passed on to consumers. Of course, it could be argued that this profit is in fact expended on (wasteful and unsuccessful) attempts to regain full market power. However, here we will assume that *all* revenues lost by firms as a result of the competitive process accrue to consumers.

Once competition between downstream firms is initiated, we have seen that for a range of merger costs E, integration takes place solely because this results in a reduction in that competition. When $\underline{t} < E < \theta m + \underline{t}$, integration occurs even though merger costs exceed the transport cost savings that result. In effect, the integrating firms in the industry are undertaking costly merger simply to avoid making transfers to consumers. Clearly, integration in such circumstances will be inefficient.

It should be pointed out that our analysis has been based on the assumption that consumers derive no added value from retailer separation. In general, consumers may benefit from the greater variety offered by integration-induced retailer differentiation. Such considerations would obviously affect our view of the efficiency implications of integration.

Finally, note that to simplify our analysis in this section, we have only considered single product firms. Given our basic assumptions, this rules out foreclosure effects. However, the changes in downstream location patterns induced by integration suggest that foreclosure effects would be restored in a richer model. Furthermore, in a multi-product setting, retailers' product portfolio choices may affect the intensity of inter-firm competition. This may further encourage additional integration-induced foreclosure.

3.6 Conclusions

In this chapter we have modelled a retailer's choices between two rival suppliers and between single-product and multi-product retailing. An investment specificity decision must also be taken - in the form of a discrete location choice. Vertical integration may encourage supplier-specific location, where an independent retailer would choose general investment. Indeed, costly merger will only take place if a change in the retailer's location decision is induced.

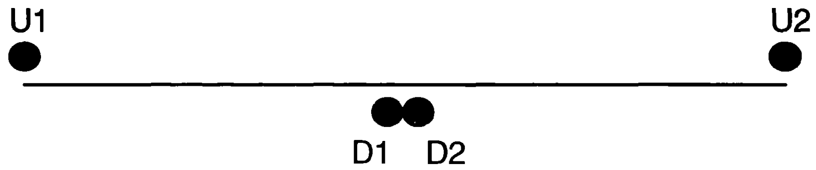
When single product retailing dominates, an independent retailer will always choose a general location, thus minimising lock-in to a given supplier. However, in precisely these circumstances a supplier-specific location is efficient. Vertical merger may then be attractive, eliminating lock-in concerns and encouraging specific investment. Of course, such integration will occur only if the benefits of appropriate investment outweigh the (exogenously given) costs of merger. Whenever bilaterally attractive merger occurs, it will be efficient in this case.

Retailer supply patterns may be sensitive to the location decision. Integration can therefore result in foreclosure of the non-integrated supplier. Whether this foreclosure is efficient or not depends delicately on the inter-relationship between the relative merits of specific versus general investment, and single versus multiproduct retailing. In the context of this simple model we explore the interaction between efficiency and anti-competitive motives for vertical integration.

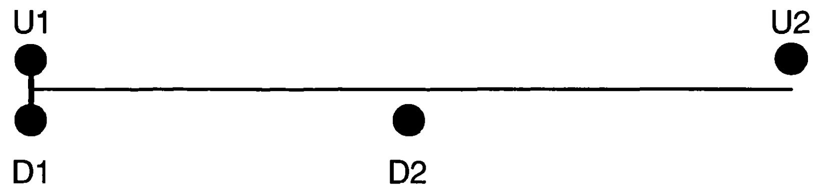
The basic model was extended to consider the impact of vertical integration on competition *between retailers*. Such competition may motivate otherwise unattractive merger. Integration, in encouraging supplier-specific investment, allows retailers to adopt competition-reducing product differentiation strategies. For moderate levels of competition, a partially integrated industry structure then results, as integration by a subset of firms is sufficient to reduce rivalrous interaction. With tougher competition, a coordinated integration process is essential. In this case, a fully vertically integrated industry results. The integration process then displays the characteristics of an integration chain.

Figure 3.1

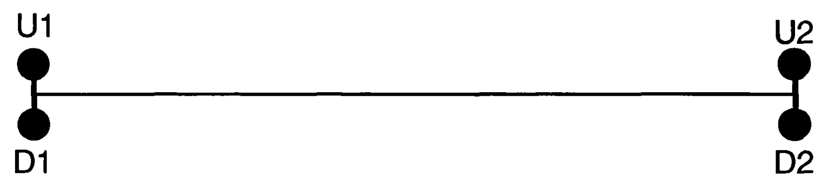
(i) Non-Integration



(ii) Partial Intagration
(U1-D1 integrated)



(iii) Full Integration



Appendix 3

Here we derive the results of the auction process, proving Lemma 3.1.

Non-integration

If unit 1 from U1: value if both units from U1 = m

value if unit 2 from U2 = 2d-t2

winning supply price = $|2d-t2-m|$.

If unit 1 from U2: value of unit 2 from D1 = 2d-t1

value of both units from D2 = m

winning supply price = $|2d-t1-m|$.

If U1 offers unit 1 for p_1 and wins it gains: p_1 + max[0,m-2d+t2]

If U1 does not bid for the unit 1 contract it gains: max[0,2d-m-t1].

Indifference implies p_1^{max} = max[0,m-2d+t1] - max[0,2d-m-t2].

If U2 supplies unit 1 at price p_2 it gains: p_2 + max[0,m-2d+t1]

If U2 loses the unit 1 contract it gains: max[0,2d-m-t2].

Indifference implies p_2^{max} = max[0,m-2d+t2] - max[0,2d-m-t1].

D will accept a bid that yields maximum two period gains. Here this implies D gains Π_D, where:

$$\Pi_D = \min\,[d\text{-}t2\text{-}p_2max + \min(d\text{-}t1,m\text{-}d)\ ,\ d\text{-}t1\text{-}p1^{max} + \min(d\text{-}t2,m\text{-}d)].$$

Let M = [d-t1-$p_1$max + min(d-t2,m-d)] - [d-t2-p_2^{max} + min(d-t1,m-d)].

Substituting in the expressions for p_1 and p_2:

M = (m-t1) - max[0, 2d-m-t1] - (m-t2) + max[0, 2d-m-t2]

If M≥0 U1 wins the unit 1 supply contract.

If M<0 U2 supplies unit 1.

(Rembering that unit 1 from U1 in case of tie).

Now $M = [m-t1 - \max(0,2d-m-t1)] - [m-t2 - \max(0,2d-m-t2)]$

Some simple algebra yields

$$M = [m-d + \min(m-d,d-t1)] - [m-d + \min(m-d,d-t2)]$$

By assumption $t1 \le t2$, thus $M \ge 0$ and U1 always supplies unit 1.

Since $M \ge 0$, $\Pi_D = m-d + \min(m-d,d-t2)$.

Now, if D pays p* for unit 1 from U1 in equilibrium its profit is given by:

$$\Pi_D = d-t1-p^* + \min[m-d, d-t2]$$

However, we know that $\Pi_D = m-d + \min[m-d, d-t2]$

Consequently,

unit 1 price 2d-m-t1, and

unit 2 price = $|2d-m-t2|$.

U1 supplies both units if $d-t2 < m-d$.

One unit is supplied by each U if $d-t2 > m-d$.

This is precisely the efficiency criterion, given location.

The payoffs are:

$$\Pi_D = m-d + \min[m-d,d-t2]$$

$$\Pi_{U1} = \max[2d-m-t1, t2-t1]$$

$$\Pi_{U2} = \max[0,2d-m-t2].$$

Vertical Integration (D-U1)

If we assume the same auction process then again input allocation will be efficient, given location. The profit of U2 remains the same, while the combined profits of U1 and D are reduced by integration costs E.

Payoffs:

$$\Pi U1\text{-}D = m-t1-E$$

$$\Pi_{U2} = \max[0,2d-m-t2].$$

References

Bolton, P. and M. D. Whinston (1990): "Incomplete Contracts, Vertical Integration and Supply Assurance". Harvard University D. P. No. 1493.

Bolton, P. and M. D. Whinston (1991): "The Foreclosure Effects of Vertical Mergers". Journal of Institutional and Theoretical Economics, 147:207-226.

Bolton, P. and M. D. Whinston (1993): "Incomplete Contracts, Vertical Integration, and Supply Constraints". Review of Economic Studies, 60: 121-148.

Bonanno, G. and J. Vickers (1988): "Vertical Separation". Journal of Industrial Economics, 36: 257-265.

Bork, R. H. (1978): "The Antitrust Paradox", Basic Books, New York.

Chandler, A. D. (1959): "The Beginnings of 'Big Business' in American Industry". Business History Review, 33: 1-31.

Chandler, A. D. (1977): "The Visible Hand". Belknap Press, Cambridge.

Dixit, A. (1983): "Vertical Integration in a Monopolistically Competitive Industry". International Journal of Industrial Organization, 1: 63-78.

Fershtman, C. and K. L. Judd (1987): "Equilibrium Incentives in Oligopoly". American Economic Review, 77: 927-940.

Gal-Or, E. (1991): "Duopolistic Vertical Restraints". European Economic Review, 35: 1237-1253.

Grossman, S. and O. D. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration". Journal of Political Economy, 94: 691-719.

Halonen, M. (1991): "Vertical Foreclosure in Cournot Oligopoly". London School of Economics. Mimeo.

Hart, O. D. and J. Tirole (1990): "Vertical Integration and Market Foreclosure". Brookings Papers on Economic Activity (Microeconomics): 205-286.

Ordover, J. A., G. Saloner and S. C. Salop (1990): "Equilibrium Market Foreclosure". American Economic Review, 80: 127-142.

Porter, M. E. (1980): "Competitive Strategy". Free Press, New York.

Reiffen, D. (1992): "Equilibrium Vartical Foreclosure: Comment". American Economic Review, 82: 694-697.

Riordan, M.H. and O. E. Williamson (1985): "Asset Specificity and Economic Organization". International Journal of Industrial Organization 3: 365-378.

Salinger, M. A. (1988): "Vertical Mergers and Market Foreclosure". Quarterly Journal of Economics, 103: 345-356.

Tirole, J. (1988): "The Theory of Indusrial Organization". MIT Press, Cambridge.

Wlliamson, O. E. (1985): "The Economic Institutions of Capitalism". Free Press, New York.

Chapter 4

Incomplete Contracts, Vertical Integration and Product Market Competition [31]

[31] This chapter is based on joint work with Maija Halonen

4.1 Introduction

In recent years there has been renewed interest in the potential adverse effects of vertical merger. In particular attention has focused on the possibility that integration will result in the foreclosure of nonintegrated rivals i.e the restriction of buyers' access to suppliers or suppliers' access to buyers. Clearly an adequate response to such questions requires explicit consideration of the effects integration has on firm organisation, and of the relationship between firm organisation and the competitive environment within which that firm operates.

In this paper we will adopt an incomplete contracts, optimal control rights approach to integration. A number of recent papers, e.g. Grossman-Hart (1986) and Hart-Moore (1990), have emphasised the vital role played by the allocation of ownership rights in motivating investment by self-interested agents. The link between ownership structure and investment-driven value creation (cost reduction), is made explicit. A recent, related paper by Bolton and Whinston (1993) has highlighted the role integration may play in the reduction of rival product value and in foreclosure effects. However these papers do not address product market competition in any depth.

In contrast, a more traditional literature has explored the effect integration has on product market competition and on the possibility of foreclosure, e.g. Bonanno-Vickers (1988), Salinger (1988), Hart-Tirole (1990), Ordover-Saloner-Salop (1990). These papers relate changes in market cost structure (and hence firm performance) to integration decisions. Strategic motives for integration are emphasised. However the precise effects of integration often appear arbitrary and unclear. For instance, merger is often assumed to lead to profit sharing and the removal of all conflicts of interest within the integrated firm.

In this paper we examine the implications vertical integration decisions have for the supply relations and cost structure within an industry. A given allocation of ownership rights effectively constitutes a specialisation decision i.e. a group of assets can be assembled to concentrate on selling the final product, to specialise in input production or to combine both tasks. The effects of integration are explicitly related to the investment incentives of self-interested agents. Further, we will examine the effect of varying the degree of product market competition on industry structure. The paper is therefore an attempt to relate overall industry structure and competitive environment to firms' internal organisation decisions.

Following Grossman and Hart (1986) we will assume that ownership confers residual control rights. If an agent owns an asset then he/she makes the decisions regarding the use of that asset, except where such decision powers are granted to others through contractual agreements. In a complex world, where it may be difficult to verify or even write the clauses of a complete contract, the allocation of ownership rights will therefore be important.

In particular, it may be impossible to write workable contracts offering agents rewards conditional on complex asset enhancing investments being undertaken. In such circumstances, the only guarantee an agent has of receiving a share in the benefits created by such investment is through ownership of the relevant assets, and hence direct control over returns. Similarly, a buyer may be unable to negotiate enforceable contracts for the purchase of input supplies. For instance, it may be impossible or prohibitively expensive for a court to decide whether prescribed quality standards have been met in the event of a dispute.A buyer's incentives to undertake product development will therefore be diluted if the supplier can command a share of the benefits, via its control of input supplies. Much, of course, will depend on the degree of competition in the input supply market and the options available to the buyer. It may however prove optimal, if product development is crucial, to allocate ownership of input production facilities to the buyer.

From the above we see that a fragmented ownership structure will lead to some investment, and hence reasonable performance, at each stage in the production chain. In contrast, concentrating ownership in the hands of a particular agent will guarantee strong investment by that agent, but will also create a class of non-owning workers with weak effort incentives. The optimal allocation of ownership rights will depend on the effectiveness of investment at various stages of production, and the value of specialisation.

Obviously, the nature of the competitive environment will play a key role in determining firm structure. On the one hand, the optimal response to vigorous competition at a particular stage in production may involve a focus at that stage, encouraging significant investment and strong performance. On the other hand, it may be preferable to concentrate activity (through an appropriate allocation of asset ownership rights) where competition is weak, thereby generating improved profit opportunities. These opposing forces are central to the ownership allocation decision in this paper.

The following example illustrates the main effects at work in the model. Consider a two stage production process. At the first (upstream) stage, a simple widget is produced. This widget is a key component in the second stage (downstream) production of a sophisticated gadget. Associated with widget production is an engineer who can improve the component's value by overhauling manufacturing equipment. Similarly, a designer working on the gadget can increase its value considerably.

Suppose that initially ownership of both upstream and downstream assets is allocated to the designer. Given his control rights, the designer can develop gadgets without the need to negotiate with outside parties. He will therefore have every incentive to devote time and effort to producing a quality product, since all benefits are retained. Clearly, if the designer could contract with the engineer for maintenance services then he would do so, since improved widgets would add value to gadget manufacture, generating potential gains for both parties. However, if enforceable agreements cannot be written, once the engineer has overhauled the widget machine she can be sacked, without compensation. Foreseeing this eventuality, the engineer will not exert any effort.

Of course, an obvious solution to the maintenance problem is to give the engineer the ownership rights to the widget machine. She would then have control of the production and sale of widgets (and therefore a share in the benefits of greater effort), creating improved incentives to overhaul the machine. However, if ownership rights are given to the engineer they must be taken away from the designer. He now has to negotiate with the engineer to acquire the widgets essential for gadget production. As a result, the benefits of better design (i.e. higher value gadgets) must be shared, since without engineer-controlled widget supplies, better designs are valueless. Consequently, the designer's incentives to innovate are dampened. The relative sizes of these effects determines which ownership structure dominates. If maintenance is largely cosmetic then we might expect the designer to retain all ownership rights. On the other hand, if effective maintenance is essential, ownership of the widget machine will be given to the engineer.

Now suppose that another integrated, gadget manufacturer starts producing in a different market. Though it can produce adequate widget substitutes in-house, it is keen to purchase the genuine, engineer-enhanced variety. The in-house facility will, however, provide useful bargaining power, since the integrated gadget manufacturer does not rely entirely on external widget purchases. Clearly the new widget sales will increase the value of a machine maintenance programme, and hence the benefits of engineer ownership and control of widget production. The dulled incentives of the designer may be offset by the added value of external sales.

But what if the new firm starts competing in the home market? Selling widgets to the rival allows it to focus on downstream production. In contrast, the home designer's incentives are dulled and he will suffer in a competitive market. If competition is weak then extra component sales will outweigh the adverse effects on the gadget market. However as competition gets tougher it may pay to concentrate on gadget design, by concentrating ownership of widget and gadget manufacturing assets with the designer. Of course, widget production will be affected, but the rival suffers from this too. In addition concentrating ownership gives the designer improved incentives to work hard on the downstream market. Finally, if competition on the downstream market gets really tough, then it offers little attraction for two gadget producers, since the fruitless fight for advantage will largely exhaust profits. It may then prove optimal for one of the firms to effectively withdraw from the final market and re-deploy its assets to concentrate on widget production.

The above example provides a flavour of our results. We are particularly interested in the relationship between industry ownership structure and two key parameters - the relative value of effort at the upstream and downstream production stages, and the degree of product market competition. Broadly we see greater upstream ownership of assets (nonintegration) the more valuable is the upstream worker's investment. Allocating ownership rights to ineffective upstream managers dilutes the effort incentives of downstream workers with little offsetting gain. If, on the other hand, upstream effort is reasonably effective then encouraging activity at both production stages is optimal.

We also observe a non-monotonic relationship between ownership structure and the toughness of competition, for a range of parameter values. With little or no competition, the value of improved input sales to outside buyers may outweigh the losses due to reduced downstream managerial effort, encouraging fragmented ownership. As competition increases however, the resulting weakness of the downstream production arm (combined with strengthened rival performance) encourages a more integrated (downstream controlled) ownership structure. When competition in the final good market is extremely strong, there is little reward for identical duopolists. In such a setting the optimal (fragmented) ownership structure effectively encourages one firm to specialise in input production while the other dominates the final good market.

4.2 The Model

We consider an industry with two upstream assets, U1 and U2, and two downstream assets, D1 and D2. Input is produced upstream for use at the downstream, final production stage. The basic upstream and downstream production processes utilise costless, constant returns to scale technology.

Each manufacturing unit is managed by a worker. Pre-production effort by the worker can increase the value of the unit's output. This effort can be thought of as improving the productive efficiency of physical assets. The manager of Di chooses the value added in final good production - $J_i \in [0, J^{max}]$. Similarly, the effort of upstream Um's manager, $K_m \in [0, K^{max}]$, raises the value added at the input stage by K_m/ϕ. Therefore ϕ is a measure of the relative effectiveness of upstream effort. Suppose downstream firm i obtains input from upstream supplier m. The final product value is then the sum of the values added at each stage in production: $v_i = J_i + K_m/\phi$.

The cost of effort level X to a manager is c(X), where $c' > 0$ and $c'' > 0$.

Assumption 4.1: $\phi \geq 1$.

This assumption ensures that a given level of effort is more effective in increasing value when applied during final good production, rather than at the input manufacturing stage. We do not claim that this is in any way a general property. Rather, we limit our attention to $\phi \geq 1$ for simplicity - focusing on the attractions of backward integration.

Once they have expended effort in value-enhancing investment, the original workers can be replaced costlessly for the remaining stages of production. A qualified replacement manager can be hired from a competitive pool to supervise post-investment operations. Once effort has been applied, access to the firm's assets rather than continued employment of the original worker secures the benefits of improved performance. The key here is to note that managerial effort is directed at enhancing the quality of physical assets, rather than at developing human capital

Adopting an incomplete contracts framework, we assume that effort requirements are too complex to be described effectively in an ex ante contract. Consequently, effort levels are chosen non-cooperatively. It is for this reason that inefficiencies can arise and that the allocation of ownership rights matters in our model.

Contractual incompleteness means that any effort compensation a manager receives must be offered *after* that effort has been undertaken. We also assume that it is very difficult to adequately describe the required input characteristics, ruling out anything but spot contracts for input trade. Effective profit sharing agreements cannot be enforced either.

Ex ante contracts can however be written on the allocation of ownership rights. Among the rights ownership confers is the power to hire and fire managers. The owner of an asset also has the rights to any residual profits generated by the asset. Ownership of assets Ui and Di is allocated between the initial managers of the upstream and downstream units, ui and di respectively. These managers allocate ownership rights to maximize joint ui-di profits, taking as given the ownership of assets Uj and Dj.

As a considerable simplification we will only consider two candidate industry configurations:

(i) *Partial Integration (PI)*

One set of assets (U2-D2) is owned by a downstream manager while the second set of assets (U1-D1) are owned separately, and

(ii) *Integration (I)*

Both sets of assets are owned by downstream managers.

Industry structure is therefore determined by U1 and D1's integration decision only. The restriction on the set of possible ownership structures partly reflects our emphasis on downstream effort incentives.

Ex post, uncertainty about the required characteristics of input is resolved and the owners of upstream and downstream assets negotiate spot contracts for the procurement of input. The contract will take the form of a two part tariff. Input will be exchanged at price zero, while the fixed fee negotiated will depend on the bargaining power of the parties. Bargaining over non-owning managers' effort compensation takes place ex post, after effort has been undertaken.

Assumption 4.2: Internally sourced input can be utilised for internal production only.

This assumption is made for convenience. In a more complete model, that would endogenise firms' technology specificity choice, this assumption could be derived from first principles.

Ultimately, production of final good occurs and the downstream firms compete in the final market. There, an overall consumer population of 1 is assumed. Consumers' demands are driven by downstream product values, and the degree of market competition. The profits of downstream firm Di are given by:

$$\pi_{Di} = \pi_i\ (v_i,\ v_j;\ \rho)$$

where v_i and v_j are final product values, and ρ is a measure of the degree of competition.

When $\rho = 0$ i.e. there is no competition, each downstream firm is a monopoly in its half of the market. Profit for Di then depends on the value of its own product only:

$$\pi_{Di} = \pi_i\ (v_i,\ v_j;\ 0) = v_i / 2$$

A number of additional assuptions on the effects of competition are also made:

Assumption 4.3: For $\rho > 0$:

$$\frac{\partial^2 \pi_i}{\partial v_i^2} < \frac{\partial^2 \pi_i}{\partial v_i\, \partial v_j} < 0$$

Assumption 4.4: $\rho \leq \rho^{max}$, where $\pi_i(J^{max}, J^{max}; \rho^{max}) - c(J^{max}) = 0$

This assumption ensures that, by making the maximum investment, integrated downstream firms can always guarantee non-negative profits whatever the competitive environment. A duopolistic industry is therefore guaranteed when both D1 and D2 are integrated.

Assumption 4.5: For $v_i \leq v_j$:

$$\frac{\partial \pi_i}{\partial \rho} < 0 \quad \text{and} \quad \frac{\partial^2 \pi_j}{\partial v_j \, \partial \rho} > 0$$

Increased competition always reduces the profit of the downstream firm with the lower value product. Furthermore, the investment incentives of the (weakly) superiof firm are increasing in the tougness of the competitive environment.

To produce the final product, downstream firms require supplies of input. An integrated firm can source input either internally, or from any independent upstream firm. Assumption 4.2 restricts non-integrated downstream firms to seeking input from independent upstream firms. Control of this essential input supply gives upstream owners some negotiating power in dealing with downstream firms. We will assume a specific bargaining structure.

If Di can be supplied with input by a non-integrated upstream firm Uk only, an even split of the gains from trade between the two parties will result:

d_i's payoff: $\dfrac{\pi_i(v_i, v_j; \rho)}{2}$

u_k's payoff: $\dfrac{\pi_i(v_i, v_j; \rho)}{2}$

where di and uk are the owner-managers of Di and Uk respectively.

If Di is supplied with input internally, control of the supply unit guarantees the downstream owner-manager 100% of downstream profits:

d_i's payoff: $\pi_i(v_i, v_j; \rho)$

Any investment by a non-owning upstream manager is, of course, sunk when input trade takes place.

Finally, when a downstream firm Di can obtain input either internally or from an independent supplier Uk, but in equilibrium is supplied by the latter, then the allocation of payoffs is given by:

di's payoff: $\dfrac{\pi_i(v_i,v_j;\rho) + \pi_i(\bar{v}_i,v_j;\rho)}{2}$

uk's payoff: $\dfrac{\pi_i(v_i,v_j;\rho) - \pi_i(\bar{v}_i,v_j;\rho)}{2}$

where $\bar{v}_i$ is the value of Di's product if the internal supply option is realised.

The gains from external over internal input supply are then shared by the downstream firm and the independent upstream supplier. The idea here is that, while negotiations with the independent supplier are incomplete, the integrated downstream firm can always make use of its own (inferior) input production facility. Clearly, whenever the internal option is chosen, Di retains all profit.

Note that all actions are observable to managers. Upstream and downstream managers' effort decisions are made simultaneously, after asset ownership has been determined. Bargaining over input procurement and effort compensation occurs under symmetric information: effort levels are observable, but not verifiable. Finally, downstream managers know the extent of their rival's input purchases when setting final good prices

4.4 Manager Payoffs and Effort Incentives

i) Preliminaries

Before addressing the issue of managerial investment incentives in detail, it will first be useful to consider the position of non-owning upstream workers. From our bargaining assumptions, it should be clear that only owner-managers will receive a return on effort investment. Consequently, no effort will be undertaken by non-owning workers.

Lemma 4.1

Non-owning managers receive no compensation for effort. Consequently no effort will be undertaken by them.

Since contracts contingent on effort cannot be written, managers can only be offered credible reward for effort *after* it has been undertaken. However, once effort has been undertaken, its value is sunk in the firm's assets, and the asset owner can reap full returns without further managerial involvement. Post-investment, a manager can therefore be replaced at no cost. As a result, managers will not be rewarded for their effort, unless they are also asset owners. Foreseeing no reward, a non-owning manager will therefore make no investment in effort.

Since asset U2 is always downstream controlled, Lemma 4.1 implies $K_2 = 0$. To ease notation we will therefore let $K_1 = K$, from now on.

It follows naturally from Lemma 4.1 and our bargaining structure that if $K > 0$, the independent upstream supplier can always generate a return selling input to the integrated downstream firm. $K > 0$ ensures that the independent upstream firm's product will always be preferred to internal supply.

Lemma 4.2

In the non-integrated case, upstream firm U1 supplies both downstream firms with input, provided managerial effort $K > 0$.

Since there are no capacity constraints on output, and exclusive dealing contracts are not enforceable, an independent upstream firm cannot commit to serve one downstream firm only. Having dealt with one downstream firm on an exclusive basis, it then has every incentive to trade with the rival, since input is traded in spot exchanges before final good sales are realised. Foreseeing this, each downstream firm will negotiate the terms of its supply contract accordingly.

Assumption 4.2 ensures that under integration, both firms will source input internally.

ii) Payoff Functions

We are now in a position to calculate the payoffs for owner-managers, under the two possible ownership regimes. [32]

Partial Integration (PI):

With a partially integrated industry structure there are three independent firms - the non-integrated upstream and downstream firms U1 and D1, owned by their respective managers, and the integrated downstream firm D2, owned by downstream manager d2. The payoffs for these owners are given by:

$$\pi_{u1} = \frac{\pi_1(v_1,v_2;\rho)}{2} + \frac{[\pi_2(v_1,v_2;\rho) - \pi_2(v_1,\bar{v}_2;\rho)]}{2} - c(K)$$

$$\pi_{d1} = \frac{\pi_1(v_1,v_2;\rho)}{2} - c(J_1)$$

$$\pi_{d2} = \frac{[\pi_2(v_1,v_2;\rho) + \pi_2(v_1,\bar{v}_2;\rho)]}{2} - c(J_2)$$

where $v_1 = J_1 + \frac{K}{\phi}$, $v_2 = J_2 + \frac{K}{\phi}$ and $\bar{v}_2 = J_2$

Under the partial integration regime the independent upstream firm U1 receives half of D1's profits. Since D1 has no alternative source of input, bargaining results in this 50:50 split. U1 also receives half of *its contribution* to the profits of the integrated downstream firm, D2. If D2 fails to come to an agreement with U1, it can source input from its own (inefficient) upstream plant. U1 and D2 therefore bargain over the incremental contribution of the superior input. As a result D2 secures the full value of its internal supply option plus 50% of the incremental value of U1 input supply.

[32] As explained, since non-owning managers do not invest, they are largely irrelevant to our analysis once ownership rights are allocated.

Since contracts contingent on effort cannot be written, bargaining for input takes place *after* effort has been undertaken. Each party therefore bears the full cost of its own investment.

Integration (I):

In the integrated setting there are only two firms. Each downstream manager also controls an upstream plant. Payoffs for the owners are given by:

$$\pi_{di} = \pi_i(v_i, v_j; \rho) - c(J_i) \qquad i = 1,2$$

where $v_i = J_i$.

With no independent upstream producers, each downstream firm must source its input needs internally.[33] Given that non-owning upstream managers exert no effort, these inputs are of basic, unenhanced quality. However, the downstream owner-managers have costless access to input and as a result retain all final market revenues.

For the remainder of the analysis we will drop the arguments of the profit functions and let $\pi_i \equiv \pi_i(v_i, v_j; \rho)$ and $\bar{\pi}_i \equiv \pi_i(\bar{v}_i, v_j; \rho)$.

iii) Incentives to invest

Competition

Each party's incentives to invest in effort are determined by their individual payoff functions. Where managers receive the full revenue increments generated by their effort, incentives are maximised. If the returns to investment are shared, incentives are diluted too. Effort levels are chosen such that, at the margin, the incremental cost of extra effort, borne solely by the investing manager, is equal to the incremental return to that manager. Below we list these relationships, for each ownership structure.

---

[33] Assumption 4.2 rules out the possibility that both downstream firms utilise a common in-house production facility.

Partial Integration (PI):

$$u1:\quad \frac{1}{2\phi}\left[\frac{\partial \pi_1}{\partial v_1} + \frac{\partial \pi_1}{\partial v_2} + \frac{\partial \pi_2}{\partial v_1} + \frac{\partial \pi_2}{\partial v_2} - \frac{\partial \bar{\pi}_2}{\partial v_1}\right] = c'(K)$$

An increase in effort by u1 raises the value of the input used by D1 and D2, and hence the value of the final products. Since u1 must negotiate with each buyer, its contribution to profits will be shared 50:50 with the downstream firm. Greater effort by the upstream manager will enhance the value of D1's product, in turn tending to increase its revenues. However, improved input also benefits D1's rival on the downstream market, D2, creating a counteracting negative force on D1's profits. Clearly, the scale of this negative effect will depend on the degree of competition in the downstream market. If competition is weak then increases in the value of D2's product have very little effect on D1's profits, and vice versa. Conversely, if competition is strong the negative effects will be quite considerable. Similar forces are at work in the impact of increased upstream investment on D2's profits. However, were D2 to utilise its internal input source, greater u1 effort would have a purely negative effect - via tougher competition from an enhanced D1 product. Greater effort by u1 therefore weakens D2's internal sourcing option, thereby increasing u1's share of actual trade profit.

$$d1:\quad \frac{1}{2}\frac{\partial \pi_1}{\partial v_1} = c'(J_1)$$

When d1 exerts greater effort, it increases its profits. Given the necessary bargaining for input, d1 receives half of the resulting profit increment, which it equates with the marginal cost of extra effort.

$$d2:\quad \frac{1}{2}\left[\frac{\partial \pi_2}{\partial v_2} + \frac{\partial \bar{\pi}_2}{\partial v_2}\right] = c'(J_2)$$

By investing in more effort, d2 not only increases overall revenues, but also increases its share of these revenues (- by raising product value with internal supply, and hence increasing "stand-alone" profits). The overall and internal option profit effects each have a half weighting in determining d2's effort.

Let the efforts of u1, d1 and d2 in the partial integration case be denoted by K^{PI}, J_1^{PI} and J_2^{PI} respectively.

Integration (I):

$$\text{di}: \quad \frac{\partial \pi_i}{\partial v_i} = c'(J_i) \qquad i = 1,2$$

Under the integrated regime each downstream firm must source input internally. However, this ensures that each downstream manager keeps the full incremental value of extra effort.

We will denote d1 and d2's investments in the integrated case by J_1^I and J_2^I respectively.

Nash equilibrium in investments is determined by these first order conditions, under partial integration and integration. We will assume that all equilibria are locally stable. Our assumptions ensure that a unique equilibrium in investments exists for both structures.

From these effort equations, some of the trade-offs driving the choice of ownership structure are clear. Under the integrated regime, downstream managers keep all of the gains from extra effort, but lose out on the benefits of upstream investment. With a partially integrated industry structure, firms benefit from upstream investment in input value, but the incentives for effort by the downstream managers are diluted. Of course, the precise nature of these trade offs will depend on the relative values of upstream investment, and on the influence of competition on the downstream market. These factors will be considered further below. First however, we will examine investment incentives in the no competition case (i.e. $\rho = 0$) in more detail.

No Competition

When there is no competition on the product market, the profits of each downstream firm are unaffected by the value of the rival's output. Market share for each firm is set at 1/2 and profits depend only on the own product value. As a result, the effort equations have a particularly simple form.

Partial Integration (PI):

$$\text{u1:}\quad \frac{1}{2\phi} = c'(K^{PI})$$

$$\text{d1:}\quad \frac{1}{4} = c'(J_1^{PI})$$

$$\text{d2:}\quad \frac{1}{2} = c'(J_2^{PI})$$

U1 serves the whole market (=1) and the benefits ($1/\phi$) of each added unit of effort are shared 50:50 with the relevant downstream firm. D1 monopolises a market of 1/2 and receives half the increment to unit value its effort generates, the other half going to the input supplier. D2 also has a market of 1/2, but keeps the full value increment of its marginal effort, since it can realise all of this with internally sourced input. Each manager optimises by choosing effort such that their individual marginal return is equated to the marginal cost.

In the absence of strategic effects ($\rho = 0$), it is clear that D2's internal sourcing option ensures its manager undertakes greater effort than D1's owner-manager. The value, and therefore the level, of upstream effort is driven by the parameter ϕ - the measure of the relative effectiveness of upstream investment. When ϕ is large an investment in upstream effort has little value. Consequently there is little return on such investments, and effort by a upstream owner-manager will be low. When ϕ is small upstream effort is highly productive and large returns ensure high effort levels.

Integration (I):

$$\text{d1:} \quad \frac{1}{2} = c'(J_1^I)$$

$$\text{d2:} \quad \frac{1}{2} = c'(J_2^I)$$

With an integrated structure, the downstream managers (with market share of 1/2) each receive the full benefit from increased effort, thus maximising their incentives to invest in such effort. Symmetry ensures that effort levels for both downstream managers are identical. Note that, for $\rho = 0$, d2's investment is identical under both industry ownership configurations.

Lemma 4.3

With no competition between downstream firms ($\rho = 0$) investment by the integrated firm owner (d2) is identical under integrated and partially integrated industry structures. This effort level is identical to that undertaken by both active managers in the integrated regime, but more than that of the owner-manager of a non-integrated downstream firm i.e.

$$J_2^{PI} = J_2^I = J_1^I > J_1^{PI}$$

We are now in a position to consider the equilibrium industry ownership structure.

4.5 Industry Ownership Structure

In the context of our simple model, the equilibrium industry ownership structure is determined by the integration decision of managers u1 and d1. If assets U1 and D1 are jointly controlled the industry will consist solely of integrated firms. While if U1 and D1 are separately controlled the industry will be partially integrated, since by assumption assets U2 and D2 are always jointly owned. Ownership of assets U1 and D1 is allocated to maximise the combined profits of managers u1 and d1.

Profits, for each ownership regime, are given below:

$$\pi^{I}_{u1-d1} = \pi^{I}_{1} - c(J^{I}_{1})$$

$$\pi^{PI}_{u1-d1} = \pi^{PI}_{1} + \frac{[\pi^{PI}_{2} - \bar{\pi}^{PI}_{2}]}{2} - c(J^{PI}_{1}) - c(K^{PI})$$

We are particularly interested in the dependence of industry structure on two key parameters - the effectiveness of upstream effort (ϕ) and the degree of product market competition (ρ). The locus of points separating the Integrated and Partially Integrated regions of our parameter space is therefore the set of (ϕ, ρ) combinations where u1-d1 profits under the two ownership structures are equal.

No Competition

We will begin by considering the case where $\rho = 0$ i.e. when there is no competition between the downstream firms. The industry ownership structure observed in equilibrium then depends on the value of the parameter ϕ.

Proposition 4.1

Suppose that $\rho = 0$ i.e. there is no competition between the downstream firms. Then there exists a $\bar{\phi}$ such that for $\phi > \bar{\phi}$ the industry will be integrated, while for $\phi < \bar{\phi}$ the industry will be partially integrated.

Proof:

The difference between u1 and d1's joint profits under the integrated and partially integrated regimes is given by:

$$\pi^{I}_{u1-d1} - \pi^{PI}_{u1-d1} = [\frac{J^{I}_{1}}{2} - c(J^{I}_{1})] - [\frac{J^{PI}_{1}}{2} - c(J^{PI}_{1}) + \frac{3K^{PI}}{4\phi} - c(K^{PI})]$$

The individual, profit maximising, effort incentive equations of the previous section imply:

$$\frac{J^{I}_{1}}{2} - c(J^{I}_{1}) > \frac{J^{PI}_{1}}{2} - c(J^{PI}_{1})$$

If $\phi = 1$:

$$\frac{K^{PI}}{2} - c(K^{PI}) = \frac{J_2^{PI}}{2} - c(J_2^{PI}) = \frac{J_1^{I}}{2} - c(J_1^{I})$$

and as $\phi \rightarrow \infty$:

$$\frac{3K^{PI}}{4\phi} - c(K^{PI}) \rightarrow 0$$

Hence for $\phi = 1$: $\pi^{I}_{u1\text{-}d1} - \pi^{P\ I}_{u\ 1\text{-}d1} < 0$, and for large enough ϕ: $\pi^{I}_{u1\text{-}d1} - \pi^{P\ I}_{u\ 1\text{-}d1} > 0$.

By continuity, there is a $\phi = \bar{\phi}$ such that: $\pi^{I}_{u1\text{-}d1} - \pi^{P\ I}_{u\ 1\text{-}d1} = 0$.
QED.

In considering the possible integration of assets U1 and D1, managers u1 and d1 face a simple trade-off. Concentrating ownership of both assets in the hands of the downstream manager ensures the best possible downstream effort incentives. Since full control of input supply is assured, the returns on effort investment are not shared with the upstream manager. Of course, the upstream manager - denied control of the upstream asset - will not exert any effort in this case. Furthermore, with integration, no input is sold to D2.

Effort will be undertaken by the upstream manager if he owns the input producing asset. Such input investment not only contributes to the value of D1's product, but in addition generates input sales to D2. Of course, upstream effort incentives are dampened by the need to share the benefits with downstream buyers. Similarly, the returns to effort by downstream manager d1 must now be shared with u1, who controls the supply of input. A dilution of downstream effort results from the separation of upstream and downstream asset ownership.

For low values of ϕ the value of upstream investment is high. The benefits of some upstream effort then outweigh the effects of dilution on d1's investment incentives. Consequently, a non-integrated ownership structure is attractive. However, when ϕ is high, encouraging upstream manager effort translates ineffectively into added input value. Separating ownership of upstream and downstream assets then buys a small upstream value contribution, at a relatively high cost in terms of diluted downstream incentives. For high values of the parameter ϕ, ownership is best concentrated in the hands of the downstream manager.

When u1 and d1 make their integration decision, they consider their own combined profits only. Yet that integration decision has important implications for D2, even when there is no competition between downstream firms. Integration effectively eliminates a valuable source of input. For D1, this is traded off against improved downstream managerial incentives. However, from D2's perspective, integration has an unquestionably negative impact. Only half of the extra revenues generated for D2 by an independent U1's superior input accrue to U1. In making their integration decision, u1 and d1 fail to take account of the other 50% of these benefits, captured by D2. Integration thus imposes an externality on D2, through the foreclosure of its most attractive input source.

To maximise overall producer surplus, integration should only occur when the downstream incentive benefits for d1 outweigh the adverse effects of reduced upstream investment and lower value input for *both* downstream firms. Note that since our demand structure allows producers to capture all the surplus in the no competition setting, an undesirable ownership structure from a producer viewpoint is also inefficient in overall welfare terms. Proposition 4.2 formally summarises this.

Proposition 4.2
When there is no competition between downstream firms ($\rho = 0$), there exists a range of values of ϕ for which the equilibrium industry ownership structure does not maximise producer (and hence total) surplus. Within this parameter range, integration by U1 and D1 results in inefficient foreclosure of D2's optimal input source.

Proof:

When $\rho = 0$, both downstream firms are monopolists. Consumer surplus is then zero, given our assumptions. Hence overall welfare is maximised when producer surplus is maximised.

By definition, at $\rho = 0$ and $\phi = \tilde{\phi}$: $\pi^{I}_{u1\text{-}d1} = \pi^{PI}_{u1\text{-}d1}$.

From Lemma 4.3: $J_2^{PI} = J_2^{I}$

Therefore:

$$\pi^{PI}_{d2} = \frac{J_2^{PI}}{2} + \frac{K^{PI}}{4\tilde{\phi}} - c(J_2^{PI}) > \frac{J_2^{I}}{2} - c(J_2^{I}) = \pi^{I}_{d2}$$

Hence at $\phi = \tilde{\phi}$: $\pi^{PI}_{d2} + \pi^{PI}_{u1\text{-}d1} > \pi^{I}_{d2} + \pi^{I}_{u1\text{-}d1}$.

Define $\phi = \hat{\phi}$ such that when $\rho = 0$ and $\phi = \hat{\phi}$: $\pi^{PI}_{d2} + \pi^{PI}_{u1\text{-}d1} = \pi^{I}_{d2} + \pi^{I}_{u1\text{-}d1}$.

By continuity, $\hat{\phi} > \tilde{\phi}$.

QED.

It may be worthwhile at this point to compare these results with those of Bolton and Whinston (1993). In their model, the analogue of the above scenario, where there is no competitive interaction at all between downstream firms, is an environment with no supply constraints. In the absence of competitive interaction, the multilateral context effectively reduces to a pair of bilateral relationships, between the input supplier and each of the two downstream firms. Integration of the sole supplier with one of the downstream firms in the Bolton and Whinston model is then always privately attractive, and also improves social welfare. First best investment is ensured from the integrated downstream manager and, since there are no externality consequences, this has no adverse impact on the non-integrated downstream firm's investment.

In contrast, our model, generates externality effects to integration even where there are no competitive pressures. Vertical Integration of U1 with D1 eliminates D2's most efficient input supplier. However, since u1 captures only half these supply benefits, it does not take full account of the overall effect on D2 profit in making its integration decision.

Furthermore, in the Bolton and Whinston model, even where supply is constrained, integration still generates first best if outside options do not bind. In this case, though both downstream firms are chasing a single unit of input, there is still no effective competition between them. The value of internal supply to the integrated input supplier does not then affect the terms of its negotiations with the non-integrated downstream firm, nor does it affect the supply decision. It should be noted, however, that for outside options *never* to bind, the Bolton-Whinston framework requires extreme assumptions on the correlation of downstream firms' values for input across states of the world.

Competition

As we increase the degree of competition between the downstream firms this general pattern will continue with the integrated structure being preferred by u1 and d1 when ϕ is high and the non-integrated structure dominating when ϕ is low. However the value of ϕ where the agents are indifferent between integration and non-integration will in general vary as the degree of competition is increased. It is to this relationship that we now turn.

We have already determined that when $\rho = 0$, u1 and d1 are indifferent between integrated and non-integrated structures when $\phi = \bar{\phi}$. Which structure dominates, however, when we increase ρ i.e. we introduce product market competition?

It will be useful to begin by considering the relative attractions of integrated and non-integrated structures for managers u1 and d1, when competition is tough.

Lemma 4.4

There exists a ρ^* such that for $\rho > \rho^*$ partial integration is preferable to non-integration for all values of ϕ.

Proof: See Appendix.

Tough competition results in considerable dissipation of profits under the symmetric structure, with both downstream firms owning and controlling input production facilities. Downstream managers undertake considerable investment in effort. However, since these investments are exactly equal, market shares remain at 1/2. Therefore, at high ρ almost all profits are dissipated under the symmetric industry structure. In contrast, a non-integrated U1 can always make positive profit as a supplier. Though separation of upstream and downstream assets may effectively eliminate D1 on the final good market, this is offset by the (admittedly small) supplier profits accruing to U1. For strong levels of competition, a non-integrated structure for U1 and D1 is therefore preferred. Rather than structuring ownership to encourage vigorous head to head competition in the downstream market (i.e. through integration), assets are more advantageously organised to extract maximum benefit from the U1-D2 *buyer-supplier* relationship.

We will now consider the other extreme, where competition is weak. It will first prove useful to compare downstream product values for $\rho = 0$ and $\phi = \tilde{\phi}$.

Lemma 4.5

At $\rho = 0$ and $\phi = \tilde{\phi}$, the following relationships between product values hold:

$$v_2^{PI} > \bar{v}_2^{PI} = v_1^{I} = v_2^{I} > v_1^{PI}$$

Proof:

The equalities follow trivially from Lemma 4.3.

At $\rho = 0$ and $\phi = \tilde{\phi}$, $\pi^{I}_{u1-d1} = \pi^{PI}_{u1-d1}$ implies that:

$$\frac{J_1^I}{2} - c(J_1^I) = \frac{J_1^{PI}}{2} - c(J_1^{PI}) + \frac{3K^{PI}}{4\tilde{\phi}} - c(K^{PI})$$

But from Lemma 4.3:

$$\frac{J_2^I}{2} - c(J_2^I) = \frac{J_1^{PI}}{2} - c(J_1^{PI}) + \frac{3K^{PI}}{4\tilde{\phi}} - c(K^{PI})$$

Re-arranging:

$$J_2^{PI} - J_1^{PI} - \frac{K^{PI}}{\tilde{\phi}} = 4\left[[\frac{J_1^{PI}}{4} - c(J_1^{PI})] - [\frac{J_2^{PI}}{4} - c(J_2^{PI})] \right]$$

$$+ 4\,[\frac{K^{PI}}{2\tilde{\phi}} - c(K^{PI})]$$

Now from the profit maximising effort conditions:

$$\frac{J_1^{PI}}{4} - c(J_1^{PI}) > \frac{J_2^{PI}}{4} - c(J_2^{PI})$$

and

$$\frac{K^{PI}}{2\tilde{\phi}} - c(K^{PI}) > 0$$

Therefore

$$J_2^{PI} > J_1^{PI} + \frac{K^{PI}}{\tilde{\phi}}$$

QED.

Consider the set of parameter values such that there is no competition ($\rho = 0$) and u1 and d1 are indifferent between integration and separation ($\phi = \tilde{\phi}$) i.e. we are at the point in parameter space where the critical boundary between Integrated and Partially Integrated regimes cuts the ϕ axis (see Figure 4.1). In the Integrated case downstream firms' products are identical, while under the Partially Integrated regime the value of D2's product (even when using internally sourced input) exceeds that of D1's product.

Therefore, when U1 and D1 are non-integrated we would expect D2 to increasingly dominate the market as competition gets tougher. On the other hand, when U1 and D1 are integrated the values of the two downstream products are identical and hence both firms will maintain a market of 1/2 (though increasing competition will reduce profits).

When $\rho = 0$, $\pi^{I}_{u1\text{-}d1} = \pi^{PI}_{u1\text{-}d1}$ at $\phi = \bar{\phi}$ (by definition). It is then simple to show that when competition increases, the change in relative profits for the two structures is given by:

$$\frac{\partial\,[\pi^{I}_{u1\text{-}d1} - \pi^{PI}_{u1\text{-}d1}]}{\partial\rho} = [\frac{\partial\pi^{I}_{1}}{\partial\rho} - \frac{\partial\pi^{PI}_{1}}{\partial\rho}] - \frac{1}{2}[\frac{\partial\pi^{PI}_{2}}{\partial\rho} - \frac{\partial\bar{\pi}^{PI}_{2}}{\partial\rho}] - \frac{1}{4}[\frac{\partial J^{PI}_{1}}{\partial\rho} + \frac{\partial K^{PI}}{\partial\rho}]$$

The first two bracketed terms give the relative direct effects of a competition increase on u1 and d1's combined profits in the integrated and non-integrated regimes. The first term is the differential direct effect on D1's profit. Lemma 5 has shown that, at $\rho = 0$, the value of D1's product is lower and that of the rival D2 is higher under partial integration. We would therefore expect this effect to be positive, enhancing the attractiveness of the integrated structure.

The second bracketed term denotes the effect of competition on U1's share of D2's profit in the partially integrated case, where U1 supplies input to D2. If, as competitive pressures grow, U1's higher value input becomes increasingly important to D2's profit, then U1's share of that profit will also be enlarged. Of course, for every £1 added to D2 profit, bargaining limits U1's share to £0.5. This input sales effect may at least partially offset the direct competition effect on D1's profit. For every sale D2 captures from D1 as competition gets tougher, U1 will share some of the benefit.

The final term in the expression above captures the *indirect* investment effects of competition on non-integrated u1-d1 profits. As our investment incentive equations have shown, in the no competition case a fragmented ownership structure induces underinvestment by both u1 and d1, when seen from a combined u1-d1 profit perspective. To the extent that competition encourages an escalation in effort investment, this will ameliorate the basic underinvestment effect.

In general it will not possible to sign the overall relative effect of an increase in the toughness of competition on integrated and non-integrated u1-d1 profits. Such an increase in competition can be viewed as reducing the prices sustainable by *both* downstream firms, for given structure (see e.g. Sutton (1991), p.9).

However, the underlying forces inducing tougher price competition need not imply reduced overall industry profit. For instance, an increase in the substitutability of the rival products will in general lead not only to an increase in competitive interaction, but also to increased sales for the higher value product. If the latter expansion effect dominates the competition effect, overall profits will rise. Provided the non-integrated supplier U1 captures a sufficient share of D2's profit increase, overall u1-d1 profits may increase, though competition has become tougher. In general, both competition and expansion effects will coexist - see Shaked and Sutton (1990). This limitation accepted, we will attempt to illustrate the implications when the competition effect is dominant.

Lemma 4.6

Suppose that, for $\rho = 0$:

$$\left.\frac{\partial \pi_i}{\partial \rho}\right|_{\rho=0} = f\,(v_j)$$

then

$$\left.\frac{\partial\,[\pi^{I}_{u1\text{-}d1} - \pi^{PI}_{u1\text{-}d1}]}{\partial \rho}\right|_{\rho=0} > 0$$

This lemma takes the initial impact of competition as depending on the *rival's* product value alone. The idea here is that very weak competition will result in few consumers switching their demand to the rival's product. Instead, competitive pressure only impacts on the *prices* that the dominant firm in each half of the overall market can charge. This is a scenario where the pure competition effect outlined above dominates.

When U1 is independently controlled (non-integration), a valuable input source is created for D2. At $\phi = \tilde{\phi}$, Lemma 4.5 has shown that this results in a higher value D2 product and a lower value D1 product relative to the integrated case, where both downstream products have identical values. Given our assumptions, it is the former effect that matters when $\rho = 0$. The competitive impact of a higher value D2 product results in greater profit reduction for the non-integrated D1, when compared with that for the integrated case. Furthermore, U1's ability to extract a profit share from D2, as its input supplier, is unaffected by marginal increases in ρ from 0, since changes in v_2 do not alter the effects of competition on D2 profit.

It will be useful to compare the effect of integration in our model with that in Hart and Tirole (1990). There, integration results in a change to the supply *price* faced by the non-integrated rival. Effectively, integration is undertaken to secure a commitment to higher supply prices for this downstream firm. In contrast, integration in our model eliminates supply of the rival entirely. Our explicit modelling of the relationship between ownership and investment incentives has shown that backward integration will reduce investment at the upstream stage, while increasing downstream investment. For the integrating parties a trade-off exists. However, integration is clearly disadvantageous for the non-integrated downstream firm. As competition begins, integration is therefore an increasingly attractive option for u1 and d1.

Taken together, the results of Lemma 4.4 (tough competition) and Lemma 4.6 (weak competition) imply a non-monotonic relationship between the key parameters ϕ and ρ, along the critical boundary between Integrated and Partially Integrated regions of our parameter space. This is summarised in the proposition below.

Proposition 4.3

If the conditions of Lemma 4.6 hold, then the relationship between equilibrium ownership structure and competition (ρ) may be non-monotonic. If $\bar{\phi}(\rho)$ denotes the locus of points in our parameter space where u1 and d1 are indifferent between full integration and non-integration, then $\bar{\phi}'(\rho)$ is negative for low values of ρ (weak competition), but positive for high levels of ρ (strong competition).

Proof: This follows directly from the results of Lemma 4.4 and Lemma 4.6.

The result is illustrated in Figure 4.1. The critical level of ϕ that separates the Integrated and Partially Integrated regions of our parameter space initially falls as ρ increases from 0 (since an integrated structure is increasingly attractive for U1 and D1). However, as competitive pressures rise further, a fragmented structure becomes more and more attractive. Eventually the Partially Integrated industry structure will dominate, and the critical value of ϕ, $\bar{\phi}$, will rise.

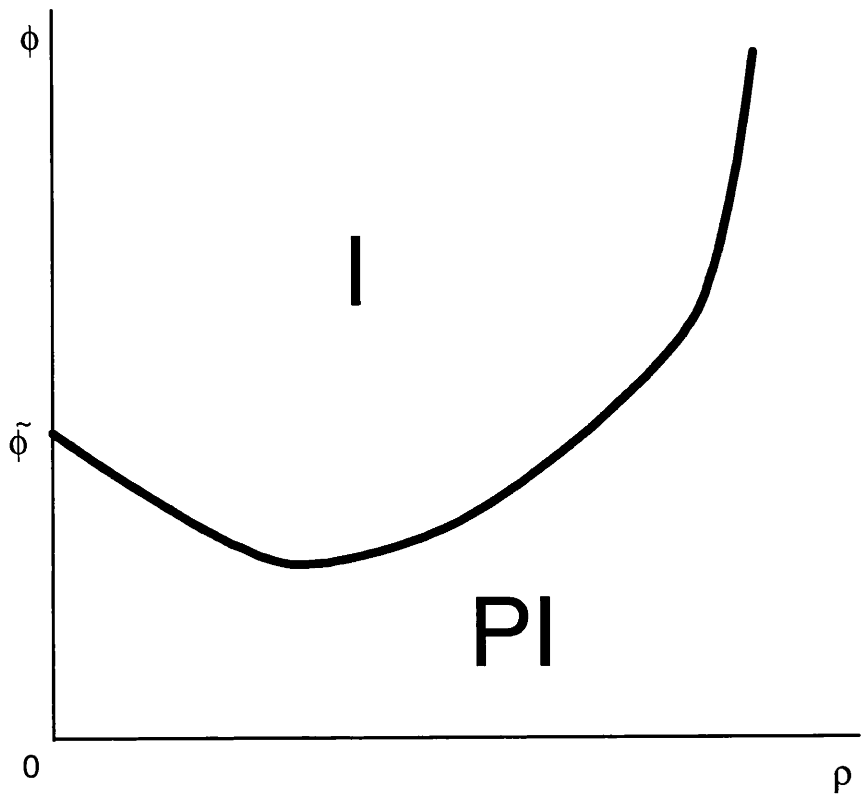
4.6 Conclusions

In this paper we have analysed the extent of vertical integration between upstream (input) and downstream (final good) producers in an industry. We have adopted an incomplete contract, residual control rights approach to integration, emphasizing the role ownership rights play in inducing non-contractible investment by worker-managers. In particular, we have examined the role of two key parameters: the relative value of upstream (versus downstream) investment, and the degree of final product market competition.

We find that when upstream investment is relatively important then control of the upstream asset should be given to its manager - yielding a fragmented industry structure. In contrast when the value of upstream investment is relatively low (or the importance of downstream investment particularly great) control of both upstream and downstream assets should be concentrated in downstream hands, yielding good downstream investment incentives.

The choice of ownership structure is therefore also a specialisation decision. Downstream control of all assets induces specialisation in final good production, while a fragmented (non-integrated) structure encourages some investment at both final and intermediate production stages. We find that for a range of parameter values the specialisation decision depends delicately on the degree of final market competition. As competition increases it may become increasingly important to concentrate on the downstream market (via an integrated structure) to "meet the competition". However, for very tough competition, head-to-head confrontation leads to large profit dissipation. It is then optimal for one group of assets be deployed in a non-integrated configuration, emphasising the supplier relationship and allowing the rival (integrated) firm to effectively dominate the downstream market. We therefore observe a non-monotonic relationship between industry structure and the degree of competition.

Figure 4.1



Appendix 4

Proof of Lemma 4.4:

(i) ρ^{max} is defined such that $\pi(J^{max}, J^{max}; \rho^{max}) - c(J^{max}) = 0$.

Our assumptions imply:

$$\frac{\partial J_i^I}{\partial \rho} = \frac{(\pi_{ij} - \pi_{jj})\pi_{i\rho}}{(\pi_{11}\pi_{22} - \pi_{12}\pi_{21})} > 0$$

Therefore:

$$\frac{d\pi^I_{u1\text{-}d1}}{d\rho} = \frac{\partial \pi_1}{\partial \rho} + \frac{\partial \pi_1}{\partial v_2}\frac{\partial J_2^I}{\partial \rho} < 0$$

where $\pi_{i\rho} = \frac{\partial^2 \pi_i}{\partial v_i \, \partial \rho}$, $\pi_{ii} = \frac{\partial^2 \pi_i}{\partial v_i^2}$ *and* $\pi_{ij} = \frac{\partial^2 \pi_i}{\partial v_i \, \partial v_j}$

Consequently, J^{max} will be selected at ρ^{max} by each firm.

It therefore follows that for any $\varepsilon > 0$, there exists ρ^* such that $\pi^I_{u1\text{-}d1} < \varepsilon$.

(ii) Now $\pi_2 (v_1, v_2; \rho^*) > 0$ at $\phi = \bar{\phi}$.

u1 can therefore make a very small investment at infinitessimal cost (remembering $c'(0) = 0$), which ensures:

$$[\pi_2(v_1, v_2; \rho^*) - \pi_2(v_1, \bar{v}_2; \rho^*)]/2 > 0$$

Hence $\pi^{PI}_{d1} > 0$, which immediately implies $\pi^{PI}_{u1\text{-}d1} > 0$.

QED.

Proof of Lemma 4.6:

i) Considering the direct effect on D1's profits:

$$[\frac{\partial \pi_1^{I}}{\partial \rho} - \frac{\partial \pi_1^{PI}}{\partial \rho}] = f(v_2^{I}) - f(v_2^{PI}) > 0$$

since $f'(\cdot) < 0$ and $v_2^{I} < v_2^{PI}$.

(ii) The competition effect on U1's share of D2's profit is given by:

$$\frac{1}{2}[\frac{\partial \pi_2^{PI}}{\partial \rho} - \frac{\partial \bar{\pi}_2^{PI}}{\partial \rho}] = f(v_1^{PI}) - f(v_1^{PI}) = 0$$

(iii) From the effort incentive equations:

$$\frac{\partial J_1}{\partial \rho} = \frac{\left[\frac{\partial^2 \pi_1}{\partial \rho\, \partial v_1}\right]}{2\, c''(J_1)}$$

$$\frac{\partial K}{\partial \rho} = \frac{\left[\frac{\partial^2 \pi_1}{\partial \rho\, \partial v_1} + \frac{\partial^2 \pi_1}{\partial \rho\, \partial v_2} + \frac{\partial^2 \pi_2}{\partial \rho\, \partial v_2} + \frac{\partial^2 \pi_2}{\partial \rho\, \partial v_1} - \frac{\partial^2 \bar{\pi}_2}{\partial \rho\, \partial v_1}\right]}{2\, \phi\, c''(K)}$$

Now:

$$\left.\frac{\partial^2 \pi_i}{\partial \rho\, \partial v_i}\right|_{\rho=0} = 0 \quad \text{and} \quad \left.\frac{\partial^2 \pi_2}{\partial \rho\, \partial v_1}\right|_{\rho=0} = f'(v_1) < 0$$

Consequently:

$$\frac{\partial K}{\partial \rho} < 0 = \frac{\partial J_1}{\partial \rho}$$

QED.

References

Bolton, P. and M. Whinston (1991): "The 'Foreclosure' Effects of Vertical Mergers". Journal of Institutional and Theoretical Economics, 147: 207-226.

Bolton, P. and M. Whinston (1993): "Incomplete Contracts, Vertical Integration, and Supply Constraints". Review of Economic Studies, 60: 121-148.

Bonanno, G. and J. Vickers (1988): "Vertical Separation". Journal of Industrial Economics, 36: 257-265.

Gal-Or, E. (1992): "Vertical Integration in Duopoly". Journal of Law, Economics, and Organization, 8: 377-393.

Grossman, S. and O. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration". Journal of Political Economy, 94: 691-719.

Hart, O. and J. Moore (1990): "Property Rights and the Nature of the Firm". Journal of Political Economy, 98: 1119-1158.

Hart, O. and J. Tirole (1990): "Vertical Integration and Market Foreclosure". Brookings Papers on Economic Activity, Microeconomics, 205-276.

Helper, S. and D. Levine (1992): "Long-term Supplier Relations and Product-Market Structure". Journal of Law, Economics, and Organization, 8: 561-581.

Holmstrom, B. and J. Tirole (1991): "Transfer Pricing and Organizational Form". Journal of Law, Economics, and Organization, 7: 201-228.

Ordover, J., G. Saloner and S. Salop (1990): "Equilibrium Vertical Foreclosure".

American Economic Review, 80: 127-142.

Salinger, M. (1988): "Vertical Mergers and Market Foreclosure". Quarterly Journal of Economics, 103: 345-356.

Shaked, A. and J. Sutton (1990): "Multiproduct Firms and Market Structure". RAND Journal of Economics, 21: 45-62.

Sutton, J. (1991): "Sunk Costs and Market Structure". MIT Press, Cambridge.

Vickers, J. (1985): "Delegation and the Theory of the Firm". Economic Journal (Supplement), 95: 138-147.

References

Bolton, P. and M. Whinston (1991): "The 'Foreclosure' Effects of Vertical Mergers". Journal of Institutional and Theoretical Economics, 147: 207-226.

Bolton, P. and M. Whinston (1993): "Incomplete Contracts, Vertical Integration, and Supply Constraints". Review of Economic Studies, 60: 121-148.

Bonanno, G. and J. Vickers (1988): "Vertical Separation". Journal of Industrial Economics, 36: 257-265.

Gal-Or, E. (1992): "Vertical Integration in Duopoly". Journal of Law, Economics, and Organization, 8: 377-393.

Grossman, S. and O. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration". Journal of Political Economy, 94: 691-719.

Hart, O. and J. Moore (1990): "Property Rights and the Nature of the Firm". Journal of Political Economy, 98: 1119-1158.

Hart, O. and J. Tirole (1990): "Vertical Integration and Market Foreclosure". Brookings Papers on Economic Activity, Microeconomics, 205-276.

Helper, S. and D. Levine (1992): "Long-term Supplier Relations and Product-Market Structure". Journal of Law, Economics, and Organization, 8: 561-581.

Holmstrom, B. and J. Tirole (1991): "Transfer Pricing and Organizational Form". Journal of Law, Economics, and Organization, 7: 201-228.

Ordover, J., G. Saloner and S. Salop (1990): "Equilibrium Vertical Foreclosure". American Economic Review, 80: 127-142.

Salinger, M. (1988): "Vertical Mergers and Market Foreclosure". Quarterly Journal of Economics, 103: 345-356.

Shaked, A. and J. Sutton (1990): "Multiproduct Firms and Market Structure". RAND Journal of Economics, 21: 45-62.

Sutton, J. (1991): "Sunk Costs and Market Structure". MIT Press, Cambridge.

Vickers, J. (1985): "Delegation and the Theory of the Firm". Economic Journal (Supplement), 95: 138-147.