802

# THE ROLE OF CONTRACTS IN INDUSTRIAL ORGANISATION THEORY

by

Patrick Bolton

London School of Economics

A dissertation submitted to the University of London for the degree of Doctor of Philosophy

June 1986

## ABSTRACT

The thesis comprises three chapters and an introduction. Chapter 1 extends the basic Principal-Agent model by allowing the principal to investigate the agent after the latter has chosen his action. The threat of investigation can be used by the principal as an incentive-scheme. As is well known, this scheme is most effective when the punishment imposed on an agent who is found shirking is as large as possible. It is shown, however, that there will be limits to the optimal size of penalties if the principal makes inspection-errors and if he cannot precommit himself to a given inspection-strategy. Furthermore, if one of these two assumptions is not verified then the principle of maximum deterrence may still apply.

Chapter 2 addresses the question of whether an incumbent seller who faces a threat of entry into his market does prevent entry by signing long-term contracts with his customers. The related question of the optimal length of contracts between the incumbent and his clients is also considered.

It is shown that such contracts do prevent entry to some extent but that they never completely preclude it. Furthermore, it is established that such contracts are socially inefficient. Finally, when the seller possesses superior information about the likelihood of entry, it is shown that optimal contracts may be of finite length, since the length of the contract may act as a signal of the likelihood of entry.

Chapter 3 deals with vertical restraints in manufacturer-retailer contractual relations. The case of a manufacturer who sells a

homogeneous good to retailers who compete in prices and "post-sales" services, is considered. It is shown that simple forms of vertical restraints, such as resale-price-maintenance and franchise-fees, dominate the optimal linear-price contract but are dominated by vertical integration. The analysis is concluded with the description of an optimal contract.

## ACKNOWLEDGEMENTS

My greatest debt is to Oliver Hart. As teacher and supervisor he has had the deepest influence on my approach to economics and on the research leading to this dissertation. His invaluable advice, his patience and his numerous encouragements have helped me through the difficult moments of this thesis. Without his continuing support this dissertation would not have been completed. I also owe a particular debt to Jean Tirole for the interest he has shown in my work, for his precious advice and for his very kind encouragements. I was very fortunate to benefit from many illuminating discussions with Eric Maskin. Finally, I am grateful to John Moore for valuable suggestions and for his warm encouragements.

It was a great pleasure to work with Philippe Aghion on Chapter II and with Giacomo Bonanno on Chapter III.

For stimulating conversations and useful comments I am grateful to Dilip Abreu, Roland Benabou, Richard Caves, Mathias Dewatripont, Mike Dicks, Robert Gertner, Jerry Green, Maria-Jose Herrero, Reija Lilja, David Scharfstein, Guglielmo Weber and Richard Zaltz.

I am extremely grateful to the London School of Economics and the British Council for financial support. I would also like to acknowledge my gratitude to the Massachusetts Institute of Technology for allowing me to visit the department of Economics.

Finally I wish to thank Sue Kirkbride and Ann Usher for typing the thesis in a very short time.

# TABLE OF CONTENTS

## INTRODUCTION

The thesis comprises three self-contained chapters, which can be read independently. Each covers a specific aspect of the role of contracts in industrial organisation. The literature on contract theory in industrial organisation (or transaction-cost economics) has developed in a number of different areas in recent years. The purpose of this introductionn is to relate each chapter to the existing body of research. Over the past twenty years the literature has grown so large that it would be difficult to provide a comprehensive survey. Most existing surveys have had to concentrate on a specific line of research (see for example Baiman (1982); Blair-Kaserman (1983); Rey-Tirole (1985a) and Williamson (1985)). We shall, therefore, be very selective and will limit ourselves to presenting only the main issues at stake and the underlying themes in the existing literature.

Transaction-cost economics is concerned with three fundamental questions: Firstly, a large volume of research has been devoted to the problem of the optimal allocation of transactions inside the firm and in the market. This includes the question, "What determines the optimal size of the firm?" This is a very important question, since it underlies much of the literature on vertical and lateral integration, mergers and take-over bids. Standard microeconomic theory takes the market-structure as given and represents the firm as little more than a production-function. Transaction cost economics, on the other hand, views the corporation as a nexus of contracts, with the corporate charter taking a constitutional role for the firm. Hence, the <<firm versus market>> question reduces to an optimal contracting problem: the optimal allocation of residual rights through contracts.

The second broad question, which transaction costs - economics tries to tackle, is the optimal internal organisation of the firm, and how this structure affects the firm's behaviour in the market. For example, how is authority and control optimally delegated within the firm? What determines the hierarchical structure of firms? This field of investigation is generally known as managerial economics. It was first developed in the early sixties by, most notably, Marris (1964); Simon (1959), Williamson (1964), Cyert and March (1963). Two main themes are prevalent: (1)) most modern corporations are characterized by the separation of ownership and control, so that the objectives of managers do not necessarily coincide with those of the shareholders. (2) As a result, modern corporations do not seek to maximise profits and, as a consequence, their behaviour in the market will be very different from what standard micro-economic theory predicts. The formal foundations of managerial economics were not laid until the Principal Agent theory was developed in the 1970's.

The third main topic of interest concerns the issue of how the various contractual practices, which firms enter into affect competition in (intermediate) industries; also, in the opposite direction how competition disciplines individual firms' opportunistic behaviour. In particular, to what extent competition may act as a substitute for sophisticated incentive-schemes. The main difficulty here is that many of the contractual practices made between suppliers and buyers are designed to enhance efficiency, (by promoting the right incentives of each party), but may simultaneously be acting as coordinating devices to enforce a cartel, or perhaps be acting as entry-barriers against potential entrants. As Williamson pointed out, this leads to the direct confrontation of two fundamental contracting traditions:

<<The two polar contracting traditions for evaluating non-standard or unfamiliar contracting practices are the common law tradition and the antitrust or inhospitability tradition. Whereas contractual irregularities are presumed to serve affirmative economic purposes under the common law tradition, a deep suspicion of anticompetitive purposes is maintained by the antitrust (or inhospitability) tradition>> [Williamson (1985), p.200].

Naturally, all of the three main issues are closely linked. For example, it is difficult to separate the <<firm-versus-market>> question from the problem of whether contractual restraints serve the purpose of maintaining a monopoly position or whether they are meant to ensure the efficient execution of transactions. Similarly, issues of internal organisation of firms have a bearing on the <<firm-versus-market>> problem. These inter-connections are reflected in the three chapters in the thesis. Thus, while chapters two and three concentrate on the third issue of the role of contracts in intermediate markets they also deal with aspects of the first question indirectly, since they compare various contractual practices with the vertical integration solution. Chapter One, on the other hand is concerned with the internal organisation and delegation literature since it studies a specific monitoring scheme (random inspection) in the context of a Principal-Agent model. Despite the obvious interconnection between these three broad issues it is still useful to consider each one in isolation.

(1) The entrepreneur coordinator versus the price mechanism

Coase (1937) was the first to raise the question of why some transactions are undertaken inside a firm while others are mediated through the price system. He phrased his question as follows:

> <<Outside the firm, price movements direct production, which is coordinated through a series of exchange transactions on the market. Within a firm, these transactions are eliminated and in place of the complicated market structure with exchange transactions is substituted the entrepreneur coordinator, who directs production. It is clear that these are alternative methods of coordinating production. Yet, having regard to the fact that if production is regulated by price-movements, production could be carried on without any organisation at all, well might one ask, why is there any organisation?>> [(Coase (1937) pp.388].

Indeed, the very existence of firms implies some form of market failure. Having recognised this, however, the next question is: Why is not all production carried out by one big firm? Coase identifies several reasons which limit the size of the firm. They are all classified under the heading <<diminishing returns to management>> and include:

(1) the costs of organising production: If, beyond a certain point, the marginal cost of organising an additional transaction within the firm is increasing, then there will be a limit to the size of the firm.

(2) As the number of transactions increases, the entrepreneur will tend to misallocate resources more.

These are obviously somewhat vague categories. The reasons for market failure given by Coase are also obscure. They include:

(1) the costs of discovering what the relevant prices are on the market,

(2) the costs of negotiating and concluding a separate contract for each exchange transaction,

(3) the costs of writing completely contingent long-term contracts.

To be sure, these are costs that are also faced by the entrepreneur inside the firm. Thus Coase has not clearly identified the reasons for market failure that lead to the existence of firms.

The major challenge for research after Coase was to identify more precisely the determinants of how transactions are allocated between firms and the market. This is what Williamson has tried to answer (see also Klein, Crawford and Alchian (1978)).

Williamson's argument is in three steps: first, he delimits the underlying causes of transaction costs, attributing them essentially to a combination of asset specificity, bounded rationality, opportunism and uncertainty (this is also the line taken by Klein, Crawford and Alchian). The main difficulty is, however, in making precise the meaning of each of these four categories. The second step is in demonstrating why, and when, internal organisation is more efficient in minimising transaction costs. The last step is to identify the limnits to the size of firms.

The main problem with asset-specificity is that, even if there is competition between buyers and sellers in intermediate markets before the signature of a contract, once the buyer and the seller have signed the contract, (and have invested in specific assets), they are locked into a bilateral monopoly relationship. If an event then occurs which has not been anticipated by the parties, and is not covered by their agreement, then they will have to find a new agreement. In this situation, each will try to exploit his newly acquired monopoly position at the expense of the other. The bargaining outcome will in general differ from the sharing-rule that the parties would have agreed on ex-ante, had they anticipated the event. As a result, if parties are aware that they might find themselves in ex-post bargaining situations they may under-invest ex-ante, since they anticipate that they will not be able to fully appropriate the returns to their investment. (It is implicitly assumed here that ex- ante investment is not contractible).

As is clear from the above argument, asset specificity only poses a problem if situations arise in the future which have not been dealt with initially in the contract. Williamson argues that such events are likely to occur both because future outcomes are uncertain and because people either cannot foresee all the possible future events or find it too costly to write a complex contract which incorporates all possible future contingencies. In other words, people have bounded rationality. Finally, neither uncertainty nor bounded rationality would cause any problems if people were not pursuing their self-interest systematically. In Williamson's vocabulary, it is because people are opportunistic that events which have been left out of the initial contract are a problem, for when they do occur agents will bargain over the sharing of the remaining surplus in such a way as to induce ex-ante inefficient investment.

Asset-specificity should be understood in a very broad sense. Whenever a party to the contract invests in assets which are more valuable to the parties to the contract than to outsiders, it invests in a specific asset. Thus, on-the-job-training or learning-by-doing are both examples of specific human capital. It is clear that in the absence of asset-specificity, ex-post opportunism disappears and consequently, the inefficiency of initial investment.

The next question to consider is, why internal organisation might be more efficient at reducing transaction-costs than an appropriate long-term contract signed by a buyer and a seller on the market? Presumably, asset specificity and opportunism are still relevant aspects of transactions carried out inside the firm. Any argument which purports to show the superiority of internal organisation over the market must rest on the notion that internal organisation offers a wider set of instruments than the market. The difficulty then is to identify instruments available inside the firm but not outside it. Williamson identifies four such instruments:

First of all he argues that internal organisation is more flexible than outside contracting. It allows a swifter adaptation to a changing environment.

Secondly, when a buyer and a seller are members of the same firm they will tend to cooperate more than if they were members of different firms.

Thirdly, the firm is better at resolving internal conflicts than a judge. Managers can exercise their authority in a more selective and precise manner.

Finally, the firm knows more about its internal operations, since internal audits are better than external audits. In other words, internalising some operations will allow the manager to relax some information constraints.

Each of the four reasons given by Williamson is vague and questionable. Thus as Grossman-Hart (1985) argue, it is not clear why internal audits ought to be superior to external audits. One might even argue that in some cases internal audits are worse since they are more susceptible to be biased by corruption. Similar criticisms can be applied to the other reasons given by Williamson.

The last step in Williamson's argument is to explain why firms do not grow larger and larger, relative to the market. This is another tricky problem which has yet to receive satisfactory answers, as Williamson himself admits. In 1967 he argued that firms were limited in size because of the <<control-loss>> phenomenon which he describes as follows in his recent book:

<<If any one manager can deal with only a limited number of subordinates, then increasing firm size necessarily entails adding hierarchical levels. Transmitting information across these levels experiences ... losses ... which are cumulative and arguably exponential in forms>> [Williamson (1985), p.134].

This argument has also been explored by Mirrlees [1976] and Calvo-Wellicz [1978], who developed models where the hierarchy is formed by several layers of Principal-Agent relationships. These studies show that when some inefficiency arises at higher levels in the hierarchy it will trickle down to all subordinates. Thus the more subordinates and hierarchical tiers there are, the larger the aggregate inefficiency. This will tend to put limits to the size of firms.

This is a plausible argument but it is unsatisfactory mainly because it is implicitly assumed that the entire firm is manageed from the top and that information is transmitted from bottom to top across all the intermediate tiers. Clearly internal organisation can be much more flexible than this. Indeed there is no reason a priori why internal organisation should not be able to replicate what is feasible on the market, whenever the market organisation becomes more efficient. In order to get a better understanding of the costs of Bureaucracy and red-tape a more detailed study of internal organisation is necessary. We shall discuss some recent research in this direction later.

Another approach in the literature has been to identify situations where the vertical integration solution can also be achieved through restrictive clauses written into contracts between input-suppliers and downstream firms. The purpose of this research, was first of all, to explain why some standard clauses such as franchise-fees, tying arrangements, exclusive territories or resale-price-maintenance are commonly used in contracts between buyers and sellers in intermediate industries (see for example the surveys by Rey-Tirole (1985a) or Mathewson-Winter (1986)). The second aim was to bring about a consistent approach of antitrust law to vertical integration and vertical restraints. If both integration and vertical restraints have the same effects then they ought to be treated the same way. Blair and Kaserman (1983), explain that in the US:

> <<The legal status of the contractual alternatives to vertical ownership integration varies from per se illegality to presumptive legality.>> [Blair and Kaserman (1983), p.154].

As a consequence, this strand of the literature made clear that the relevant comparison is not between a vertically integrated firm and a competitive market, but rather between vertical integration and the alternative contractual arrangements, which are available in the market. This, in turn, makes it much harder to explain vertical integration. Indeed, it is extremely difficult to get away from the economic equivalence between vertical integration and some alternative optimal contract.

Most of the literature on vertical restraints not only shows that there exists an alternative contractual solution to vertical integration but also that such an alternative contract is usually of a very simple form. It only involves one or a combination of the standard clauses, which were mentioned above. Secondly, this literature shows that these vertical restraints are privately and socially efficient ((see Mathewson-Winter (1986)). There are two exceptions, however. One is Rey-Tirole (1985b), who show that when there is aggregate demand uncertainty or if there are shocks on costs, then none of the above standard clauses will in general be efficient. The second exception is Chapter 3 in this thesis. Here it is demonstrated that vertical integration dominates contracts with standard clauses, such as resale-price maintenance or franchise-fees and that as a consequence these are inefficient contractual arrangements. This does not imply that there does not exist an alternative contractual arrangement, equivalent to vertical integration. In fact we derive one such alternative contract.

One of the reasons why vertical control by ownership is equivalent to vertical control by contract in this literature is because the domain of feasible contracts is very large and the situations described are very simple. In more complicated environments, when parties are forced to write incomplete contracts, this equivalence ought to disappear. This is the point made by Williamson, but as was explained earlier, he did not provide reasons that are sufficiently precise to distinguish exactly when vertical integration is better than a contractual alternative. Recently, however, Grossman-Hart (1985) have attempted to provide a formal model of the costs and benefits of vertical integration along Williamsonian lines. They define integration as the power to exercise control over the assets owned by the firm. To avoid ex-ante inefficient investment caused by ex-post opportunism, it may be optimal (depending on how bad ex-post opportunism is and how specific ex-ante investment is) to vertically integrate so as to be able to control ex-post the other agent's assets. They argue that vertical integration is a better solution than a contract that specifically allocates rights to control assets in situations

<<where there are many aspects of a firm's operations, each of which may be important in a different contingency, and thus the costs of assigning specific rights of control, ex ante are much higher than the costs of assigning generalized control>> [Grossman-Hart (1985), p.11].

In other words, vertical integration is equivalent to a contract which assigns generalised control over all assets. Finally, Grossman-Hart identify the following costs of vertical integration:

> <<the owner of firm 1 will have the power to intervene in firm 2 in ways which may distort the incentives of firm 1's manager. Moreover, the owner cannot commit himself to intervene <u>selectively</u> in his subsidiary's operations since by their very definition residual rights refer to powers that cannot be specified in advance.>> [Grossman-Hart (1985), p.7].

The inability to intervene selectively has also been emphasised by Williamson in his recent book. It remains necessary to get a clearer understanding of how this might be related to the size of the firm, and why this problem gets worse as the firm grows larger. This requires a closer study of the internal organisation of the firm, an area to which we shall now turn.

## (2) The Internal Organisation of Firms and the Delegation Problem

Economists have become more and more interested in issues of internal organisation, partly because in many industries some firms have grown to be very large and complex organisations and the share of these corporations in the economy has grown larger and larger. Furthermore, many of these firms have diversified their activities in different markets and across national boundaries. They are usually widely held corporations, with a very large number of shareholders, each owning only a small fraction of the firm's assets and with most shareholders not taking an active part in running the corporation. Such modern corporations are characterized by the separation of ownership and control.

This separation has been considered very important by managerial economists who argue that because of it, large corporations do not behave like profit-maximising firms. Instead, it is thought that since managers run the corporation and because shareholders are free-riding on monitoring management, executives are free to pursue their own interests (which are different from the shareholders' interests), subject to a minimum profit constraint. The executives' objectives are typically, <<expense preference>> or <<emoluments>> (Williamson (1963)); the growth of the firm (Marris (1964)), or revenue maximisation (Baumol (1959))). More generally it is argued that managerial firms tend to grow larger than entrepreneurial (or profit-maximising) firms. Furthermore, the benefits that accrue to management from the consumption of perquisites are described as rents that managers extract from shareholders:

<<Emoluments represent rewards which if removed would not cause the managers to seek other employment>> [Willliamson (1963), p.1035].

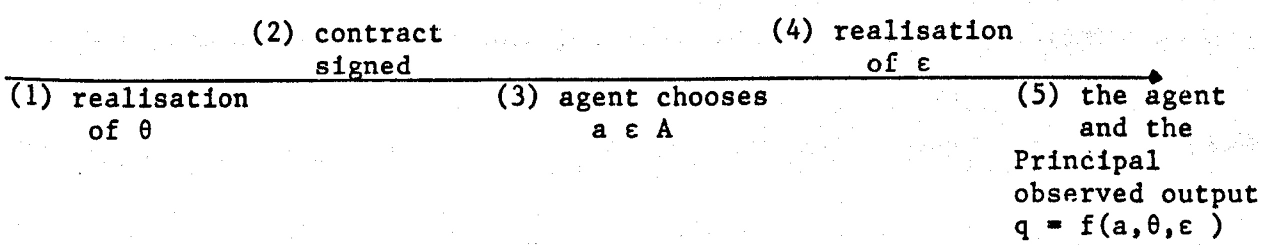
Most of the conclusions (and the arguments), of the early literature have been severely criticised. Jensen and Meckling (1976) have summarized one of the main objections to managerial economics as follows:

<<In practice it is usually possible by expending resources to alter the opportunity the owner-manager has for capturing non-pecunary benefits. These methods include auditing, formal control systems, budget restrictions and the establishment of incentive compensation systems which serve to more closely identify the manager's interests with those of the outside equity holders... >> [Jensen-Meckling (1976), p.323].

More generally, even if there is a separation of ownership and control is it clear that product and capital-market competition will not discipline the firm to behave as if it was maximising profits? We shall deal with this question in the next section. Despite the criticisms, it is now widely recognised that the delegation problem between shareholders and management _is_ important and that it requires better treatment.

The new foundations of managerial economics have been provided by the Principal-Agent theory. According to this theory, the shareholders are aware that management will try and pursue different objectives from theirs and therefore will write a contract with the manager that gives him the right incentives to run the firm in the shareholders' interest. For example, the manager may be offered shares of the company or stock options, which bring his objectives more closely in line with those of shareholders. (Alternatively, this theory says that managers are aware of the incentive problem and are prepared to write contracts which give the shareholders guarantees that they will run the firm in the owners' interests, in exchange of a higher remuneration). In fact, the Principal-Agent theory has also provided new foundations for most recent research on internal organisations. We will thus, briefly, review the basic principal-agent model before turning to the numerous applications. (There are two complementary surveys on the Principal-Agent literature; one by Baiman (1982) and the second by Hart and Holmstrom (1986)).

Two basic Principal-Agent models can be distinguished. One is called the hidden action model and the other the hidden information model. The main difference between the two models has been presented in the following terms by Caillaud-Guesnerie-Rey-Tirole (1984): Suppose that the profit generated by the agent, q, is a function of the agent's action $a \in A$ and two random variables, $\theta$ and $\varepsilon$. Thus $q = f(a,\theta,\varepsilon)$. Suppose furthermore that the realisations of these two random variables are known to the agent as follows:



Then the hidden action model corresponds to the case where both the Principal and the Agent know $\theta$ before the signature of the contract, but where the Principal only observes q (and not a and $\varepsilon$). Furthermore, the agent has different preferences over actions than the principal. The hidden knowledge model, on the other hand, corresponds to the situation where the principal does not observe $\theta$, but where he can observe (or infer from other observations) the agent's action.

In many situations both aspects are combined but it is useful to consider them in isolation. The hidden action model seems most natural as a representation of the shareholder-manager delegation problem, since the conflict of interest arises over the way the firm should be run and managers have superior information over their choice of action. Consequently, this has been the most commonly used model to study delegation problems inside the firm.

So far we have not said anything about how the conflict of interest is formalised. In the hidden action model what is emphasised is the trade-off between incentives and risk-sharing while in the hidden-knowledge model the trade-off is between truthful information revelation and the cost of screening. This latter cost may take different forms depending on what variable is used to screen the agent. These types of conflict are different from the one described in the early managerial theory of the firm. One weakness of this early literature was that most variables that entered the managers' utility function were observable and verifiable. Thus they were contractible and the divergence of interest between managers and shareholders could then be eliminated through an appropriate contract. In addition, the trade-off between incentives and risk-sharing explains why for example part (but not all) of managers' renumeration is through shareholdings. This is a widely observed practice which previous managerial theories failed to explain (see Jensen-Meckling (1976), pp,330-331).

The Principal-Agent theory does, however, have some drawbacks. Firstly, it does not explain why in practice linear incentive schemes of the form $R = w + s.q$ are so common (here R = manager's renumeration; w = base wage; q = profits; s = share of profits that accrue to the manager). Optimal incentive schemes in the Principal-Agent theory are usually much more complicated and severe restrictions on the distribution-function over $\varepsilon$ are needed simply to guarantee the monotonicity of the optimal incentive scheme. The point is that the optimal sharing rule is extremely sensitive to the form of the distribution function over $\varepsilon$. Thus Hart and Holmstrom (1985) write:

> <<The fact that we can view the optimal incentive scheme as responding to inferences is intuitively very appealing but also problematic for placing restrictions on s(q). The connection between q as physical output and as statistical information is very tenuous. In fact the physical properties of q are rather irrelevant; all that matters is the distribution of the <<posterior>> (or the likelihood ratio) as a function of the agent's action.>>

A notable exception to these problems is the recent paper by Holmstrom and Milgrom (1985). They explain that linear schemes may be optimal in a special multi-period Principal-Agent model, where the agent is assumed to have an exponential utility function.

The second and probably more serious weakness of the hidden-action model is that it cannot generate useful predictions as to the optimal choice of action by the agent and how his choice of action differs as the incentive-problem gets worse. This makes it very difficult to determine in what respect the behaviour of managerial firms differs from that of a profit-maximising firm. Holmstom and Weiss (1985) have developed a model where they show that managerial firms tend to under invest (in bad states) and as a result this may lead to greater variability in aggregate output and investment. The trick in their model, however, is to transform the hidden-action problem into a standard screening problem. Another study by Marcus (1982), has shown that because managers in managerial firms cannot diversify away the risk imposed on them by the contract signed with shareholders, they tend to both underinvest in risky projects and waste resources in reducing the variability in profits.

Despite these weaknesses, the Principal-Agent model has been fruitfully used to explain how the separation of ownership and control may imply a determinate debt-equity structure of the firm. Jensen and Meckling (1976) offer a very interesting discussion along these lines and suggest the following explanation of the determinateness of the debt-equity ratio. If the owner-manager finances part of his investment project through equity he will not receive all the benefits obtained by running the firm efficiently. He will thus have a tendency to slack. On the other hand, if he finances his investment project through debt, he will not bear all the costs if the project fails but he will get all the benefits if it succeeds. As a result he will tend to take a more risky course of action, which is against the interest of bond holders. Debt financing will then become more costly as the size of the debt becomes large. The optimal debt-equity ratio will then be determined when the cost of an additional unit of debt equals the agency-cost of an additional unit of equity. More recently, Grossman-Hart (1983) have provided another explanation of debt-equity ratios based on the idea that bankruptcy may act an an incentive scheme for the manager. If the latter dislikes the event of bankruptcy he can precommit himself to run the firm efficiently by incurring debt: the existence of debt implies the possibility of bankruptcy and the risk of going bankrupt (given the level of debt) can only be reduced if the manager runs the firm efficiently. The optimal debt-equity ratio is then determined as the outcome of the manager's optimal commitment problem. Both these studies, and others (see Ross (1977); Leland and Pyle (1977)) have stressed the importance of asymmetric information in explaining the optimal debt-equity ratio. One weakness in all these models, however, is that they

> <<beg the question why capital structure needs to be used for incentive purposes when direct incentive schemes would appear cheaper ...>> [Hart-Holstrom (1985), p.24].

This brings us to the question of what information regarding the agent's performance, ought to be taken into account in a direct incentive scheme?

Holmstrom (1979) answered this question by providing a very powerful sufficient statistic theorem. Suppose for example that instead of basing the manager's remuneration solely on profits, $q = p.x$ (where $P$ = price and $x$ = quantity sold), shareholders decided to base his remuneration on price and quantity variances. Then to disaggregate between price and quantity leads to an improvement if and only if disaggregation influences the likelihood ratio. That is, let $f(p.x|a)$ be the density of profits given the action, a, chosen by the agent and let $h(p;x|a)$ be the joint density of price and quantity given a; then disaggregation provides additional useful information about the agent's action if the following equality is not satisfied:

$$\frac{fa(px|a)}{f(px|a)} = \frac{ha(p,x|a)}{h(P,x|a)} \quad \text{almost everywhere} \quad a \in A.$$

(see Baiman-Demski (1980)).

Thus, if profits q are a sufficient statistic for (q,P) or (q,x) then there is no point is disaggregating. Profits are a sufficient statistic, for example, when to every profit outcome there corresponds a unique price-quantity pair.

The main consequence of Holmstrom's theorem is that the standard one-period Principal-Agent model where the only performance measure is profits, is too restrictive. Much additional information ought to be used in an optimal contract. As a result much research has been devoted to studying how repeated relationships and relative-performance evaluation could reduce the incentive problem by providing more information to the Principal (see for example Rogerson (1985) on repeated Principal-Agent relationships and Mookherjee (1983) on relative performance evaluation).

A third line of research has been devoted to optimal investigation policies. When the principal can observe another signal of the agent's action (or possibly the action choice of the agent), by paying a fixed cost, what is the optimal investigation policy and how should the transfer to the agent depend on the observation of this new signal? This problem has particularly interested theoretical accountants who have been concerned about two specific aspects of investigation policies, the decision-facilitating role of new information and the decision-influencing role [for a survey see Demski-Kreps (1982)].

The decision-facilitating problem is one where the investigator must decide whether or not to acquire further information about a process, when the information he has already obtained about it is imperfect and does not reveal whether the process is still in control or not. This is formally equivalent to a product-quality test problem, where the investigator must decide, on view of the results of a preliminary test, whether to carry out another test or not.

The decision influencing problem is basically an incentive problem where the principal must decide how to make best use of the threat of inspection to discipline the agent. Typically, in most investigation problems inside the firm, the two aspects are combined. For example, shareholders may decide to audit management both for incentive purposes and to acquire additional information which will help to allocate new investment. Gjesdal (1981) has compared the value of information for decision-facilitating and decision-influencing purposes in a general one-period Principal-Agent model. Not surprisingly he finds that the value of an information system is not the same in general, when used for decision-facilitating purposes and when used for incentive reasons.

One of the main difficulties with optimal random investigation policies, when used for incentive purposes, is that it pays to impose a penalty that is as large as possible on an agent who is found shirking. This is the principle of maximum deterrence. The problem is that this principle is rarely applied in practice and it is hard to find straightforward explanations for this state of affairs. Chapter 1 takes a closer look at this problem and investigates whether the possibility of punishing an innocent agent might imply a bound on penalties and thus invalidate the principle of maximum deterrence. The answer is somewhat surprisingly negative. This result still holds when the principal makes inspection-errors. (Note that the results obtained depend very much on the model specification. For example Polinsky-Shavell (1979) obtain the opposite conclusion in a different model (see Chapter 1)).

To close this section we shall briefly mention other applications of the Principal-Agent paradigm to issues in internal organisation. Firstly, Holmstrom (1982) has used a Principal-multi-agent model to formalise the idea suggested by Alchian-Demsetz (1972) that the main role of the firm is to ascertain the marginal products of team members. Thus if the production of a given commodity or service is undertaken by a team, then the market will not be able to distinguish each team member's contribution. As a result, Alchian-Demsetz argue that each member will have an incentive to shirk and free-riding can only be reduced if a manager specialises in the activity of monitoring each member's performance.

Holmstrom gives a slightly different interpretation of the role of the entrepreneur-manager. He shows that any sharing rule (of aggregate profits) which satisfies budget balance must yield an inefficient outcome (i.e. agents are free-riding). The role of the entrepreneur-manager is then to break this budget balance and become a residual claimant, in order to achieve an efficient outcome. In other words, the manager sets up a Groves-revelation mechanism to induce each worker to truthfully reveal his action choice. It is well known that he can only do this by breaking budget-balance.

A second application of the Principal-Agent theory has been to study hierarchies inside the firm. In section 1 we have mentioned some work by Williamson (1967), Mirrlees (1976) and Calvo-Wellicz (1978) along these lines. What these studies have overlooked however is that as soon as we have a three-tier relationship of the form, Principal-Supervisor-Agent, (which in practice characterizes most hierarchies), there is scope for collusion between some members in the hierarchy against upper (or lower) tiers. This problem has been studied by Tirole (1985), who shows that the incentive problem gets worse when one takes into account collusive behaviour. In particular, it is shown that the natural coalition is between the supervisor and the Agent (more generally, collusion tends to take place <<at the organisation's nexus of informed parties>> [Tirole (1985), p.43]. Thus, the supervisor tends to act as an advocate for the agent. In addition, repetition of the relationship is now less desirable (or even undesirable) since it facilitates collusive behaviour.

Finally, collusion in organisations may explain why they tend to be run by rules leaving little discretionary power to intermediate tiers, since then supervisors and agents have little scope for cheating the Principal.

### (3) Competition and Long-term Contracts

Compared to the vast literature on the question of markets versus hierarchies and on the internal organisation of firms, relatively little research has been undertaken on how competition reduces the scope of opportunistic behaviour in contractual relations. On the other hand, competition is often invoked in informal discussions about managerial slack as a disciplinary force which suffices to bring managerial firm's behaviour in line with that of a profit-maximising firm. Thus, in his review of Galbraith's book: <<The New Industrial State>>, Solow wrote:

> <<It is possible to argue - and many economists probably would argue - that many management controlled firms are constrained by market forces to behave in much the same way that an owner-controlled firm would behave, and many others acquire owners who like the policy followed by the management>> [Solow (1967), p.103].

The disciplinary force of competition has been emphasized both in the capital market (see Manne (1965)) and in the product market (see Machlup (1967)). More recently, Fama (1980) suggested a reputation-argument, whereby managers and other agents inside the firm are efficiently disciplined by competition and the desire to establish a reputation on the market for their services.

Vice-versa, as far as the effect of long-term contracts on competition is concerned, another informal argument often heard, is that long-term contracts between a buying firm and a selling firm cannot prevent competition from working efficiently. For example, Robert Bork put forward the following argument:

> <<The problem is to know what exclusion is improper. All business activity excludes. A sale excludes rivals from that piece of business. Any firm that operates excludes rivals from some share of the market. Superior efficiency forecloses. Indeed, exclusion or foreclosure is the mechanism by which competition confers its benefits upon society. The more efficient exclude the less efficient from the control of resources and they do so only to the degree that their efficiency is superior.>> [Bork (1978), p.137].

There have been several attempts recently in formalising the idea that competition plays a disciplinary role in various markets. Most studies conclude that, in general competition is not a sufficient disciplinary force to obtain efficiency. Worse than this even, it is possible to show that competition may have adverse effects in reducing managerial slack. In a similar vein Chapter 2 demonstrates that long-term contracts may have adverse effects on competition. That is, market foreclosure may lead to social inefficiencies. Several authors have also developed arguments that portray the role of certain contractual clauses as coordinating devices of cartels. We shall first review the literature on the effect of competition on incentives and then that regarding the anticompetitive aspects of contractual relations in intermediate industries.

The seminal paper by Grossman-Hart (1980) has uncovered several difficulties with the argument that the threat of take-overs is sufficient to discipline management. Firstly, it is pointed out that take-overs may fail because of a free-rider problem among shareholders. The reason is that a raider is only willing to take over the firm if he can increase profits after the raid. Existing shareholders are aware of this and may want to hold on to their shares in order to benefit from the increase in profits, hoping that a sufficiently large number of other shareholders is willing to sell and hence make the raid feasible. But every shareholder will make the same reasoning so that no single shareholder will be prepared to sell and thus the take-over attempt is bound to fail. Because of free-riding there may not be any disciplinary threat of take-overs at all. Grossman-Hart suggest that shareholders can improve on this state of affairs by somehow committing themselves not to get the full benefit of the take-over if it takes place. That is by diluting their property-rights. From a social welfare point of view, of course, the threat of a take-over should be maximised by having maximum dilution but this does not agree with the shareholders' private interest. They would get a very low tender price if dilution was large. Thus, in general take-overs will not lead to social efficiency.

A second point implicit in Grossman-Hart's analysis is the idea that the take-over mechanism is not a substitute for an optimal incentive scheme but rather supplements an optimal delegation contract. If there was no incentive-scheme then whenever there is separation of ownership and control, the manager would slack, thus profits would fall. If this is anticipated at the time that the raider wants to sell his shares again, then the latter could not make a profit simply by buying up a firm, reorganising it and selling it back. He can only make a profit if he introduces a better incentive scheme than the old one when he takes over the firm. This leaves open the question of why existing shareholders did not introduce the optimal incentive-scheme in the first place? There may be many reasons for that and recently Scharfstein (1985a) has explicitly modelled how the existing incentive scheme can become outdated. He assumes that management is likely to become better informed than shareholders overtime and can then use its informational advantage to slack. Scharfstein then shows that an optimal incentive scheme involves the threat of a take-over but that from a social efficiency point of view again this threat will be <<too small>> since shareholders tend to set tender prices <<too high>>.

Schleifer and Vishny (1985) have extended the Grossman-Hart model by explicitly modelling the raiders' search process of target-firms and competition among raiders. In their model an attempt is made to explain the commonly observed take-over resistance tactics used by incumbent managers. Also, Harris-Raviv (1986) have recently modelled the role of the financial structure of the firm as a means to reduce the threat of a take-over: In the short-run, an owner-manager can more easily prevent a take-over (when this is in his interest) by substituting debt for shares held by other shareholders.

If competition in the capital market cannot completely discipline management, neither can competition in the product market. The disciplinary role of product-market competition has often been considered in evolutionary terms and a recent exposition of this approach can be found in Nelson-Winter (1982). An old question in economics is whether or not one can validate the hypothesis of profit-maximisation by appealing to some natural selection argument, whereby only profit-maximising firms can survive. This is indeed how Friedman (1953) justified profit-maximisation. Now Nelson-Winter (1982) show that this need not be true: there are many decision-rules (some very inefficient in particular environments) which can survive along with profit-maximisation rules in long-run equilibrium. The reason is that many rules may work as well as profit maximisation in a stable environment but very poorly in new circumstances not often encountered. Then, as long as there is no systematic perturbation in the long-run these rules will survive. The difficulty with this literature, for our purposes is that it does not yield any predictions as to the influence of competition in the product-market on the contracting process inside the firm, since this process is not explicitly modelled.

The first formal attempt in this direction was by Hart (1983) who considers a market where a fraction of managerial firms competes with entrepreneurial firms (e.g. profit-maximising firms). In his model the incentive problem is extreme in two respects: firstly because the manager's effort and output are not observable by shareholders. Thus all the latter can insist on is some minimum dividend payment, independent of the profits realised by the firm. Secondly, managers are assumed to have very special tastes. Their utility function has the form: $U(I) - V(a)$ where $U' = \infty$ for $I < \bar{I}$ and $U' = 0$ for $I > \bar{I}$. Shareholders are uncertain about the firm's costs which can be high or low, but these costs are correlated among firms. This correlation is sufficient for competition to play a role in reducing slack in managerial firms. The reason is that when costs in a managerial firm are low, they will also be low in some entrepreneurial firms. As a result the latter will expand output. This lowers equilibrium prices and managerial firms' profits. Thus shareholders can give a smaller rent to low-cost firms to get their managers to truthfully reveal their costs.

Finally, Hart argues that if one assumes that it is more costly to set up an entrepreneurial firm than a managerial firm (because of monitoring costs) then product-market competition cannot eliminate managerial slack (this can only be achieved by increasing the fraction of entrepreneurial firms in the market), for if it did, then there would be no incentive for setting up an entrepreneurial firm.

One may wonder whether competition could not have a more disciplinary effect on managers if the latter were assumed to have more reasonable tastes. Scharfstein (1985b) has shown that, surprisingly, with more plausible managerial tastes, competition in the product market can make things worse. Instead of the utility function specified by Hart, he assumes that:

$$U(I) = \begin{cases} \alpha + \beta I & \text{if } I \geq 0 \quad ((\beta > 0) \\ -\infty & \text{if } I < 0 \end{cases}$$

(when $\beta=0$, his utility function reduces to Hart's utility function). In this more general setting, Scharfstein explains that the following problem may arise:

<<Now consider the effect of an increase in competitive pressure from entrepreneurial firms. This leads to a drop in the price in the bad state (when costs are high) since entrepreneurial firms are more efficient, thus ceteris paribus, the profits of managerial firms in the bad state go down. It then becomes easier for managers in the good state (when costs are low) to shirk since the profit target they must meet is lower. This effect feeds back into the design of the optimal contract, resulting in greater slack in the bad state. Rather than mitigating the incentive problem competitive market pressure exacerbates it>> [Scharfstein (1985b), p.2].

This phenomenon could not arise in Hart's model, since any change in I (the profits that the manager gets) does not affect the managers utility, (unless $I < \bar{I}$). Therefore, neither does it affect his incentive to slack.

A final area concerns the disciplinary role of competition among managers and workers inside and outside the firm. This has also been recently formalised. In particular, Fama's argument that managerial slack will be eliminated because managers are concerned about their reputations, has been analysed by Holmstrom (1982b) in a simple dynamic model. He assumes that the manager's output is a noisy function of ability and effort, that no contingent contracts are available, and most importantly, that ability is initially unknown both to the market and to the manager himself. Since contracts are non-contingent, the only reason why a manager would want to increase his supply of effort today is to influence the market's future beliefs about his ability. The supply of effort is then most valuable when little is known about the manager's ability. As a result managers tend to supply too much effort when they are starting their career and little is known about them and too little effort after successive observations of output make the market's beliefs much more precise. Efficiency, on the other hand, requires a constant supply of effort over time. The basic point of Holmstrom's argument is that reputation formation is valuable only temporarily. Furthermore, if considerations of risk-aversion by managers were also taken into account there would be even greater inefficiency. Indeed, Holmstrom provides two examples where reputation-considerations when managers are risk-averse, induce the latter to choose safer projects than shareholders would like.

We shall now turn to the other side of the coin, namely how long-term contracting in intermediate industries affects the competitive process.

There has been a long debate in the antitrust literature about exclusive dealing contracts and their effect on the competitive process. A focal point in this debate has been Judge Wyzanski's controversial decision in the famous case:

<<United States versus United Shoe Machinery Corporation>>
(see Bork (1978; Caves (1984) and Posner (1976)).

United Shoe was producing shoe-machinery which it then leased to shoe-manufacturers. Over the years it has developed a complex system of leases whereby only a small fraction of shoe-manufacturers saw their leases expire at any given time. United Shoe's market share was more than 80% and it was thought that the leasing system developed by United Shoe was set up to prevent entry. Judge Wyzanski decided that this allegation was well-founded and declared that United Shoe had violated the Sherman act.

His decision has been a subject of continued debate, despite the fact that after his ruling United Shoe's market share dropped considerably. Judge Wyzanski's critics have basically argued that long-term contracts could not have an entry preventing effect which is socially harmful, since it is hard to conceive why a buyer (seller) would be prepared to perpetuate a seller's (buyer's) monopoly position by signing a long-term contract with him. The same point was made again in a number of other famous legal cases like <<Federal Trade Commission v. Motionpicture Advertising Service Co.>> or<< Standard Oil Co. of California and Standard Stations v. United States.>> (see Bork (1978).

In a formal model set up in Chapter 2, we demonstrate that exclusive dealing contracts will have an entry-preventing effect for much the same reasons as why the take-over mechanism does not work perfectly in the capital market (one may view the raider as a potential entrant into the market represented by the firm he wants to raid).

The crux of the argument is that whenever an incumbent seller signs a long term contract with a buyer, the latter must pay damages to the seller if he breaches the contract and switches to the entrant. These damages are like an entry fee which the entrant must pay to the contractors. (The latter can then split this fee between them in whichever manner they desire). Thus the buyer and the seller as a coalition will set this entry fee in the same way as a monopolist sets his price when he does not know the reservation-values of his customers: not too high (in order not to discourage entry too much) and not too low (in order to get a high enough entry fee). Of course social efficiency requires that this entry-fee should be set at a level where a more efficient entrant is not prevented from entering. Private incentives, on the other hand generate higher entry fees than the socially efficient level.

This is a very general principle. It explains the inefficiency of the take-over mechanism in Grossman-Hart (1980)) and Scharfstein (1985a), for example. It may also be an explanation for severance pay in labour contracts. It also explains why in many markets (take for example the housing market) buyers are required to make down-payments before the final delivery of the goods. (There are of course alternative explanations like Williamson's hostage theory, when there are specific assets involved, which would also explain down-payments).

As Posner and Bork pointed out, the entry preventing nature of such contracts depends on the length of the contract. Indeed, in <<FTC v Motion Picture Advertising Company>> the judge ruled that exclusive dealing contracts of limited length (one or two years) were permissible. Given the importance of the length of the contract in this context, one may wonder what determines the optimal length of the contract? This turns out to be a very difficult question. On the one hand a well known principle says that (absent transaction costs) parties should sign the longest possible contract whenever they engage in mutually advantageous trade. On the other hand, most contracts in practice are signed for an explicit finite duration. Clearly,

<<transaction costs>> and <<bounded rationality>> must be important reasons for why contracts are of finite duration but this information is not of much help since <<transaction costs>> is a notably vague category and we do not have a satisfactory theory of <<bounded rationality>>. We have therefore attempted to follow another route in Chapter 2. We seek to explain the finite length of contracts through asymmetric information at the time of signature of the contract:

Suppose that initially the seller has superior information about the probability of entry than the buyer. Then the length of the contract becomes a signal for the seller's information:

If the latter knows that the probability of entry is small he can signal his information by signing a short-term contract. The reason is that it would be too costly for a seller who faces a high probability of entry to mimick a seller who faces a low probability of entry, by signing a short-term contract. The length of the contract is then increasing with the likelihood of entry. We thus have a simple theory of contract length which allows us to sidestep the difficult question of what is <<bounded rationality>>.

Exclusive dealing contracts are not the only impediment to competition. Contracts in intermediate industries may also be socially harmful because they serve as coordinating devices for producer cartels. Of course, << price fixing>> arrangements or <<concerted refusals to deal>> are illegal, but seemingly innocent clauses may have the same indirect effect as these explicit cartel-coordination practices. It is not surprising then that there have always been two different approaches to non-standard contractual arrangements, one, - (the Coasian view) - emphasizing the private (and possibly social efficiency) aspects of various contractual provisions, the other - (the anti-trust view) - trying to find anti-competitive motives behind these various clauses.

Resale-price-maintenance (RPM) is a classic example of such a contractual restraint. Indeed, in his seminal article, Telser (1960) gave both an efficiency explanation for RPM and an anti-competitive explanation. In Chapter 3, we summarize his efficiency explanation and also provide an alternative efficiency explanation to Telser's, based on post-sales service-competition by retailers.

Here we briefly mention his other explanation. Telser argued that cheating on a manufacturers' cartel was more difficult if all manufacturers imposed RPM on their retailers than if they did not. The reason is that deviation from the cooperative strategy is more easily detected by the other cartel members, when there is RPM, since by cutting the wholesale price a chiseler can only expand into other outlets and thus substantially reduce the market-share of his competitors. This will be easily detected. If, on the other hand, he also allowed retailers to lower his retail price then he could claim that the increase in his market share was due to an increase in demand induced by the fall in the retail price (and not the wholesale price which is assumed to be unobservable). Unfortunately, this argument has not been formalised yet and it is unclear whether it would stand on its feet if it was analysed more carefully.

Recently Salop (1986) argued that two other contractual provisions the "clause of the most favored nation" and the "meeting the competition clause" may serve as cartel coordination devices.

The most favoured nation clause (MFN) guarantees a buyer the same price as the lowest price conceded by the seller to any other buyer. Such a clause can reduce competition among sellers considerably, since it makes it very costly for sellers to engage in fierce price-competition over new clients.

The meeting the competition clause (MCC) guarantees a buyer the same price as the lowest price offered in the market by any seller. Usually, it is the entry-preventing effect of MCC's that is emphasized. If an incumbent seller adopts an MCC he will eliminate the threat of entry since he thereby makes sure that an entrant will never have any customers. In Chapter 2 we show, however, that for entry-deterrence purposes such clauses are inefficient since they completely preclude entry and thus prevent the buyer and the seller from cashing in on the entry fee. Salop, however, argues that they facilitate cartel-coordination:

> <<Buyers are automatically given the rival's lower price until all firms raise their prices. This eliminates the transitional losses that might otherwise deter price-rises. It also eliminates the rival's transitional gains and with it the incentive to delay a matching price increase.>> [Salop (1986), p.281].

It is certainly also possible to give efficiency-explanations for the existence of these types of clauses (see, for example, the discussion following Salop's paper in Stiglitz-Mathewson (1986)), thus making the antitrust judge's task extremely difficult. The interactions between competition and contracting in intermediate industries are extremely complex and have as yet been insufficiently well explored.

# CHAPTER 1

# RANDOM INSPECTION AND OPTIMAL PENALTIES IN THE PRINCIPAL-AGENT RELATIONSHIP

## INTRODUCTION

The purpose of this chapter is to reconsider the problem of the optimal size of penalties to be imposed on a shirking agent when the principal is allowed to inspect the agent's choice of action, ex-post. This problem has received a lot of attention in a wide variety of contexts. To our knowledge, Becker (1968) was the first to address this issue in his seminal paper on Crime and punishment. More recently, a number of authors have dealt with this problem in the context of adverse-selection models of the insurance market (Stiglitz (1975)) and the labour-market (Guasch and Weiss (1980, 1981, 1983)); Nalebuff and Scharfstein (1985)). There are also a series of studies in the theoretical accounting literature, most notably by Baiman and Demski (1980a, 1980b) and in the Regulation literature (Baron and Besanko (1984)).[1]

One of Becker's main points was to show that when law-enforcement is costly, society can reduce enforcement costs to a minimum by imposing fines on offenders, which are as large as possible (equal to the wealth of the offender), provided that no innocent person is punished. The deterrence effect of large fines is stronger, so that society can reduce the likelihood of detection of offenses and thereby save resources. When agents are risk-neutral this result is true both in cases where it is optimal to deter everybody from commiting an offense and in situations where it is Welfare-maximising to have some agents commiting the offense. (In Becker's own terminology, when the social value of the gain to these offenders from the offense exceeds the social cost).

Polinsky and Shavell (1979) have shown, however, that when agents are risk-averse, then in this second case it is no longer true that the optimal penalties should be as large as possible. The reason is that the lower the probability of detection and the higher the fine, the more risk is imposed on those who have a positive net social value when commiting the offense.[2] As a result those individuals may be over-deterred.

Of course, overdeterrence is a good thing when society wants to prevent everybody from commiting certain offenses such as theft, rape or murder. In these cases risk-aversion only strengthens Becker's conclusion. Similar results have been established by Stiglitz (1975), Townsend (1979), Guasch and Weiss (1981) and Nalebuff and Scharfstein (1985).

Economists are generally unhappy with this result. First of all, this principle of maximum deterrence is not applied in most Western legal systems. The principle that punishment should fit the crime is usually the rule and it directly contradicts the idea of maximum deterrence. Secondly, if this proposition was taken seriously, almost all incentive-problems that have been studied over the past twenty years would find a very simple solution: with unbounded penalties the first-best outcome can be approximated arbitrarily closely by making the probability of detection (and thus inspection costs) arbitrarily small.

To avoid this conclusion some authors assume that penalties are bounded; the size of the penalty cannot exceed some number $0 < K < \infty$ and the agent's utility U(K) is bounded away from minus infinity. This is usually justified by appealing to some form of limited liability (see Baiman and Demski (1980a,b); Sappington (1983); Baron and Besanko (1984)). The first-best is no longer attainable then, but the principle of maximum deterrence still holds since the optimal solution in these models is usually that the optimal penalty is as large as possible. One is then faced with the question: What determines the constraint on penalties? If the efficiency of the contract can be improved by increasing the size of the penalty, why not contract around the limited liability constraint by increasing the size of the collateral for example?

Some authors like Stigler (1970) or Harris (1970) have argued that maybe one reason for why the principle of maximum deterrence should fail is that there is always the risk of punishing someone who is innocent. We propose here to analyse this argument in greater detail, in the setting of a Principal-Agent model. We also address the question of commitment to a given inspection policy. All the above-mentioned studies have assumed that the Principal can commit himself to a given inspection strategy. While this is a reasonable assumption to make when one deals with Crime and Punishment it may not be in other contexts.

The Paper is organised as follows: In section 2 we present the model and discuss the maximum deterrence result under the assumptions of commitment and non-commitment to an inspection strategy. In section 3, we consider the problem of random inspection, when the Principal makes type-one and type-two errors when he inspects. Our main conclusion is that, contrary to what intuition suggests, the risk of punishing an innocent person, no matter how high it is, provided that the principal does not systematically punish an innocent agent when he inspects, does not imply that optimal penalties will be bounded, even when the agent is very risk-averse (so long as he is not infinitely risk-averse). Section 4 presents some concluding remarks.

## SECTION 2: RANDOM INSPECTION WITH NO OBSERVATION ERRORS

It is now well known that the first-order approach to the Principal-Agent problem is unsatisfactory unless one is prepared to make severe assumptions about the distribution function over output (see Mirrlees (1974, 1975, 1979) and Rogerson (1985)). We did not want to restrict ourselves, at the outset, to very special distribution functions, so we follow the approach by Grossman-Hart (1983).

The Principal hires an Agent to perform a certain task. For any action, a, chosen by the Agent from his action set, A, there are n possible profit (or output) outcomes, $(q_1,\ldots,q_n)$ that occur with probability $(\pi_1(a),\ldots,\pi_n(a))$; where $\pi_i(a) > 0$, for all i and $\sum_{i=1}^{n} \pi_i(a) = 1$. Usually one assumes that the principal does not observe a, but that $q_i$ is publicly known. Thus, he can make the payment to the agent contingent on the observation of output. Let $t_i$ be the monetary transfer to the agent when the principal observes $q_i$.

The Agent's preferences are represented by a Von Neuman-Morgenstern utility function U(t,a), which is assumed to be separable in income and actions: $U(t,a) \equiv V(t) - a$.[3] He is willing to work for the principal only if he gets a reservation utility, $\overline{U}$.

The principal is assumed to be risk-neutral for simplicity. Furthermore, it is assumed that the Agent's utility function, his action set and the function $\pi$: $A \to S$ (where $S = \{x \in R^n / x_i \geq 0$ and $\sum_{i=1}^{n} x_i = 1\}$) are common knowledge.

Thus the Principal solves the standard program:

$$\max_{\substack{t_i \in [\underline{t},\overline{t}] \\ a \in A}} \sum_{i=1}^{n} \pi_i(a)\,(q_i - t_i)$$

subject to:

$P_1$

(IR) $$\sum_{i=1}^{n} \pi_i(a)\, V(t_i) - a \geq \overline{U}$$

(IC) $$\sum_{i=1}^{n} \pi_i(a)\, V(t_i) - a \geq \sum_{i=1}^{n} \pi_i(\hat{a})\, V(t_i) - \hat{a}$$

for all $\hat{a}$ in A.

We follow Grossman and Hart in assuming:

A.1: $V(.)$ is continuous, concave and strictly increasing on the open interval $(\underline{t}, +\infty)$, where $\underline{t} > -\infty$, but $\lim_{t \to \underline{t}} V(t) = -\infty$.

A.2: Let $M = \{v/v = V(t) \text{ for some } t \in (\underline{t}, +\infty)\}$. Then $(\overline{U}-a) \in M$, for all a in A.

The interpretation behind $\lim_{t \to \underline{t}} V(t) = -\infty$, is that the agent suffers an infinite loss, in utility, when all his wealth is taken away from him.

We shall modify the program $P_1$ slightly, by allowing the principal to inspect the agent's action, ex-post. We shall begin by assuming that he can observe the Agent's action exactly by paying an inspection cost $C > 0$. In the contract, the Principal must now specify a transfer to the agent when he inspects, which will be a function of his observation, $s(a)$; a transfer when he does not inspect, $t_i$ and an inspection rule. This rule will in general be a function of the principal's output observation: for every outcome $q_i$ the principal must specify a probability of inspection $p_i \in [0,1]$. We assume, for the moment, that the principal can precommit himself to a given inspection policy.[4]

He now faces the following program:

$$\text{Max} \quad \sum_{i=1}^{n} \pi_i(a) \; \{(q_i - t_i)(1-p_i) + p_i(-C-s(a))\}$$

$(t_i)$, $s(a) \quad \varepsilon \; (\underline{t} \; +\infty)$

$a \; \varepsilon \; A$

$P_2 \qquad p_i \; \varepsilon \; [0,1]$

subject to:

$$(\text{IR}) \quad \sum_{i=1}^{n} \pi_i(a) \; \{(1-p_i)V(t_i) + p_i V(s(a))\} - a \geq \overline{U}$$

$$(\text{IC}) \quad \sum_{i=1}^{n} \pi_i(\hat{a}) \; \{(1-p_i)V(t_i) + p_i V(s(\hat{a}))\} - \hat{a} \leq \overline{U} \,,$$

for all $\hat{a}$ in A, $\hat{a} \neq a$.

It is immediate from the IC-constraint what the form of the optimal contract will be. Define $a^*$ to be the first-best action and consider the worst case for the principal, where in the optimal contract all $p_i$ are strictly positive. Then the principal can implement $a^*$ and make all $p_i$ arbitrarily small by letting $s(a)$ tend to $\underline{t}$ for all $a \neq a^*$ and setting $s(a^*) = t_i = V^{-1}(\overline{U}+a^*)$. Such a contract is incentive-compatible and satisfies the IR-constraint.

Moreover, this contract approximates the first-best outcome since the agent chooses a*, is perfectly insured and the expected inspection costs of the principal, $\sum_{i=1}^{n} \pi_i(a^*)\, p_i.C$, are negligible.

(The above contract is optimal, a forteriori; when the principal can set some $p_i$ equal to zero.) Of course, if the agent's utility-function was bounded below, the first-best would not be approximated but the principle of maximum deterrence would still hold.

How is the optimal contract modified if we do not allow the principal to precommit to a given inspection-policy? In the absence of commitment, the principal and the agent play a sequential game, where the timing of moves is illustrated below:

| | | | |
|---|---|---|---|
| (1) contract signed: $c=\{s(a);(t_i)\}$ | (2) agent chooses $a \in A$ | (3) occurrence of $q_i$ | (4) Principal chooses $p_i \in [0,1]$ |

Clearly, inspection by the principal will only be credible when $q_i$ is observed, if

(1) $\qquad t_i \geq C + s(a)$

where a is the action implemented by the contract.

When (1) holds with equality, the principal will be indifferent between inspecting and not inspecting. Without loss of generality, we shall assume that he will then choose to inspect with probability one.

Let $F_a = \{i/t_i \geqq C + s(a)\}$ and $F_a{}^c$ be the complement of $F_a$. Without precommitment the principal faces the following program:

$$\max_{\substack{\{(t_i);s(a)\}\\ a\varepsilon A}} \sum_{i\varepsilon F_a} \pi_i(a)(q_i - s(a) - C) + \sum_{i\varepsilon F_a^c} \pi_i(a)(q_i - t_i)$$

subject to:

$P_3$

$$\text{(IR)} \qquad \sum_{i\varepsilon F_a} \pi_i(a)V(s(a)) + \sum_{i\varepsilon F_a^c} \pi_i(a)V(t_i) \geqq \bar{U}+a$$

$$\text{(IC)} \qquad \sum_{i\varepsilon F_a} \pi_i(\hat{a})V(s(\hat{a})) + \sum_{i\varepsilon F_a^c} \pi_i(\hat{a})V(t_i) - \hat{a} \leqq \bar{U}$$

for all $\hat{a}$ in A.

Here again, whenever the optimal contract involves inspection in some state with positive probability (i.e. F is non-empty), we can see from the IC-constraint that an optimal contract will be such that $-s(\hat{a})$ is as large as possible, for all $\hat{a} \neq a$. Thus, the principle of maximum deterrence applies whether there is commitment or no commitment to an investigation-policy by the principal. The first-best outcome, on the other hand, will not be approximated in general, since F is non-empty. This is altogether not very surprising and suggests that we must bring new features into the model if we want to obtain a result where optimal penalties are bounded. We propose to do this in the next section, by introducing observation-errors by the principal when he inspects.

## SECTION 3: RANDOM INSPECTION WITH TYPE-ONE AND TYPE-TWO ERRORS

In this section we investigate the case where the principal may only imperfectly observe the agent's action when he inspects. We shall proceed as follows: first we solve for the optimal contract in the simplest possible example. Then, whenever possible, we shall explain how our results are modified when the example is generalised.

In this example the agent's action set is given by $A = \{a_0, a_1\}$, where $a_0 < a_1$. That is, the agent can either work hard or slack. Furthermore, it is assumed that the principal cannot make the contract contingent on the output observation, $q_i$. The only information the principal obtains about the agent's action-choice is the signal he observes when he inspects the agent. The second assumption is verified in situations where the principal must pay the agent before he observes $q_i$. For example, when the agent is a building-constructor, the principal often finds out only 10 or 20 years after the completion of the building, what the quality of the construction is, but while construction is underway, he may randomly inspect the agent. Moreover, this is the relevant example to consider if one is interested in law-enforcement issues.

Suppose now that the agent has chosen action $a_i$, $(i = 0,1)$, then the probability that the principal will observe action $a_i$ when he inspects is strictly less than one:

$$Pr(\tilde{a} = a_i / a_i) < 1$$

where $\tilde{a}$ is the signal observed by the principal. We define:

$$\beta_1 = Pr(\tilde{a} = a_1/a_1)$$

$$\beta_0 = Pr(\tilde{a} = a_0/a_0) \ .$$

We assume that the principal can precommit to a given inspection policy, $P \in [0,1]$ and that $\beta_1, \beta_0$ are common knowledge. Also, for the contract to be enforceable by a court we must assume that the signal $\tilde{a}$ observed by the principal when he inspects is public knowledge.

Now when the principal offers a contract $c = \{t, s(a), P\}$ to the agent, the expected payoff of the agent when he chooses action $a_1$ and $a_0$ respectively is given by:

$$EU(c,a_1) = (1-P)V(t) + P(\beta_1 V(s(a_1)) + (1-\beta_1)V(s(a_0))) - a_1$$

(2)

$$EU(c,a_0) = (1-P)V(t) + P(\beta_0 V(s(a_0)) + (1-\beta_0)V(s(a_1))) - a_0$$

For notational convenience, let $s(a_0) \equiv s_0$ and $s(a_1) \equiv s_1$.

We can restrict the analysis, without loss of generality to the case where $\beta_1 + \beta_0 > 1$.[6]

The incentive-problem is real only if the principal optimally wants to implement $a_1$. We shall assume that for any optimal contract it is best for the principal to implement $a_1$. Now, the principal's problem is to choose, $t$, $s_1$, $s_0$, and $P$ to solve the program:

$$\min_{\substack{t, s_0, s_1 \in (\underline{t}, +\infty) \\ P \in [0,1]}} \quad (1-P)t + P(\beta_1 s_1 + (1-\beta_1)s_0 + C)$$

subject to:

$P_4$ (IR) $(1-P)V(t) + P[\beta_1 V(s_1) + (1-\beta_1)V(s_0) \geqq \overline{U} + a_1$

(IC) $P[V(s_1) - V(s_0)] \geqq \dfrac{a_1 - a_0}{\beta_1 + \beta_0 - 1}$

We shall solve $P_4$ in two stages. First we fix P and solve for the optimal transfers as functions of P : $\{t^*(P); s_1^*(P); s_0^*(P)\}$. Then we will determine the optimal probability of inspection, P.

When P is fixed we have a program that is equivalent to the cost-minimisation problem in Grossman and Hart (1983). As they noted, $P_4$ is not a convex program, however, assumptions A1 and A2 permit us to regard $v \equiv V(t)$; $v_1 \equiv V(s_1)$; $v_0 = V(s_0)$ as the control variables of the principal. $P_4$ is then rewritten as:

$$\min_{\{v, v_0, v_1 \in M\}} \quad (1-P)h(v) + p(\beta_1 h(v_1) + (1-\beta_1)h(v_0) + C)$$

subject to:

$P_5$ (IR) $(1-P)v + P(\beta_1 v_1 + (1-\beta_1)v_0) \geqq F$

(IC) $P(v_1 - v_0) \geqq k$

where $h \equiv V^{-1}(.)$; $k = \dfrac{a_1 - a_0}{\beta_1 + \beta_0 - 1}$ ; $F = \overline{U} + a_1$.

$P_5$ involves the minimisation of a convex function (h(.) is convex since V(.) is concave) subject to two linear constraints and from Proposition 1 in Grossman-Hart (1983) we know that an optimal solution to $P_5$ exists. A solution must satisfy the first-order conditions and is such that (IR) and (IC) are binding. From the first-order conditions we obtain the following equation:

$$(3) \qquad h'(v) = \beta_1 h'(v_1) + (1-\beta_1)h'(v_0)$$

And from the (IC) and (IR) constraints and (3) we can solve for v, $v_0$, $v_1$ to obtain:

$$(4) \qquad v_1 = \frac{F - v(1-P) + k(1-\beta_1)}{P}$$

$$(5) \qquad v_0 = \frac{F - v(1-P) - \beta_1 k}{P}$$

$$(6) \qquad h'(v) = \beta_1 h'\left[\frac{F-v(1-P)+k(1-\beta_1)}{P}\right] + (1-\beta_1)h'\left[\frac{F-\beta_1 k-(1-P)v}{P}\right]$$

Proposition 1: For any given $P \in (0,1]$ a unique solution v exists to (6).

Proof: The LHS of (6) is strictly increasing in v and the RHS is strictly decreasing in v. Furthermore, for any $P \in (0,1]$, for values of v close to $(F - \beta_1 k)$ the RHS of (6) is strictly greater than the LHS. Similarly, for v close to $F + (1-t_1)k$, the LHS is strictly greater than the RHS. It follows by the continuity of h'(.), that there must be a value v* that satisfies (6) for any given $P \in (0,1]$. This value is unique since the LHS is strictly increasing in v and the RHS strictly decreasing in v, for any $P \in (0,1]$.

Thus equation (6) defines an implicit function $v = v(P)$, so that we can write the solutions to $P_5$ as functions of P: $v_1(P)$; $v_0(P)$ and $v(P)$.

Proposition 2: a) for all $P \varepsilon (0,1]$ we have

$$F + k(1-\beta_1) > v(P) > F - \beta_1 k$$

b) $\lim_{P \to 0} v(P) = F + k(1-\beta_1)$

Proof: we have $\dfrac{F - v(1-P) + k(1-\beta_1)}{P} > \dfrac{F - v(1-P) - \beta_1 k}{P}$

since $k > 0$.

Now h(.) is strictly increasing, convex. This implies, from (6) and $0 < \beta_1 < 1$, that:

$$h'\left(\frac{F - v(1-P) + k(1-\beta_1)}{P}\right) > h'(v) > h'\left(\frac{F - v(1-P) - \beta_1 k}{P}\right)$$

thus,

$$\frac{F - v(1-P) + k(1-\beta_1)}{P} > v > \frac{F - v(1-P) - \beta_1 k}{P}$$

and

$$F + k(1-\beta_1) > v > F - \beta_1 k$$

This establishes a).

$$\lim_{P \to 0} h'\left(\frac{F + k(1-\beta_1) - (1-P)v(P)}{P}\right) = +\infty$$

unless, $\lim_{P \to 0} v(P) = F + k(1-\beta_1)$

(from a) we know that v(P) cannot be greater than $F + k(1-\beta_1)$).

Similarly, $$\lim_{P \to 0} \frac{F - \beta_1 k - (1-P)v(P)}{P} = -\infty .$$

Now h(.) is strictly convex increasing; thus $h'(-\infty) > 0$ and $h'(+\infty) = +\infty$. It follows that (6) can only be satisfied for all values of $P \in (0,1]$ if we have:

$$\lim_{P \to 0} v(P) = F + k(1-\beta_1)$$

This establishes b).

It follows from proposition 2, that the first-best outcome cannot be approximated here, unless $\beta_1=1$, which we have ruled out. The reason is that

$$\lim_{P\to 0} h(v(P)) > h(F) = V^{-1}(\overline{U}+a_1)$$

In other words, in the second-best contract wage costs are higher for the principal. This is not surprising in view of the fact that the agent must be compensated here for the risk of being punished when he is inspected. Nalebuff and Scharfstein (1985) have obtained an equivalent result in a model of self-selection. They show that if the tests to which agents are submitted are not perfectly accurate, then the first-best cannot be approximated.

The second important conclusion to be drawn from proposition 2 is that as P tends to zero the transfer $h(v(p))$, to the agent when the principal does not inspect, does not become very large. Now any optimal contract $c^* = \{v^*(P);\ v_1^*(P);\ v^*{}_0(P)\}$ must satisfy the equation:

$$h'(v^*(P)) = \beta_1 h'(v_1^*(P)) + (1-\beta_1)h'(v_0^*(P))$$

And proposition 2, tells us that

$$\lim_{P\to 0} h'(v^*(P)) = \lim_{P\to 0} \{\beta_1 h'(v_1^*(P)) + (1-\beta_1)h'(v_0^*(P))\}$$

$$= h'(F+k(1-\beta_1)) < +\infty$$

In other words, the expected wage, when inspection takes place, $\{\beta_1 . h(v_1(P)) + (1-\beta_1)h(v_0(P))\}$, is bounded above as P tends to zero.

We are now ready to move to the second stage of the principal's minimisation problem:

$$\min_{P\varepsilon(0,1]} \quad (1-P)h(v(P)) + P\{\beta_1 h(v_1(P)) + (1-\beta_1)h(v_0(P))\} + P.C$$

This can be rewritten as:

$$(8) \qquad \min_{P\varepsilon(0,1]} \qquad \phi(P) + P.C$$

And from (8) the following proposition follows:

<u>Proposition 3</u>: If the optimal probability of inspection is different from one, then it is a strictly decreasing function of the costs of inspection, C.

<u>Proof</u>: (9) $\phi(P_2^*) + P_2^* C_2 \leqq \phi(P_1^*) + P_1^* C_2$

(10) $\phi(P_1^*) + P_1^* C_1 \leqq \phi(P_2^*) + P_2^* C_1$

Adding (9) to (10) we obtain:

$$(P_1^* - P_2^*)\ (C_2 - C_1) \geqq 0$$

but $(C_2 - C_1) < 0$ hence $P_1^* \leqq P_2^*$.

Next, $\phi(.)$ is differentiable and if there is an interior solution $P^*$ to (8) (i.e. $P^* \neq 1$), such a solution must satisfy:

$$\phi'(P^*) = -C$$

It follows that if $P_1^*$ and $P_2^*$ are interior solutions then $P_1^* < P_2^*$.

Proposition 3 tells us that the higher the inspection costs, the lower will be the probability of inspection. Thus one may wonder whether the optimal probability of inspection will be arbitrarily close to zero for some sufficiently high inspection cost? If this turns out to be indeed the case then it follows that the principle of maximum deterrence would still hold (see the (IC) constraint in $P_4$). This would however be an uninteresting conclusion if it turns out that P is arbitrarily close to zero only if $C = +\infty$.

Proposition 4: If $\underline{t} > -\infty$, there exists $\bar{C} < +\infty$, such that if $C > \bar{C}$ a solution to $\min_{P\varepsilon(0,1]} \phi(P) + PC$ does not exist.

Proof: from proposition 2, we know that

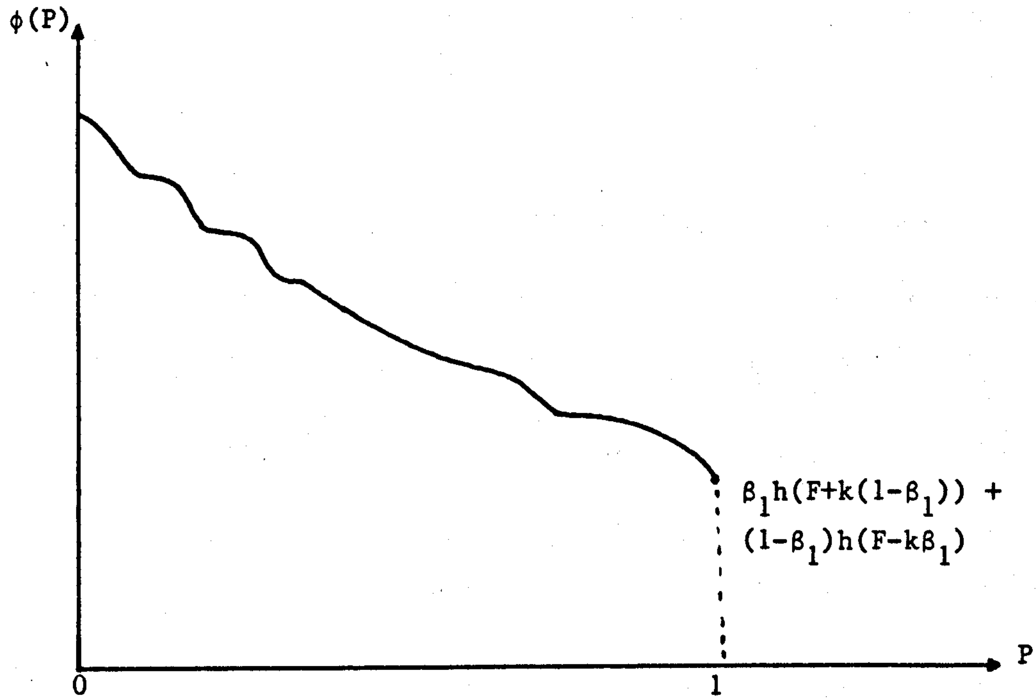
$$\lim_{P \to 0} \phi(P) = h(F + k(1-\beta_1))$$

Next, by the envelope theorem we have:

$$\phi'(P) = \beta_1 h(v_1(P)) + (1-\beta_1)h(v_0(P)) - h(v(P))$$

so that $\lim_{P \to 0} \phi'(P) > -\infty$.

Given the above information about $\phi(P)$, we obtain the following figure:




We know that $h(.) > \underline{t}$ and by assumption $\underline{t} > -\infty$, it follows that $\phi'(P) > -\infty$ for all $P \varepsilon (0,1]$.

Now define $\bar{C}$ such that:

$$\inf_{P\varepsilon(0,1]} \phi'(P) = -\bar{C}, \quad \text{then } \bar{C} < +\infty.$$

Also from proposition 3, if $C > \bar{C}$, we must have

$$\phi'(P) = -C < \inf_{P\varepsilon(0,1]} \phi'(P) = -\bar{C}$$

This is clearly not possible, so that if $C > \bar{C}$, a solution to

$\min_{P\varepsilon(0,1]} \phi(P) + PC$ does not exist.

In fact, for $C > \overline{C}$ we have an open-set problem similar to the one in section two. If the principal faces high fixed inspection costs, it may be optimal for him to inspect with probability P, arbitrarily close to zero.

There is a trade-off for the principal between facing high inspection costs, PC, or facing high expected wage costs, $\phi(P)$. If he lowers the probability of inspection, he must offer the agent a more risky inspection wage $\{\beta_1 h(v_1(p)) + (1-\beta_1)h(v_0(p)\}$, and since the agent is risk-averse this implies that he will have to pay the agent a higher expected wage. On the other hand, by lowering P, he lowers his expected inspection costs PC.

Now, as the variance of the inspection wage increases, the expected wage does not shoot off to infinity (this was established in proposition 2). The reason is that there are two counter-balancing effects, one of which dominating the other as P becomes small: on the one hand, the increased risk due to a more variable inspection wage increases the expected inspection wage, but on the other hand, there is a reduction in risk due to a smaller probability of inspection. The latter effect dominates the former as P tends to zero.

Observe that our result would no longer be true if $\underline{t} = -\infty$,, for then we have,

$$\lim_{P \to 0} \phi'(P) = -\infty$$

and it would not be optimal for the principal to impose arbitrarily high penalties on the agent, unless $C = +\infty$. This point was also noticed by Nalebuff and Scharfstein (1985).

So far we have not said anything about the type-one and type-two errors nor have we put restrictions on the degree of risk-aversion (relative or absolute) of the agent. In this respect, proposition 4 is very general. Nalebuff and Scharfstein (1985) show that if as $C \to +\infty$, the accuracy of the test becomes perfect (i.e. $\beta_1 \to 1$) then the first-best outcome can be approximated. This is also true in our model, since as $C \to +\infty$ and $\beta_1 \to 1$, we have $P^* \to 0$ and $\{\phi(P) + PC\} \to h(F) = V^{-1}(\overline{U}+a_1)$, (provided, of course that $P.C \to 0$). Moral-hazard models are quite different from self-selection models and it is remarkable that, as far as optimal inspection (or testing) contracts are concerned they yield identical conclusions.

One may wonder how robust proposition 4 is to changes in the model. Note first-of-all that it does not depend on the size of the type-one and type-two errors. The form of the utility-function of the agent, however, is important. For example, utility-functions of the HARA-family, considered in Baiman-Demski (1980) will not do, mainly because they do not satisfy the condition that $V(\underline{t}) = -\infty$, for some $\underline{t} > -\infty$.[5]

Next, the restriction to two actions, $A = \{a_0, a_1\}$, does not appear to be important. This is a conjecture since we have not generalised proposition 4 to the case of n actions. A priori, there is no reason however for this result to break down when the agent has access to more than two actions.

Also, it seems unlikely that when the principal is allowed to make the contract contingent on output ($q_i$), this result would fail to be true. Baiman and Demski (1980) have derived optimal inspection policies in this case, when the agent is assumed to have a utility-function over income of the HARA-family. Under the restrictions that the distribution-function over output satisfies the monotone-likelihood-ratio property, that the monitoring technology is conditionally independent of the production technology (i.e. $\tilde{a}$ is conditionally independent of $q_i$) and that penalties are bounded, they obtain a striking result. Namely that when the agent is relatively risk-averse the optimal investigation policy is to inspect with probability one whenever the profit-outcome, $q_i$, is less than or equal to some prespecified level $\bar{q}$, and not to inspect at all when profits are higher than $\bar{q}$. Baron and Besanko (1984) have obtained a similar result in a regulation model, where the planner can inspect ex-post the cost-realisation of the regulated firm, after observing the firm's performance.

This result strongly depends on the monotone-likelihood-ratio-property, since the lower the profit outcome the more likely it is under this assumption, that the agent was slacking, so that there is more to be gained from investigating. Probably the restriction on the form of the agent's utility-function is also important, as Holmstrom (1980) noted. For example, with utility-functions of the form specified here, the optimal contract will in general not be deterministic, as in Baiman and Demski, in order to save on the inspection costs. Indeed, for the same reasons as in our model, it may be optimal to inspect with an arbitrarily small probability.

More importantly, the assumption that the principal can precommit to a given inspection-policy seems to be crucial to obtain proposition 4. Thus, in the example considered here, when the principal wants to implement $a_1$ and cannot precommit himself to a given inspection policy, the principle of maximum deterrence breaks down: With no commitment, we must have

$$t \geqq \beta_1 s_1 + (1-\beta_1)s_0 + C$$

for inspection to take place, at all. At best, we have $t = \beta_1 s_1 + (1-\beta_1)s_0 + C$. Hence, no matter what probability of inspection the principal chooses, he will not be able to save on his inspection-cost, C. Previously, the main reason for increasing the size of the penalty was to save on expected inspection costs, P.C. Now, the principal will not be able to save on costs and he will increase the expected wage he has to pay to the agent, by raising the penalty for shirking. Thus, in this case the principle of maximum deterrence breaks down.

To conclude this section we want to point out another interpretation of the model considered here. Suppose that by paying a sufficiently high inspection cost the principal can observe the agent's action choice perfectly accurately, when he inspects, but that the agent "trembles" slightly in his choice of action. Then $(1-\beta_1)$ would be the agent's choice-error when he wanted to choose action $a_1$, and $(1-\beta_0)$ his choice-error when he wanted to choose action $a_0$. Formally, this problem is identical to the one considered in this section, so that the conclusion would be that when the agent "trembles" this does not necessarily imply that penalties will be bounded in an optimal contract.

## SECTION 4: CONCLUSION

The main purpose of this chapter was to examine the claim that the principle of maximum deterrence would be violated when there is a positive probability of punishing someone who is innocent. The conclusion reached is that this is not necessarily true. Proposition 4 demonstrates that in some cases it may be optimal to punish an agent who shirks as severly as possible, even if there is a risk of punishing someone who is innocent. To reach this conclusion it was important to assume that the principal could precommit himself ex-ante to a given investigation policy. We explain that the combination of inspection-errors and no-commitment possibilities for the principal is necessary in our model to obtain an outcome where penalties are bounded.

FOOTNOTES

1 See also Townsend (1979); Polinsky and Shavell (1979); Gale and Hellwig (1985).

2 In their paper the offense is double-parking. The objective is to avoid an outcome where everybody will double-park without preventing those for which the private benefit from double-parking exceeds the social cost.

3 The results obtained where can be generalised to utility functions of a more general form: $U(t,a) = G(a) + K(a)V(I)$ (see Grossman and Hart (1983)).

4 One may ask what it means that the principal can precommit himself to a given probability of inspection $P_i$?
The Principal has a randomisation device, which can be formalised as follows: Consider the interval $[a,b] \in R$, where $a<b$. Everytime the device is activated it produces an outcome $\theta \in [a,b]$. Assume that $\theta$ is uniformly distributed on $[a,b]$. The Principal determines a sub-interval $[a,b']$, (where $b' \in [a,b]$) when he chooses $P_i$. That is, $P_i$ is defined by:

$$P_i = \frac{b'-a}{b-a}$$

What is necessary for the principal to be able to commit himself to $P_i$, is that the randomisation device described above be public knowledge and that $\theta$ be publicly observable.

5 V(x) is a utility-function belonging to the HARA family if:

$$V(x) = \frac{(1-\gamma)}{\gamma}\left(\frac{\beta x}{1-\gamma} + \eta\right)^{\gamma}$$

where $\gamma \neq 1$, $\beta > 0$; $\eta = 1$ if $\gamma = +\infty$.

# CHAPTER 2

## "ENTRY-PREVENTION THROUGH CONTRACTS WITH CUSTOMERS"

### INTRODUCTION

Most of the literature on entry prevention deals with the case of two duopolists (the established firm and the potential entrant) who compete with each other to share a market, where one of the duopolists (the incumbent) has a first-move advantage.[1] This basic paradigm has been studied under various assumptions: about the strategy-space of the players; the information-structure of the game; and the time-horizon. Recently, the model has been enlarged to allow for several entrants, several incumbents, several markets and third parties.[2]

We propose here to extend the entry-prevention model in one other direction, which to our knowledge has not yet been formalised, namely, we consider whether optimal contracts between buyers and sellers deter entry and whether they are suboptimal from a Welfare point of view. It has been pointed out by many economists that contracts between buyers and sellers in intermediate-good industries may have significant entry-prevention effects and that such contracts may be bad from a Welfare point of view.[3]

On the other hand, it is a widespread opinion among antitrust practitioners that contracts between buyers and sellers are socially efficient.[4] There have been a number of antitrust cases involving exclusive dealing contracts and often the decision reached by the Judge has led to considerable controversy. One famous case, United States vs.

United Shoe Machinery Corporation, illustrates quite clearly the nature of the debate: The United Shoe Machinery Corporation controlled 85 percent of the shoe-machinery market and had developed a complex leasing system of its machines to shoe-manufacturers, a leasing system against which, it was thought, other machinery manufacturers would have difficulty competing. The Judge ruled that these leasing contracts were in violation of the Sherman act; his decision has been repeatedly criticised by leading antitrust experts (see Posner (1976) and Bork (1978)). The main argument against the decision has been expressed by Posner (1976, pp.203) as follows:

> "The point I particularly want to emphasize is that the customers of United would be unlikely to participate in a campaign to strengthen United's monopoly position without insisting on being compensated for the loss of alternative and less costly (because competitive) sources of supply".

Exactly the same point is made by Bork (1978, pp.140), who concludes that when we find exclusive dealing contracts in practice, then these contracts could not have been signed for entry-deterrence reasons.

Both authors are right in pointing out that the buyer is better off when there is entry and that he will tend to reject exclusive dealing contracts that reduce the likelihood of entry unless the seller compensates him by offering an advantageous deal. Nevertheless, we show that contracts between buyers and sellers will be signed for entry-prevention purposes.

When the buyer and the seller sign a contract they have a monopoly power over the entrant. They can jointly determine what fee the entrant must pay in order to be able to trade with the buyer; that is to say, if the buyer signs an exclusive contract with the seller and then trades with the entrant, he must pay damages to the seller. Thus he will only trade with the entrant if the latter charges a price which is lower than the seller's price minus the damages he pays to the seller. These damages, which are determined in the original contract (liquidated damages) act as an entry-fee the entrant must pay to the seller. We show that the buyer and the seller set this entry-fee in the same way that a monopoly would set its price, when it cannot observe the willingness to pay of its customers. Thus, the main reason for signing exclusive contracts, in our model, is to extract some of the surplus an entrant would get if he entered the seller's market.

These contracts are not socially optimal, for they sometimes block the entry of firms which may be more efficient than the incumbent seller. Entry is blocked because the contract imposes an entry-cost on potential competitors. This cost takes two different forms: an entrant must either wait until contracts expire or induce the customers to break their contrant with the incumbent by paying their liquidated damages.

The waiting cost is larger, other things being equal, the longer the contract. We are thus led to study the question of the optimal length of the contract. It is a well-known Principle in economics that if agents engage in mutually advantageous trade it is in their best interest to sign the longest possible contract. A long-term contract can always replicate what a sequence of short-term contracts achieves.

This principle, however, sharply contrasts empirical evidence: In practice most contracts are of an explicit finite duration. Many economists have been puzzled by this obvious discrepancy between the theory and empirical evidence, and several authors have attempted to provide an explanation for why contracts are of a finite duration; most notably Williamson (1975, 1979) and Harris-Holmstrom (1984).

We argue here that looking only at the length of the contract does not make sense. What is important is, to what extent the contract locks the parties into a relationship. If the parties were free to leave the relationship at any time and at no cost, then no matter how long the contract, signing a contract in this instance would not change anything with respect to the no-contract situation. Thus we are led to make the distinction between the nominal length of the contract (the length that is specified in the contract (optimally it is infinite)) and the effective length of the contract (the actual length that the parties expect the relationship to last at the time of signing). Liquidated damages constitute an implicit measure of the effective length of the contracts.

This Chapter is organised as follows:

Section 1 outlines the model in full detail. Section 2 looks at optimal contracts between a single customer and the incumbent when both parties to the contract know exactly the likelihood of entry. It is shown that optimal contracts never completely preclude entry, but that they reduce the likelihood of entry. It is also shown that the effective length of the optimal contract is inversely related with the probability of entry. Section 3 analyses optimal contracts when there is asymmetric information about the probability of entry. It is shown that under these circumstances both the effective and the nominal length of the contract may act as a signal of the probability of entry. Finally, Section 4 investigates how the possible signature of long term contracts by other customers may influence a typical customer to also sign a long term contract. It is shown, in particular, that even if before the signature of any contract the probability of entry is equal to one, the incumbent is still able to have all customers accept a contract where he charges the monopoly price. The reason is that the incumbent can play all his customers against each other.

Since the contracts we describe are particularly relevant for intermediate-good industries we have labelled the customers in our model as downstream firms and the incumbent as an upstream monopolist.

## SECTION I: THE MODEL

We consider a two-period model, where a single producer of an intermediate good, supplies one unit of that good to n downstream firms. The latter are all identical and are assumed to operate in isolated markets. They all have a reservation price of 1, for the unit of input purchased. The intermediate good producer, has zero unit costs of production. He faces a threat of entry into his market. We assume (mainly for simplicity) that there is no entry in each of the downstream firms' markets. We thereby rule out the possibility for an entrant to enter into both the upstream market and the downstream markets.

The way entry is thought of here is that there is an investor who faces many profit opportunities in different markets, and who cannot enter more than one market at a time. Define $\pi^A$ as the highest profits an investor can make by entering a different market than the intermediate good market and $\pi$ as the profits the entrant can get in the intermediate good market. We assume that $\pi^A$ is a random variable distributed on the interval $[\underline{\pi},\bar{\pi}]$, with density $f(\pi^A)$. Then, the probability of entry into the upstream market is:

$$\psi = \Pr(\pi > \pi^A)$$

We have attempted to model in a very simple way the view of the world where there are many investors at each period of time who try to invest their funds in the markets where they hope to get the highest returns. The distribution of profits over markets, however, changes stochastically over time. Therefore, entry into a given market may also be stochastic. In this story we implicitly assume that investors do not have an unlimited access to funds and/or that there are diminishing returns to managing more investment projects. If neither of these assumptions hold then investment will take place until the marginal return on the last investment project is equal to the interest-rate. Many good reasons have been given for why investors only have a limited access to funds (see for example, Stiglitz and Weiss (1981) or Williamson (1971)).

The profits $\pi$ are the post entry equilibrium profits to the entrant. We consider the following post entry game: If no long-term contracts have been signed in period 1 between the incumbent and the downstream firms, there will be Bertrand-competition between the entrant and the incumbent in period 2. The post entry equilibrium price then is equal to zero, if we assume that the entrant has unit costs of production that are less than or equal to the incumbent's unit costs. Then the entrant's profits $\pi$ are zero. As Dasgupta and Stiglitz (1984) have pointed out: "ex ante competition" is driven out by "ex post competition". The incumbent can then charge the monopoly price, P = 1, without facing any threat of entry.

We will assume that the entrant's unit costs of production are the same as the incumbent's, but that $\underline{\pi} < 0$. More specifically, we assume that $\underline{\pi} = -\theta$ and $\bar{\pi} = 1-\theta$, where $1 > \theta > 0$. So that $\pi^A$ is distributed on $[-\theta, 1-\theta]$. Then $\pi^A$ represents the opportunity cost of entry. This choice of support is made entirely for convenience and nothing crucial depends on this assumption, as will become clear later.

Several interpretations can be given for the entrant's opportunity cost and we do not want to be more specific at this point. We wish to emphasize, however, that entry will only occur if $\pi^A$ is negative; in other words, if the entrant is in some sense more efficient than the incumbent.

For simplicity, we assume that $\pi^A$ is uniformly distributed on $[-\theta, 1-\theta]$,[5] so that when no long term contracts have been signed between the incumbent and the downstream firms the probability of entry is:

$$\psi = \int_{-\theta}^{0} f(x)dx = \theta$$

Furthermore, it is assumed that the entrant has complete information over the incumbent's cost function and that he knows the demand the incumbent faces. Similarly, the incumbent knows everything about the entrant. We also assume throughout sections 2 and 4 that the downstream firms know everything about the potential entrant so that there is symmetric information about the probability of entry.[6]

We will now describe how the incumbent can change the post-entry game by signing long term contracts with the downstream firms. The idea is that by manipulating the post entry game, in period 2, the incumbent can change the probability of entry. In section 2, we consider the case where there is only one downstream firm, so that we have a "chain of monopoly" problem with potential entry into the upstream market. In section 4 we study the negotiation game between the incumbent and several downstream firms.

The sequence of moves in the negotiation game between the incumbent and a single downstream firm is as follows: In period 1 the upstream firm makes the first move by proposing a long-term contract to the downstream firm. The latter then can only accept or reject the contract offered. In other words, the incumbent "sets the contract". If the downstream firm rejects the offer then, the incumbent charges the monopoly price P = 1, if entry does not occur, otherwise the entrant supplies the intermediate good at price, P = 0. In other words, both parties agree to leave "options open" and to sign a short term contract in period 2, which specifies that, P = 1, if there is no entry and, P = 0, if there is entry. If the downstream firm accepts the offer, the post entry game is changed and the probability of entry is different.[7]

This is a very simple bargaining game, where the incumbent has all the bargaining power. He has all the bargaining power because we assume that he makes the first offer, that the downstream firm makes no counter-offers and finally that if the incumbent's offer is rejected he does not make another offer. Such a negotiation game would make the best sense in a situation where there are many downstream firms and where it is costly for the incumbent to make an offer; therefore we see this bargaining solution as particularly relevant to the case analysed in section 4, where there are many downstream firms.

The virtue of the solution adopted above lies in its simplicity and none of the results obtained in sections 2 and 3 depend on it: What is important in this game is that the upstream firm and the downstream firms can get together and use their first-move advantage to extract some of the expected surplus of the entrant. This will become clear in section 2.

## SECTION 2: WHAT IS THE FORM OF THE OPTIMAL CONTRACT?

We assume in this section that there is only one downstream firm. First, we will define the set of feasible long term contracts between the incumbent and the downstream firm. Having done that we will characterize the optimal feasible contracts for the incumbent.

In general a long term contract between the incumbent and the buyer can be quite complicated; It may be contingent on the event of entry and on the realisation of the entrant's opportunity cost, $\pi^A$. It may also be contingent on the contract the entrant offers to the buyer in period 2. Finally it may include side payments, which are independent of whether trade between the incumbent and the downstream firm takes place or not.

To simplify matters we will make several assumptions which do restrict the set of feasible contracts:

1) the realisation of the entrant's opportunity cost, $\pi^A$, is neither observable by the incumbent and the downstream firm nor verifiable by a court. This assumption rules out the possibility of writing contracts contingent on $\pi^A$,
2) because of transaction costs and because the entrant's offer may be difficult to verify, the contract will not be contingent on the entrant's contract offer in period 2,[8]
3) the incumbent and the downstream firm can precommit not to renegotiate the contract in period 2.

The last assumption is the strongest and we will indicate wherever it is relevant how our results change if we allow for renegotiation. Given the above assumptions we can define a long term contract, c, to be a collection of four prices $\{P, P^e, P_o, P_o^e\}$, where:

$P$ = the price the buyer must pay if he buys one unit of the good produced by the incumbent and there is no entry into the upstream market.

$P^e$ = the price the buyer must pay if he buys one unit from the incumbent and there is entry into the upstream market.

$P_0$ = the price the downstream firm must pay to the incumbent if it does not buy the unit of input and there is no entry.

$P_0^e$ = the price the downstream firm must pay to the incumbent if it does not buy the unit of input and there is entry.

Notice that we have not included any side-payments in the contract. In fact we have verified that allowing for side-payments would not modify any of our results and it would only make the analysis more cumbersome.[9]

Note also that the following long term contract, $C = \{P_0 = P_o^e = 0\ ;\ P^e = 0;\ P = 1\}$ is equivalent to signing a short-term contract in Period 2. Thus, there is a long-term contract that can replicate the outcome that a sequence of short-term contracts achieves.

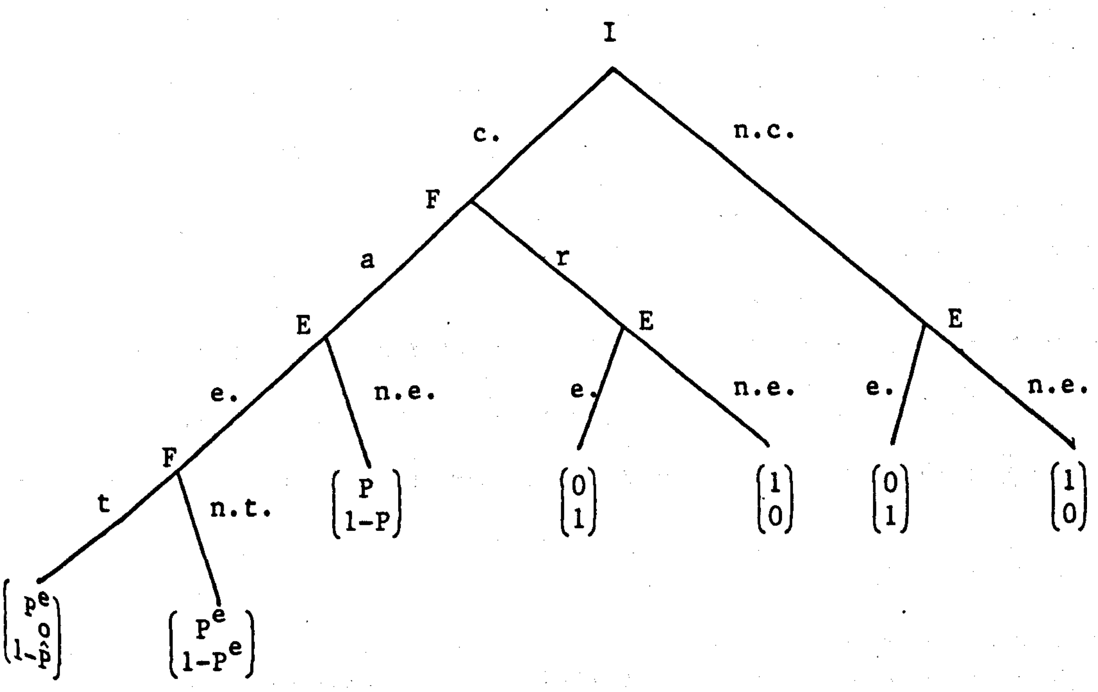
We will now show what the form of the optimal contract is when:

1) both the incumbent and the downstream firm are risk neutral.
2) the downstream firm is risk averse and the incumbent is risk neutral.

3) the incumbent is risk averse and the downstream firm is risk neutral.

The game between the incumbent, the downstream firm and the entrant is represented in the game tree below. We will adopt the following notation:

First the players are labelled by: I ≡ incumbent; F ≡ downstream-firm; E ≡ entrant. Second, the moves of the players are labelled as follows: For the incumbent, c ≡ contract offer; n.c. ≡ no contract offer. For the entrant, e ≡ entry; n.e. ≡ no entry. For the downstream firm: a ≡ accept; r ≡ reject; t ≡ trade with the entrant; n.t. ≡ not trade with the entrant.



The first entry in the column vectors represents the incumbent's payoff and the second, the downstream firm's payoff. We explain the payoffs as follows: If no long term contract is signed (this arises either because no contract was offered or because the offer was rejected) all options are left open and in period 2 the incumbent either charges $P = 1$, if no entry takes place, or $P = 0$, when there is entry. The downstream firm trades with the incumbent when there is no entry and otherwise it trades with the entrant at the Bertrand equilibrium price, $P = 0$.

If a long term contract is signed and no entry takes place then the downstream firm has the option of buying or not buying the unit of input from the incumbent. So long as $P < 1$, it will always buy from the incumbent. One can think of contracts where $P > 1$ and $(P-P_0) > 1$; in this case the downstream firm will refuse to buy and will pay $P_0$ to the incumbent. We exclude contracts where $P > 1$ so that it is not necessary to specify $P_0 > 0$. It will become clear below that there is no loss of generality in doing so.

If a long term contract is signed and entry takes place, F has the choice of either trading with the incumbent or with the entrant. If it trades with the latter its payoff is $1 - \hat{P}$, where $\hat{P}$ represents the sum of the entrant's price, $\tilde{P}$, and the price, $P_0^e$; so that $\hat{P} = \tilde{P} + P_0^e$. Clearly, F will only trade with the entrant if $1 - \hat{P} > 1 - P^e$.

Finally, the probability of entry in period 2 is $\psi$ if no long term contract is signed in period 1 and $\psi'$ when a long term contract has been signed.

We define $\psi'$ as follows:

The downstream firm is indifferent between trading with the entrant or the incumbent when $\hat{P} = P^e$, hence the most the entrant can hope to get is $P^e - P_0^e$ so that $\psi'$ is defined as:

$$(2.1) \qquad \psi' = \Pr(P^e - P_0^e > \pi^A)$$

$\pi^A$ is uniformly distributed on $[-\theta, 1-\theta]$ so that,

$$(2.2) \qquad \psi' = \int_{-\theta}^{P^e - P_0^e} f(x)dx = P^e - P_0^e + \theta$$

Clearly, whenever $P^e \neq P_0^e$, we have $\psi' \neq \psi$.

## I. The Downstream Firm and the Incumbent are Risk Neutral

Suppose that the incumbent offers a contract such that if it is accepted by the downstream firm, entry will occur with probability $\psi' = 0$; that is, once the contract is signed entry is blocked. For example if the incumbent offers a contract where $P_0^e = +\infty$, entry will be blocked once the contract is signed, for the downstream firm would always trade with the incumbent.[10] Then the downstream firm's payoff is $1 - P$ and it will accept such a contract only if:

$$(2.3) \qquad 1 - P > \psi.1 + (1 - \psi).0$$

The RHS of (2.3) represents the downstream firm's expected payoff when it rejects the offer. When (2.3) holds with equality F is indifferent between accepting and rejecting. If it accepts the offer we have $P = 1 - \psi$, and P is equal to the expected payoff to the incumbent if F rejects. So when (2.3) holds with equality both firms are indifferent between signing a contract and not signing one. We conclude that there are no gains to signing such a long term contract.

We will now show that there are contracts such that there is a positive gain to signing them. Consider a contract such that if it is accepted by F, $\psi' > 0$. Then the downstream firm's payoff is $\psi'(1 - P^e) + (1 - \psi')(1 - P)$, if it <u>always</u> trades with the incumbent and $\psi'(1 - \hat{P}) + (1 - \psi')(1 - P)$ if it trades with the entrant when entry takes place.

Hence, if the entrant offers a price $\hat{P} = P^e$, he will be able to trade with F and he need not offer a lower price in order to attract F (we assume that when F is indifferent between trading with I or E it will trade with E). Thus the expected payoff to F when it signs a contract is: $\psi'(1 - P^e) + (1 - \psi')(1 - P)$. In order for F to accept the contract we must have:

$$\psi'(1 - P^e) + (1 - \psi')(1 - P) \geq \psi.1 + (1 - \psi).0 \tag{2.4}$$

From (2.2), (2.4) can be rewritten as

$$(P^e - P_0^e + \theta)(P - P^e) + 1 - P \geq \theta \tag{2.5}$$

Now the problem for the incumbent is to choose $c \in C$, to maximise his expected payoff subject to (2.5).

So the incumbent solves:

$$\left|\begin{array}{ll} \max\limits_{\{P_0^e, P, P^e\}} & (P^e - P_0^e + \theta)(P_0^e - P) + P \\ \\ \text{subject to:} & (P^e - P_0^e + \theta)(P - P^e) + 1 - P \geq 0 \end{array}\right. \tag{2.6}$$

At the optimum the constraint must hold with equality so that we have:

$$(2.7) \qquad P = (P^e - P_0^e + \theta)(P - P^e) + 1 - \theta$$

Substituting for P in the objective function we obtain the unconstrained problem:

$$(2.8) \qquad \max_{\{P_o^e, P^e\}} (P^e - P_o^e + \theta)(P_o^e - P^e) + 1 - \theta$$

It is easy to see from (2.8) that any optimal contract must be such that:

$$(2.9) \qquad P_o^e = P^e + \frac{\theta}{2}$$

Furthermore from the constraint in (2.6) we must have:

$$(2.10) \qquad P(1 - \frac{\theta}{2}) = 1 - \theta - P^e . \frac{\theta}{2}$$

Equations (2.9) and (2.10) characterise the set of optimal contracts. Any contract that satisfies these two equations yields an expected payoff of $\frac{\theta^2}{4} + 1 - \theta$, to the incumbent and a payoff to the downstream firm of $\theta$.

Note that when $P = P^e$ the optimal contract is such that, $P_o^{e*} = 1 - \frac{\theta}{2}$ and $P^* = P^{e*} = 1 - \theta$.[11]

We summarise these results in the following proposition:

Proposition 2.1:

(1) The set of optimal contracts for the incumbent is characterised by the equations: (i) $P_o^e = P^e + \frac{\theta}{2}$

(ii) $P(1 - \frac{\theta}{2}) = 1 - \theta - P^e . \frac{\theta}{2}$

(2) The probability of entry when an optimal contract is signed is given by $\psi'^{*} = \frac{\theta}{2}$. It is strictly positive and strictly less than the probability of entry when no long term contract is signed ($\psi = \theta$).[12]

(3) The incumbent's expected payoff when an optimal long term contract is signed is: $\frac{\theta^2}{4} + 1 - \theta$. When he does not sign a contract his expected payoff is $1 - \theta$. So the incumbent is strictly better off when he signs a long term contract. Furthermore he is indifferent between signing a contract where the price is independent of the event of entry ($P = P^e$) and signing a contract where the price is a function of entry ($P \neq P^e$).

Proposition 2.1 can be interpreted as follows: By choosing $P_o^e$ appropriately the upstream monopolist can extract a fraction $\frac{\theta^2}{4}$ of the entrant's expected surplus. In fact, if the contract could be made contingent on the entrant's opportunity cost $\pi^A$ the incumbent would be able to extract all of he entrant's surplus. To choose $P_o^e$ appropriately involves choosing $P_o^e$ to be strictly positive and also not to be too high in order to have a probability of entry $\psi'$, which is not too low.

This idea that the incumbent and the downstream firm can get together and extract some of the surplus of the entrant is very general: it does not depend on the particular bargaining solution adopted here and it does not depend on the type of distribution function over profit alternatives which we have assumed.

In Appendix 1 we show that all the qualitative conclusions obtained in Proposition 2.1 hold for any arbitrary continuous distribution function which has a support that is bounded below.

Diamond and Maskin (1979) have obtained a similar result in the context of a model of search with breach of contract, where they assume the Nash bargaining solution. They explain that: "once shares (of the surplus) in new deals become tied to previous damage payments, a pair of individuals in a contract has some monopoly power over potential partners", "Damages ($P^e_o$ in our model) cannot be raised without limit because higher damages mean that breach is less likely and only when breach occurs can monopoly power be exerted".

Remark 1: We have established in Proposition 2.1 that there is a whole class of optimal contracts. There are, however, good reasons for restricting the set of feasible contracts to the class of contracts that satisfy: $P = P^e$. Indeed there are two major problems in specifying a price which is contingent on entry. First of all, "entry" may be a very complicated event to describe. This is the case when we allow for the possibility that an entrant may enter with a non-homogeneous good. The question arises then of what commodities qualify as "entrants"? Even if a list of such commodities can be defined and written into the contract an entrant would always have an incentive to choose a slightly differentiated good which is not included in the list whenever, $P > P^e$, for then he can charge P instead of $P^e$. And if $P^e > P$, there would be an incentive for the incumbent to claim that entry has occurred, whenever there is an ambiguity about the event of entry. In short, the event of entry may be difficult to observe, let alone to verify. Then there is a strong case for restricting oneself to contracts where $P = P^e$.

Secondly, even if the entrant could only produce the same commodity as the incumbent the following problem arises: Suppose that $P > P^e$, then the downstream firm has an incentive to bribe someone to "enter" only to force the incumbent to lower its price. Similarly, when $P < P^e$, the incumbent has the same incentive, so that the parties could only agree on a contract where $P = P^e$.

Because of these two problems we believe that contracts where $P \neq P^e$ are seldom well defined. This is a matter of opinion, however, and in the end it is an empirical matter to see where contracts with a price contingent on entry do apply. We believe, nevertheless, that results which entirely rely on the property that $P \neq P^e$, should be taken carefully.

Remark 2: The results established in Proposition 2.1 have been shown to be true only in a restricted set of contracts and the reader may wonder whether these results would still hold if we considered more general contracts. It turns out that what is crucial is the non-observability of $\pi^A$. Given that $\pi^A$ is not observable the incumbent cannot do better than signing an optimal contract as described in Proposition 2.1. To see this, it suffices to look at the incumbent's problem like an information revelation problem: The incumbent chooses the optimal contract subject to the constraint that it is incentive-compatible for the entrant to reveal his true type, $\pi^A$, Then the incentive compatibility- constraint implies that $P_o^e$ must be independent of $\pi^A$. Thus, the incumbent cannot extract more than $\frac{\theta}{4}$ from the entrant. Also, he cannot extract more than $1-\theta$ from the buyer.

Remark 3: Notice that the effective length of the contract is here given by:

$$\ell(c^*) \equiv 1.\frac{\theta}{2} + (1 - \frac{\theta}{2}).2$$

$$\equiv (2 - \frac{\theta}{2})$$

Thus, the effective length of the contract is between 1 and 2 periods and it is a decreasing function of the probability of entry, $\theta$. In other words, the more likely entry is, the less the buyer wants to be locked in the relationship with the incumbent.

It is useful for later discussions to define the following partition on the set of feasible long term contracts, $C = \{(P_o^e, P, P^e)/P \leq 1\}$.

Let $C_A$ and $C_B$ be two subsets of C defined by:

$$C_A = \{(P_o^e, P, P^e) \in C/P_o^e > P^e\}$$

$$C_B = \{(P_o^e, P, P^e) \in C/P_o^e \leq P^e\}$$

Then for all contracts in $C_A$ we have $\psi' < \psi$ and for all contracts in $C_B$, $\psi' \geq \psi$. Furthermore, when both parties to the contract are risk neutral the incumbent weakly prefers short term contracts to long term contracts which are in $C_B$: when he signs a short term contract his payoff is 1 - θ. When he signs a long term contract which is in $C_B$ he gets (from (2.8)) $(P^e - P^e_o + \theta)(P^e_o - P^e) + 1 - \theta$. This latter expression is not greater than 1 - θ, whenever $P^e_o < P^e$.

Proposition 2.1 tells us that an optimal contract always belongs to $C_A$. On the other hand we will show in the next section that when the incumbent is better informed about entry than his customer then in most cases the optimal second best contract is constrained to be in $C_B$.

The above partition also highlights the fact that optimal long term contracts are less "flexible" than short term contracts. Jones and Ostroy (1984) explain that "flexibility is a property of initial positions. It refers to the cost or possibility of moving to various second period positions" (pp.16). In our model a contract $c \in C$ is more or less "flexible" depending on whether it increases or decreases the likelihood of entry; that is, depending on whether it increases or decreases the entrant's "cost of moving to various second period positions" (here the cost is $P^e_o - P^e$).

It is not surprising then that optimal contracts are not Pareto optimal. Efficiency would require that entry ought to occur whenever the entrant is more efficient than the incumbent. Hence a Pareto optimal outcome is achieved when the incumbent only signs a short term contract in period 2 with the buyer or when a long term contract $c \in C_B$ is signed such that $\psi' = \psi$.[13]

Finally, the model in this section may provide an alternative foundation of the "limit pricing" theory of entry prevention to the one given by Milgrom and Roberts (1982), in situations where contracts contingent on entry are not feasible. In these situations our model predicts that the pre-entry price is the same as the post entry price and that the price agreed upon in the contract is lower than the monopoly price. Also, entry is prevented (to some extent) when a long term contract is signed. One can find all these features in the limit pricing literature (see for example Bain (1956) and Modigliani (1958)). There is an important qualification, however. We have assumed at the beginning of this section that there will be no renegotiation of long term contracts in period 2 but if the incumbent cannot precommit to not renegotiate, situations may arise in period 2 where renegotiation is profitable. For example suppose that $P^e = P = 1 - \theta$, $P_o^e = 1 - \frac{\theta}{2}$, and that a firm enters the market with opportunity cost $\pi^A = -\frac{\theta}{4}$. Then if the entrant wants to attract the downstream firm it has to charge a price, $P^e - P_o^e = -\frac{\theta}{2}$ and it will make negative profits. Thus, the best offer an entrant can make is $P^e - P_o^e = -\frac{\theta}{4}$. If there is no renegotiation the incumbent serves the buyer and gets $1 - \theta$, but through renegotiation the incumbent may be able to get part of the entrant's surplus, if he lowers $P_o^e = 1 - \frac{\theta}{2}$ to $P_o^{e\prime} < 1 - \frac{3\theta}{4}$.

Thus, renegotiation may be profitable and if the entrant anticipates that renegotiation will take place when he enters there may be a higher probability of entry than $\psi' = \frac{\theta}{2}$. Notice, however, that renegotiation will only change $P_o^e$ and the probability of entry but not P. It will also affect the size of the expected surplus that the incumbent can hope to extract from the entrant; a priori, however, it is not clear whether it will increase or decrease the size of the expected surplus. A careful discussion of renegotiation is beyond the scope of this paper and all that can be said at this stage is that renegotiation is likely to reduce the rigidity of long term contracts but not to eliminate it completely.

## II. The Downstream Firm is Risk Averse and the Incumbent is Risk Neutral

When one or both of the parties to the contract is risk averse then none of the conclusions reached so far are modified if the risk averse party seeks insurance with an insurance company. Sometimes, however, it may not be possible for the parties to the contract to seek outside insurance. The reasons that are usually given for this impossibility are mainly, unobservability of the state of nature by the insurance company and collusion by the parties to the contract against the insurance company. There may also be moral hazard problems, but these are not relevant here.[14] While observability may be an important issue in labour contracts when shocks are idiosyncratic, it seems much less of a problem here (so long as the event of entry is well defined), since entry into a market is an easily observable event. Thus, the usual arguments given for why outside insurance may be impossible are much less compelling here. We do not have any other explanations for why outside insurance ought to be assumed away and we present the insurance aspect of the contract mainly for the sake of completeness.

We assume that the owner of the downstream firm has a Von-Neumann-Morgenstern utility function $U(.)$, where $U'(.) > 0$ and $U''(.) < 0$.

Now the downstream firm accepts a contract offer when:

$$(2.11)\qquad (P^e - P^e_o + \theta).U(1 - P^e) + (1 - P^e + P^e_o - \theta).U(1 - P) > \theta.U(1) + (1 - \theta).U(0)$$

This inequality can be rewritten as follows:

$$(2.12)\qquad U((P^e - P^e_o + \theta)(1 - P^e) + (1 - P^e + P^e_o - \theta)(1 - P) - r_{12}) > U(\theta - r)$$

where $r_{12}$ and r are the risk premiums corresponding to the two lotteries:

$$L_{12} = \begin{bmatrix} \text{outcome} & \text{probability} \\ 1 - P^e & \psi' = P^e - P^e_o + \theta \\ 1 - P & 1 - \psi' \end{bmatrix}$$

and

$$L = \begin{bmatrix} 1 & \psi = \theta \\ 0 & 1 - \psi = 1 - \theta \end{bmatrix}$$

Inequality (2.12) is equivalent to:

$$(2.13)\qquad (P^e - P^e_o + \theta)(1 - P^e) + (1 - P^e + P^e_o - \theta)(1 - P) - r_{12} > \theta - r.$$

And the incumbent now chooses $P^e_o$, P and $P^e$ to maximise his expected payoff subject to (2.13). We solve for the optimal contract and establish the following proposition:

Proposition 2.2: If the downstream firm is risk averse and the incumbent is risk neutral the unique optimal contract is characterised by:

(1) $P = P^e = 1 - \theta + r$

(2) $P^e_o = 1 + r - \frac{\theta}{2}$

And the incumbent's expected payoff under the optimal contract is:

$$\frac{\theta^2}{4} + 1 - \theta + r.$$

Proof: (see Appendix).

Compared to the case where both parties were risk-neutral we have reduced the set of optimal contracts to a unique contract.

III) The Downstream Firm is Risk Neutral and the Incumbent is Risk Averse

We assume here that the owner of the upstream monopoly has a Von-Neumann-Morgenstern utility function V(.), such that, V'(.) > 0 and V"(.) < 0. We are able to establish the following proposition:

Proposition 2.3: The unique optimal contract for the incumbent is characterised by:

(1) $P^e_o = P = 1 - \theta + \frac{\theta^2}{4}$

(2) $P = P^e + \frac{\theta}{2}$

Proof: (see Appendix).

## SECTION 3: ASYMMETRIC INFORMATION ABOUT THE LIKELIHOOD OF ENTRY

In sections 1 and 2 it was assumed that both the incumbent and the downstream firm know the true probability of entry. This is not always realistic and one would expect that usually the incumbent is better informed about the likelihood of entry than his customers. For example, if the incumbent is a high-tech-firm and is the only one to have the know-how to produce a given intermediate good, then it is likely to be much better informed than its customers about the possibility that a potential competitor will be able to acquire this know-how and thus produce the intermediate good. Hence, in this section we assume that the incumbent has some private information about the likelihood of entry.[15]

To keep the analysis simple we assume that there is only one downstream firm, that both the incumbent and his customer are risk-neutral, and that the probability of entry is either "high" or "low". The incumbent knows the true probability of entry but the downstream firm does not. Since the price of the intermediate good specified in the contract is inversely related to the probability of entry, the incumbent has an incentive to lie whenever he faces a "high" probability of entry. Clearly, the downstream firm will not believe that the incumbent faces a "low" probability of entry unless the latter designs a mechanism such that his private information is always truthfully revealed.

The situation described here is akin to what Myerson (1983) calls "mechanism design by an informed Principal". Here the incumbent is the informed principal and the downstream firm, the agent. In fact our model is a special case of the model of the informed Principal by Maskin and

Tirole (1985).

The purpose of this section is to study how asymmetric information imposes restrictions on the form of the contract the incumbent can offer to his customers. Three main conclusions are reached:

Firstly, if contracts where the price is contingent on the event of entry are feasible then asymmetric information puts no restrictions on the form of the contract: In other words the symmetric information optimal contract can always be implemented.

Secondly, when the price is not contingent on entry ($P = P^e$), the incumbent signals his type through an appropriate choice of the compensation price, $P^e_o$. The basic idea is that, in order to signal that the probability of entry is "low", the incumbent offers a very low $P^e_o$. That is, the incumbent signals that entry is unlikely by not making entry too costly for the entrant, or equivalently by restricting himself to low profits when there is entry. Thus asymmetric information puts restrictions on $P^e_o$, such that entry is more likely than under symmetric information. It follows that asymmetric information is welfare improving. What is meant in this last sentence is that if one compares the aggregate expected surplus of the incumbent, the downstream firm and the entrant in the symmetric information case and in the asymmetric information case, then the aggregate expected surplus in the latter case is higher than in the former. That is not to say that a planner would be able to manipulate the distribution of information in the economy in order to achieve a more efficient outcome.

Finally, when the difference between the "high" and the "low" probability of entry is sufficiently large the second-best long-term contracts are in the class $C_B$. Then the incumbent is indifferent between signing a long-term contract in the class $C_B$ and signing a short-term contract. Thus, when he is indifferent, the incumbent can signal the probability of entry through the nominal length of the contract.

We define $\bar{\psi}$ to be the "high" probability of entry and $\underline{\psi}$ to be the "low" probability of entry: $0 < \underline{\psi} < \bar{\psi} < 1$. In the language of Myerson we have an informed principal with two types: $\bar{\psi},\underline{\psi}$. The principal's strategy space is defined by the set of contracts, $C = \{c = (P_o^e, P, P^e)/P \leq 1\}$. The downstream firm, on the other hand, has two actions: "accept" and "reject", so that its action set is $A = \{a,r\}$. The downstream firm has prior beliefs over the incumbent's type: let m be the probability that the incumbent is of type $\bar{\psi}$. We assume that m is common knowledge.

As in the previous sections we assume that the entrant's opportunity cost $\pi^A$ is uniformly distributed on some interval. When the probability of entry is "high" we assume that $\pi^A$ is distributed on $[-\theta,1-\theta]$. When the probability of entry is "low" we assume that $\pi^A$ is distributed on $[-k\theta,1-k\theta]$, where $0 < k < 1$. Thus $\bar{\psi} = \theta$ and $\underline{\psi} = k\theta$.

When the incumbent makes a contract offer c this offer may or may not reveal information about his type to the downstream firm: we define $\beta(c) \equiv \Pr(\psi = \bar{\psi}/c)$, as the probability that the incumbent is of type $\bar{\psi}$ given that contract c has been proposed.

Let $U(c,a,\bar{\psi})$ be the payoff to the downstream firm if it accepts contract c and the incumbent's type is $\bar{\psi}$. Then we define the downstream firm's expected utility when a contract c is offered as:

$$\beta(c)U(c,a,\bar{\psi}) + (1 - \beta(c))U(c,a,\underline{\psi})$$

and it will accept contract c if and only if the following inequality is satisfied:

$$\beta(c)U(c,a,\bar{\psi}) + (1 - \beta(c))U(c,a,\underline{\psi}) > \beta(c)\bar{\psi} + (1 - \beta(c))\underline{\psi} \tag{3.1}$$

We will first consider the case where the incumbent offers contracts that are contingent on the event of entry: $c = (P^e_o, P, P^e)$. We know from section 2 that when a long term contract has been signed the probability of entry is $\bar{\psi}' = \theta + P^e - P^e_o$, provided that $P^e_o < P^e + \theta$ otherwise $\bar{\psi}' = 0$. Then we may write:

$$U(c,a,\bar{\psi}) \equiv (\theta + P^e - P^e_o)(P - P^e) + 1 - P \quad \text{and}$$

$$U(c,a,\underline{\psi}) \equiv (k\theta + P^e - P^e_o)(P - P^e) + 1 - P.$$

Inequality (3.1) is then rewritten as:

$$[\beta(c)(\theta + P^e - P^e_o) + (1 - \beta(c))(k\theta + P^e - P^e_o)](P - P^e) + 1 - P > \beta(c)\theta + (1 - \beta(c))k\theta \;. \tag{3.2}$$

Next we define the incumbent's payoff function when he is respectively of type $\underline{\psi}$ and $\bar{\psi}$ as:

$$V(c,\bar{\psi}) = (\theta + P^e - P^e_o)(P^e_o - P) + P$$

$$V(c,\underline{\psi}) = (k\theta + P^e - P^e_o)(P^e_o - P) + P$$

Now the incumbent's maximisation problem, when he is of type $\underline{\psi}$ say, is:

$$\left| \begin{array}{l} \max\limits_{c \in C} \quad V(c,\underline{\psi}) = (k\theta + P^e - P^e_o)(P^e_o - P) + P \\ \\ \text{subject to (3.2)} \end{array} \right.$$

Thus we obtain the following unconstrained maximisation problem:

$$\max_{c \in C} \quad V^*(c,\underline{\psi},\beta(c)) \equiv (k\theta + P^e - P^e_o)(P^e_o - P) + [\beta(c)(\theta + P^e - P^e_o) +$$
$$(1 - \beta(c))(k\theta + P^e - P^e_o)].(P - P^e) + 1 - \beta(c)\theta - (1 - \beta(c))k\theta.$$

When the incumbent's type is $\bar{\psi}$ we can similarly define $V^*(c,\bar{\psi},\beta(c))$.

Now, we may have two types of solutions to this problem: either the optimal contract is separating (i.e. $\beta(c) = 1$ or $\beta(c) = 0$) or the optimal contract is pooling or semi-separating (i.e. $0 < \beta(c) < 1$).

Two contracts $(\underline{c},\bar{c})$ are defined to be separating when the following inequalities are satisfied:

$$(3.3) \quad \left| \begin{array}{ll} V^*(\bar{c},\bar{\psi},\beta = 1) > V^*(c',\bar{\psi},\beta(c')) & \text{for all } c' \in C. \\ V^*(\underline{c},\underline{\psi},\beta = 0) > V^*(c',\underline{\psi},\beta(c')) & \text{for all } c' \in C. \end{array} \right.$$

We will now show that there exist two contracts $(\underline{c}^*, \bar{c}^*)$ which form a separating perfect Bayesian equilibrium and which are respectively optimal symmetric information (O.S.I) contracts for the $\underline{\psi}$-type and the $\bar{\psi}$-type incumbent. We know from Proposition 2.1 that an OSI-contract for the $\bar{\psi}$-type is such that:

$$\text{(I)} \quad \begin{cases} P_o^e = P^e + \dfrac{\theta}{2} \\ P\left(1 - \dfrac{\theta}{2}\right) = 1 - \theta - P^e \dfrac{\theta}{2} \end{cases}$$

Similarly an OSI-contract for the $\underline{\psi}$-type is such that:

$$\text{(II)} \quad \begin{cases} P_o^e = P^e + \dfrac{k\theta}{2} \\ P\left(1 - \dfrac{k\theta}{2}\right) = 1 - k\theta - P^e \dfrac{k\theta}{2} \end{cases}$$

When a contract c' yields beliefs $\beta(c') = 1$ and when $\underline{c}$ and $\bar{c}$ are OSI-contracts the inequalities (3.3) are rewritten:

$$\text{(3.4)} \quad \begin{aligned} &\frac{\theta^2}{4} + 1 - \theta > (\theta + P^e - P_o^e)(P_o^e - P) + (\theta + P^e - P_o^e)(P - P^e) + 1 - \theta \\ &\frac{k^2\theta^2}{4} + 1 - k\theta > (k\theta + P^e - P_o^e)(P_o^e - P) + (\theta + P^e - P_o^e).(P - P^e) \\ &\qquad\qquad + 1 - \theta \end{aligned}$$

Let $v \equiv P_o^e - P^e$ and $u \equiv P - P^e$ then we can write:

$$\text{(3.5)} \quad \begin{aligned} &\frac{\theta^2}{4} > (\theta - v)v \\ &\frac{k^2\theta^2}{4} + \theta(1 - k) > (k\theta - v)v + (1 - k)\theta u. \end{aligned}$$

Clearly both inequalities are satisfied for all contracts c' such that

$u < 1$.

Similarly, when a contract c' yields beliefs $\beta(c') = 0$ and when $\underline{c}$ and $\bar{c}$ are OSI-contracts the inequalities (3.3) are rewritten:

$$(3.6)\qquad \begin{aligned} &\frac{\theta^2}{4} - \theta(1 - k) > (\theta - v)v - (1 - k)\theta u \\ &\frac{k^2\theta^2}{4} > (k\theta - v)v \ . \end{aligned}$$

Now, both inequalities are satisfied for all contracts c' such that $u > 1$. Thus, any OSI-contract $\underline{c}$ such that $\underline{u} \equiv \underline{P} - \underline{P}^e > 1$ and any OSI-contract $\bar{c}$ such that $\bar{u} < 1$, form a separating Bayesian equilibrium with the following beliefs about off-equilibrium contract offers by the incumbent: $\beta(c') = 1$ whenever c' is such that $u' < 1$ and $\beta(c') = 0$ whenever c' is such that $u' > 1$. The beliefs specified above are of course Bayesian-consistent. So far we have only shown that there exists a perfect Bayesian equilibrium which yields the symmetric information outcome. There are of course many other perfect Bayesian equilibria, however, there is one additional reason for why the particular equilibrium we have considered is of special interest, namely, that it is the only one that satisfies Kreps' intuitive criterion (see Kreps (1984)). To see this, notice that all separating equilibria are (weakly) dominated by this equilibrium. Furthermore, all pooling and semi-separating equilibria are not stable in the sense that to offer the pooling contract is not a best reply for the $\underline{\psi}$-type incumbent versus the set of pooling (or semi-separating) contracts. We summarise the above discussion in the following proposition:

<u>Proposition 3.1</u>: When the incumbent can write contracts where the price is contingent on the event of entry ($P = P^e$), the symmetric information outcome can be attained regardless of what the incumbent's type is.

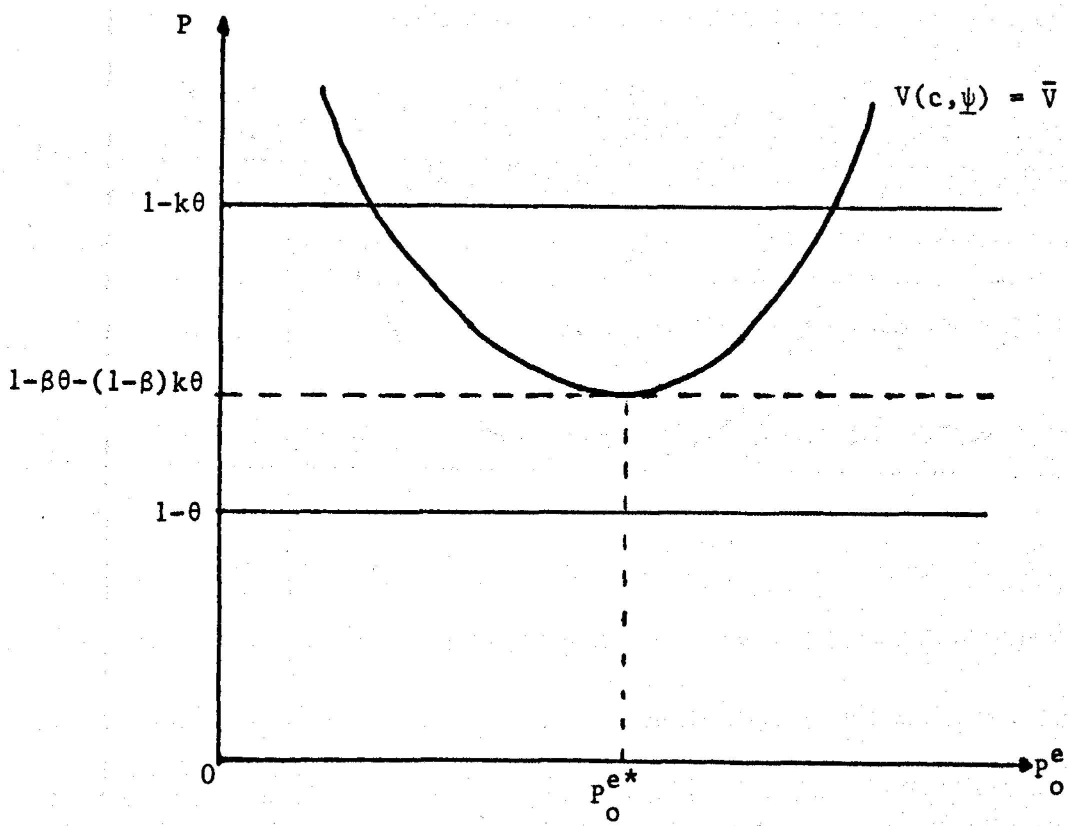
It is altogether not surprising that the OSI-outcome can be attained since the agent is assumed to be risk-neutral and since the Principal has several instruments (u and v) he can use to reveal his type. When the downstream firm is assumed to be risk-averse and when it cannot get insurance from a third party Proposition 3.1 is no longer true.[16]

We will now turn to the analysis of situations where contracts which are contingent on entry are not feasible. Then, the presence of asymmetric information will put restrictions on the form of the optimal contract, as will become clear below.

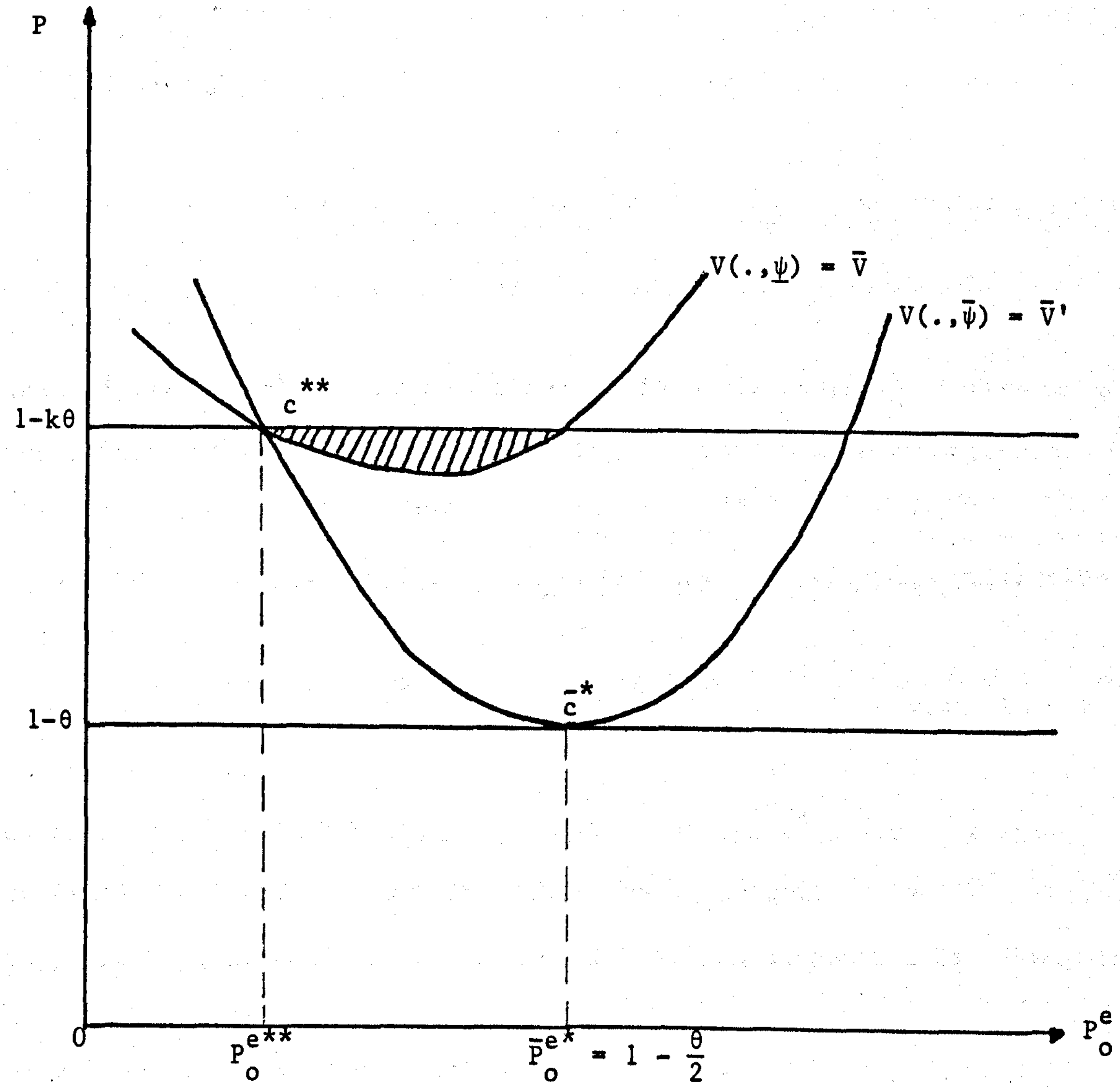
The incumbent's maximisation problem, when he is of type $\underline{\psi}$ say is rewritten as:

$$(3.7)\quad \begin{cases} \max\limits_{c \in C} \quad V(c, \underline{\psi}) \equiv (k\theta + P - P_o^e)(P_o^e - P) + P \\ \\ \text{subject to } 1 - P \geq \beta(c)\theta + (1 - \beta(c))k\theta. \end{cases}$$

Thus the constraint that the principal faces implies that P must lie between $1 - \theta$ and $1 - k\theta$. The solution to (3.7) is depicted in the figure below:




It is straightforward to verify that the Spence-Mirrlees condition: $\frac{\partial}{\partial k}\left(-\frac{V_P}{V_{p_o^e}}\right) > 0$ is satisfied and that as a consequence we obtain the following figure when we superimpose the $\bar{\psi}$-type's indifference curves $(V(c,\bar{\psi}) = \bar{V})$ and the $\underline{\psi}$-type's indifference curves:




Here $\bar{c}^*$ is the OSI-contract where $\bar{P}_o^{e*} = 1 - \frac{\theta}{2}$, and $\bar{P}^* = 1 - \theta$. Notice that the $\bar{\psi}$-type incumbent can always implement the contract

$\bar{c}^*$ and this contract dominates all other contracts c, which generate beliefs $\beta(c) = 1$.

As is usual with signalling models, there is a plethora of Bayesian equilibria and our model is no exception to this rule. Any pair of contracts $(c,\bar{c}^*)$ where c is such that $P = 1 - k\theta$ and $0 < P_o^e < P_o^{e**}$ (see figure above) constitutes a separating equilibrium. Furthermore any point in the shaded area in the figure above may be a pooling or semi-separating equilibrium of the signaling game.[17] However, following Kreps (1984) and Maskin-Tirole (1985) we may refine the Bayesian-equilibrium by using dominance and stability arguments and thus single out a unique separating equilibrium $(c^{**},\bar{c}^*)$. Henceforth we will concentrate our attention on this unique equilibrium with the justification that all other Bayesian equilibria are either dominated by $(c^{**},\bar{c}^*)$ (if they are separating equilibria) or unstable in the sense defined in Kreps (1984) (if they are pooling equilibria). (For a thorough discussion on this point see the above-mentioned two references.)

The contract $c^{**}$ is defined by the equation:

$$(3.8) \qquad V(c^{**},\bar{\psi}) = V(\bar{c}^*,\bar{\psi})$$

and it is straightforward to check that the contract $c^{**}$ is such that $P^{**} = 1 - k\theta$ and $P_o^{e**} = 1 - k\theta + \frac{\theta}{2}\sqrt{(1-k)\theta}$. Comparing the second-best contract $c^{**}$ with the OSI-contract for the $\underline{\psi}$-type $\underline{c}^*$ (where $\underline{c}^* = (\underline{P}^*=1-k\theta;\ P_o^{e*} = 1 - \frac{k\theta}{2}$ ) it is immediate that asymmetric information puts constraints on the size of the compensation price $P_o^e$ since:

$$1 - k\theta + \frac{\theta}{2} - \sqrt{(1-k)\theta} < 1 - \frac{k\theta}{2}$$

Thus we obtain our second conclusion of this section that, when contingent contracts are not feasible, asymmetric information puts restrictions on $P_o^e$ when the incumbent is of type $\underline{\psi}$, with the consequence that total welfare is improved (since there is more entry).

Finally, notice that when the difference between the "high" and the "low" probability of entry is not too small ($k < 1 - \frac{\theta}{4}$) then $P_o^e \equiv 1 - k\theta - \sqrt{(1-k)\theta} < P = 1 - k\theta$. In other words, when $k < 1 - \frac{\theta}{4}$ the second-best long term contract is of the class $C_B$ and the $\underline{\psi}$-type incumbent is indifferent between offering some form of long-term contract, where $P_o^e$ is not greater than P, or offering a short-term contract,

We have thus established that the nominal length of the contract may serve as a signal of the probability of entry. This result confirms the following basic intuition:

The downstream firm must argue in the following way, whenever it is offered a contract: "if the incumbent wants to sign a contract of a long duration, he must be worried about entry so that I infer from this that the probability of entry is high and I will only accept to sign this contract if he charges a low price. If on the other hand the incumbent offers a contract of a short duration he must not be worried so much about entry, so that I will be willing to accept a higher price".

In other words, the downstream firm recognises that "duration" may be an important element in determining to what extent the incumbent is preoccupied about entry. As Posner has put it in another context: "whether such a contract will have any exclusionary effect depends on its duration. If the contract is terminable on short notice as in the Standard Stations case, the exclusionary effect will normally be zero, since the distributor is free to take on a new supplier at any time." (Posner (1976), pp.201).

On the other hand, we saw in this and the previous section that a long term contract may be more or less exclusionary; and that some long term contracts (in the class $C_B$) have no more exclusionary effects than short-term contracts. Thus the above intuition and Posner's comment are not strictly true since they ignore the fact that long-term contracts may be more or less "flexible".

## SECTION 4: NEGOTIATION WITH SEVERAL DOWNSTREAM FIRMS

In this section we want to model the idea that when there are several downstream firms they will all achieve a worse outcome if they negotiate non-cooperatively with the incumbent than if they formed a coalition of buyers and then negotiated with the incumbent. This is true even though the incumbent has all the bargaining power in either case.

Remember that we have assumed in section 1 that the downstream firms all operate in isolated markets and that they are all identical. In other words, the downstream firms do not compete with each other in the product market. This is a restrictive assumption, however, it allows us to considerably simplify the analysis of the negotiation process when there are several downstream firms. Also, it highlights the mechanism by which the incumbent can play all downstream firms against each other.

Suppose that there are n downstream firms, which all purchase one unit of input and have reservation price 1. When entry occurs, suppose that the entrant sells to g downstream firms, where $g < n$. Then the entrant's profit is defined by $\pi(g,n)$. In other words, the entrant's profit is a function of the total size of the market, n, and the total number of customers served, g. To be able to proceed we have to specify how the entrant's return depends on g and n. We will adopt the following formulation of $\pi(g,n)$:

Assume that there is a fixed cost of entry $D > 0$ and that the entrant's unit cost is a random variable $\tilde{c}_e$ distributed on $[\underline{c},\bar{c}]$. The incumbent's unit cost is zero (see footnote (6)). If no long term contracts have been signed between the incumbent and the downstream firms there will be Bertrand competition when entry takes place, so that the entrant's profit function is defined by $\pi(g,n) = -(g.c_e + D)$.

Assume that $\underline{c} = -\theta$ and that $\bar{c} = 1 - \theta$, and finally that $\tilde{c}_e$ is uniformly distributed on $[\underline{c},\bar{c}]$ with density $f(x) = 1$. Entry only takes place if the entrant makes positive profits; hence, when no long term contracts have been signed the probability of entry is given by:

$$(4.1) \qquad \psi = n\int_{-\theta}^{0} f(x)dx - D = n\theta - D, \text{ provided that}$$

$0 < n\theta - D < 1$ so that we may write $\psi = \min\{1, \max(0, n\theta - D)\}$. (When entry takes place and there are no contracts between the incumbent and his customers, then the entrant maximises profits by serving the whole market.) We will assume that $n\theta - D \in (0,1]$.

We have adopted a rather special profit function for the entrant here, however, we will explain later that the results obtained in this section are still valid for much more general specifications of $\pi(g,n)$.

Suppose now that $g$ downstream firms have accepted long term contracts where $P_0^e = +\infty$. Then if entry occurs, the entrant will not serve these $g$ customers and the size of the entrant's market is at most $(n-g)$. Then the probability of entry is:

$$(4.2) \qquad \hat{\psi} = (n-g)\int_{-\theta}^{0} f(x)dx - D = (n-g)\theta - D,$$

provided of course that $(n-g)\theta - D > 0$, otherwise $\hat{\psi} = 0$.

It is immediate that $\hat{\psi} < \psi$ if $g > 0$. Furthermore, $\frac{d\hat{\psi}}{dg} < 0$ and there may exist a $g^*$ such that $\hat{\psi} = 0$, for all $g > g^*$. We shall assume that $g^* < n$ exists.

Clearly, the incumbent can now (by signing contracts where $P_0^e = +\infty$ with $g^*$ firms) impose the monopoly price $P = 1$ on the remaining $(n-g^*)$ firms. Any downstream firm who rejects a long term contract now faces the threat that it will have to pay its reservation price $P = 1$, if $g^*$ other firms have accepted a contract where $P_0^e = +\infty$.

We will show below that if all downstream firms negotiate non-cooperatively the incumbent can use this threat to impose the monopoly price on all downstream firms. Furthermore, the incumbent can make sure that entry will take place with positive probability, so that he can extract a positive expected surplus from the entrant. It follows that the incumbent is strictly better off than a monopolist who did not face a threat of entry.

This is a striking result and as we shall demonstrate below it does not depend on the negotiation procedure adopted by the incumbent: the latter could either visit each downstream firm sequentially or he could simultaneously mail his contract offers to all downstream firms:[18] in either case he can achieve the same payoff.

Whenever a downstream firm accepts a contract of the above form, it imposes an externality on all other downstream firms, because it reduces the probability of entry. Hence, even if downstream firms do not compete with each other on the product market, the fact that, loosely speaking, they compete for good contracts on the input market gives the incumbent enough leverage to impose the monopoly outcome on all the firms, despite the existence of a threat of entry on the input market.

A) Simultaneous offers

For the sake of illustration we will first present the case where there are only two downstream firms. Without loss of generality we can restrict ourselves to contracts which are not contingent on entry. The negotiation game proceeds as follows: first, the incumbent sends contract offers to both downstream firms simultaneously. The latter can either accept or reject the offer. Then, the incumbent reveals to both firms what the outcome of the negotiation process is and the contracts that have been accepted are implemented. We assume here, as in section 2, that there will be no renegotiation at any point after the first round of offers.

Given the above specification of the negotiation game, the most general contract form is given by $c = (P^r_{oi}, P^r_i; P_{oi}(c_j); P_i(c_j))$, where $P^r_{oi}$ and $P^r_i$ are respectively the compensation price when firm i does not trade with the incumbent and the price when firm i trades with the incumbemt, given that firm j has rejected its contract offer. And $P_{oi}(c_j)$; $P_i(c_j)$ are respectively the compensation price when firm i does not trade with the incumbent and the price when firm i does trade with the incumbent, given that firm j has accepted its contract offer $c_j$.

Note that $P_{oi}$ and $P_i$ are a function of $c_j$. We have implicitly assumed that $P^r_{oi}$ and $P^r_i$ cannot be a function of $c_j$. This is the case, for example, if a court cannot verify $c_j$ when the latter has been rejected by firm j. Also, if these two prices were contingent on $c_j$ then the incumbent may have an incentive to make "crazy" offers to firm j, which the latter would reject, but which make the contract with firm i particularly attractive to the incumbent. In any case, it will become clear below that this assumption is not at all restrictive here; the incumbent would not be able to improve his payoff if he could write more general contracts.

When the incumbent has made an offer to each firm the latter are playing a non-cooperative game where each has two pure strategies: "accept" or "reject". We can represent this game in the payoff-matrix below:

| | | Firm 2: accept | Firm 2: reject |
|---|---|---|---|
| Firm 1 | accept | $(1-P_1(c_2))$<br>$(1-P_2(c_1))$ | $(1-P_1^r)$<br>$\psi'$ |
| | reject | $\psi''$<br>$(1-P_2^r)$ | $\psi$<br>$\psi$ |

The payoffs are explained as follows:

1) If firm 1 accepts the contract it is charged a price $P_1(c_2)$ if firm 2 also accepts a contract. Then if entry occurs, firm 1 cannot do better than $1-P_1(c_2)$ since the entrant can attract firm 1 by charging a price no greater than $P_1(c_2)$ and paying the compensation price $P_{01}(c_2)$.

2) If firm 1 accepts the contract and firm 2 rejects its contract, then firm 1's payoff is $1-P_1^r$.

3) If firm 1 rejects the contract offer it has to pay P = 1 if no entry takes place and P = 0 if entry occurs. The probability of entry is given by $\psi''$, if firm 2 accepted its contract, where $\psi''$ is defined as:

(4.3)
$$\psi'' = \max\{2\theta + P_2^r - P_{02}^r - D;\ \theta - D;\ 0\}$$

↑ probability of entry when the entrant decides to serve both firms 1 and 2.

↑ probability of entry when the entrant decides to serve only firm 1 who did not contract.

4) If firm 2 also rejected its contract offer, the probability of entry is $\psi$, where $\psi = 2\theta - D$.

Firm 2's payoffs are similarly defined and we have:

(4.4)
$$\psi' = \max\{2\theta + P_1^r - P_{01}^r - D;\ \theta - D;\ 0\}$$

We are able to establish:

Proposition 4.1: When $2\theta > D > \theta$, the optimal payoff for the incumbent is:

$$2 + \left(\frac{2\theta - D}{2}\right)^2$$

In other words the incumbent is able to both implement the monopoly solution and to extract the maximum expected surplus from the entrant.

Proof: Note first that by choosing $(P_{oi}^r - P_i^r)$ large enough the incumbent can insure that $\psi' = \psi'' = 0$. Then, provided that $1-P_i(c_j) > 0$ and $1-P_i^r > \psi$, "accept" is a dominant strategy for both firms. Hence the incumbent can approximate the monopoly solution arbitrarily closely.

Next, it remains to determine the optimal $P_{oi}(c_j)$: given that both firms will accept their contract offer and that $P_i(c_j) \geq 1$, the probability of entry is defined by:

$$(4.5) \qquad \psi''' = \max \left\{ \int_{-2\theta}^{1-P_{01}(c_2)+1-P_{02}(c_1)} f(x)dx; \int_{-\theta}^{1-P_{oi}(c_j)} f(x)dx; 0 \right\}$$

Note that if $P_{oi} < \theta + 1$ then $\psi''' = 2\theta + 2 - (P_{01} + P_{02}) - D$. Thus consider the following program for the incumbent:

$$(4.6) \qquad \max_{(P_{01};P_{02})} (2\theta + 2 - (P_{01} + P_{02}) - D)(P_{01} + P_{02} - 2) + 2$$

$$\text{subject to: } P_{oi} < 1 + \theta$$

The solution to this program is $P_{01} = P_{02} = 1 + \frac{\theta}{2} - \frac{D}{4}$ and the constraint in (4.6) is redundant. Also, from (4.6) we obtain that the optimal payoff to the incumbent is: $\left(\frac{2\theta - D}{2}\right)^2 + 2$.

Notice that the incumbent could have achieved the same payoff by restricting himself to contracts of the form $c = (P_{01}, P_1, P_0^r, P_1^r)$. Using the same argument as above Proposition 4.1 can be generalised to the case where there are n downstream firms:

<u>Proposition 4.2</u>: When $n\theta > D > \theta$ the optimal payoff for the incumbent is $n + \left(\frac{n\theta - D}{2}\right)^2$.

Remark 1: When $D < \theta$, the proof of propositions 4.1 and 4.2 does not go through, for then $\psi'$ and $\psi''$ are strictly positive (since $\theta - D > 0$) and in order for "accept" to be a dominant strategy we must have $P_i(c_j) < 1$. When $D = 0$, the incumbent cannot do better than implementing the optimal contract, $c_i^* = (P_i(c_j) = 1 - \theta;\ P_{oi} = 1 - \frac{\theta}{2})$.[19]

If all downstream firms were negotiating cooperatively they would insist on a price $P_i(c_j) = 1 - \psi$, where $\psi = n\theta - D$, so that the incumbent's total payoff would be $n(1 - n\theta + D) + (\frac{n\theta - D}{2})^2$ provided $\psi = n\theta - D < 1$, otherwise, if $\psi = 1$, the incumbent could only hope for a total payoff of 1/4.

Remark 2: In the proof of proposition 4.1 it appears that if entry occurs the optimal solution is for the entrant to serve the whole market. This is an unfortunate feature of the model and it is the consequence of the assumptions we made about $\pi(g,n)$. We have assumed basically that

$$\frac{d\pi(g,n)}{dg} > 0 \text{ for all } g \in [0,n].$$

The entrant has an increasing returns to scale technology so that his profits are maximised when he serves the whole market. Consequently the incumbent maximises the expected surplus he can extract from the entrant by letting him serve the whole market when he enters. If the entrant had a U-shaped average cost curve we would in general have market-sharing.

The reader may have the impression that the results obtained above heavily rely on the simultaneity of the offers and that if downstream firms were able to negotiate sequentially with the incumbent the latter would not be able to impose the monopoly price on all firms. It turns out, however, that when downstream firms negotiate sequentially, the same outcome can be achieved by the incumbent.

B) Sequential offers

Now the incumbent negotiates first with one firm, then moves on to the next one, etc... . As in the previous subsection we start with the case where there are two downstream firms. We assume that each firm knows whether it is the first or the second to negotiate with the incumbent and if it is the second firm, we also assume that it knows the outcome of the negotiation game with the first firm. If neither of these assumptions hold, we are in a situation which is equivalent to the one just described, where the incumbent makes simultaneous offers.

Let $F_1$ be the first downstream firm to be offered a contract and $F_2$ be the second downstream firm to negotiate with the incumbent. Suppose that $2\theta > D > \theta$, then we will show that the incumbent can achieve the same payoff as in Proposition 4.1.

When the incumbent offers a contract $c = (P_{01}, P_{02})$ to $F_1$ the latter must determine how much it would obtain if it rejected the offer in order to decide on whether to take the offer or not. If $F_1$ rejects the offer, $F_2$ will be informed about it, so that $F_2$ will insist on a contract where $1 - P_2 > 2\theta - D$ (which is the probability of entry when no contract is signed). Then the incumbent will offer $F_2$ a contract of the form $(P_{02}, 1 - 2\theta + D = P_2)$. What is the optimal compensation price $P_{02}$ for the incumbent? $P_{02}$ is chosen to solve:

(4.7)
$$\begin{cases} \max_{P_{02}} & \psi . P_{02} + (1 - \psi)(2 - 2\theta + D) \\ \text{subject to} & \psi \geq 0 \end{cases}$$

$$\text{where} \quad \psi = \int_{-2\theta}^{P_2 - P_{02} - D} f(x)dx = 1 - P_{02}$$

It is straightforward to verify that the optimal price $P_{02}$ is such that $\psi = 0$. It follows that if $F_1$ rejects the incumbent's offer it gets an expected payoff of zero (since the probability of entry when it rejects is $\psi = 0$). Hence, $F_1$ is willing to accept an offer where $P_1 = 1$.

Given that $F_1$ accepted an offer $(P_{01},1)$, $F_2$ will accept any offer $(P_{02},P_2)$, where:

(4.8)
$$1 - P_2 \geq \psi_1 = \int_{-2\theta}^{1-P_{01}-D} f(x)dx = 1 - P_{01} - D + 2\theta$$

Suppose then that $F_2$ accepts a contract where (4.8) holds with equality (that is, it accepts the contract $c = (P_{02},P_2 = P_{01} + D - 2\theta)$), then the question is, what the optimal prices $P_{02}$ and $P_{01}$ are for the incumbent. To determine $P_{02}$ the incumbent solves:

(4.9)
$$\begin{cases} \max_{P_{02}} & \psi_2(P_{02} + P_{01}) + (1 - \psi_2)(1 + P_{01} + D - 2\theta) \\ \text{subject to} & \psi_2 \geq 0 \end{cases}$$

$$\text{where} \quad \psi_2 = \int_{-2\theta}^{1-P_{01}+P_{01}+D-2\theta-P_{02}-D} f(x)dx = 1 - P_{02}$$

So that the incumbent's problem reduces to:

(4.10)
$$\begin{cases} \max_{P_{02}} & (1 - P_{02})(P_{02} - 1 - D + 2\theta) \\ & \text{s.t.} \quad P_{02} \leq 1 \end{cases}$$

The solution is $P_{02}^{*} = 1 - \frac{D-2\theta}{2}$, and the incumbent's payoff is:

(4.11) $$\left(\frac{2\theta-D}{2}\right)^2 + 1 + P_{01} + D - 2\theta$$

Clearly the incumbent maximises his payoff by setting $P_{01}$ as high as possible, subject to the constraint that $\psi_1 = 1 - P_{01} - D + 2\theta \geq 0$, (otherwise, the entrant would not find it profitable to try and serve firm 1, so that the incumbent would not obtain $P_{01}$). So let $P_{01} = 1 - D + 2\theta$, then the incumbent's payoff is:

(4.12) $$\left(\frac{2\theta-D}{2}\right)^2 + 2.$$

Thus, we have established that the incumbent can achieve the same payoff, when he negotiates sequentially with 2 firms than when he negotiates simultaneously. The above argument can be generalised to n firms. For the interested reader see the Appendix.

## SECTION 5: CONCLUSION

This paper provides a simple model of entry-prevention through contracts with customers in intermediate good industries. We have identified two important factors which may impede entry when contracts are signed: The compensation price a customer must pay to the incumbent when he does not trade with the incumbent, and possibly, the duration of the contract. We found that when there is symmetric information about the probability of entry and/or the contract signed with the customer is completely contingent, then the incumbent will always sign a contract of the longest possible nominal duration with his customers.

When there is asymmetric information and when the contract signed is incomplete, situations may arise, where the nominal duration of the contract is an increasing function of the probability of entry. The reason is that nominal contract length may act as a signal of the probability of entry. In section 3, however, we emphasize the idea that the exclusionary effect of a contract cannot be assessed simply by looking at its duration. What is important is the overall "flexibility" of the contract; in other words the effective length of the contract.

One feature of our model, which at first sight may seem disturbing is that once an (incomplete) long term contract is signed between the incumbent and the downstream firm, the former is better off when entry occurs than when it does not occur. Generally, one thinks that entry hurts the incumbent and it seems that our model does not capture this basic fact. This is not true, however, since from an ex-ante point of view (before a contract is signed), a higher probability of entry always hurts the incumbent. Also, if no long term contract is signed, entry is worse for the incumbent than no entry.

In section 4 we have identified an important case where the threat of entry benefits the incumbent. That is, the incumbent is strictly better off than a natural monopoly. The reason is that the incumbent can exploit the competition between downstream firms to impose a high price on each one (in fact, the monopoly price $P = 1$) and at the same time set the compensation price, $P_o^e$, in an optimal way so as to extract the highest possible expected surplus from the entrant. This outcome is possible even if ex-ante there is a probability of entry, $\psi = 1$ !

Several extensions to our model may be interesting, like allowing for entry in the downstream markets and introducing several incumbents. Our model could also be used as a starting point to study how the flexibility of contracts varies with the level of the economic activity. In order to do that one would have to explain how k (the difference between $\bar{\psi}$ and $\underline{\psi}$) varies with this level.

Stiglitz (1984) for example argued that in a recession the threat of entry may be less important. Then according to our model, contracts negotiated in recession periods would be of a shorter duration (or more generally more flexible). Thus, there would be more price-flexibility in recessions.

In fact, it very much depends on how one interprets entry. One could equally well argue that in a recession firms want to maintain the same volume of sales in a shrinking market, so that entry (interpreted as competition from other firms) would be more likely and therefore contracts of a longer duration would be signed, i.e. we would have more rigidity in recessions. This is a very loose story, of course, but we feel that price-stickiness can be explained through incomplete contracts of a more or less long duration (or more generally, of a more or less high degree of flexibility) and that the rigidity of these contracts depends on the threat of entry by competitors. The threat of entry in turn depends on the level of economic activity.

As a final comment we would like to emphasize that our model has a wide range of possible applications. For example, it may be applied to situations where a union inside a firm negotiates with the management over wage and employment contracts and where there is a positive probability that outside workers may show up in the future. Another application may be to R&D and the diffusion of innovations: an incumbent may prevent (or delay) the diffusion of new technology owned by an entrant, if he has signed long-term contracts with his customers. This in turn may have adverse effects on investment in R&D.

FOOTNOTES

(1) See for example, the seminal contributions by Spence (1977) and Dixit (1979, 1980).

(2) For a recent survey see Fudenberg-Tirole (1984).

(3) Spence (1977, pp. 544) for example briefly mentioned contracts as a method for impeding entry; see also Williamson (1979). Furthermore, there is a literature on barriers to entry and vertical integration that is relevant to our discussion, since most of the time what vertical integration achieves in this literature, can also be done through an appropriate contract. (See Blair and Kaserman (1983).)

(4) This position has been forcefully defended by Bork (1978), for example.

(5) The choice of a uniform distribution function f(x) is entirely for the sake of computational simplicity and none of our qualitative conclusions depend on it.

(6) More precisely, the incumbent and the downstream firms know the true probability of entry, but they do not observe the precise opportunity cost of the entrant.

(7) Notice that our model does not exactly describe the situation of the United Shoe Machinery case. There the incumbent signs long-term contracts with the buyer but entry also takes place in the buyer's market. We could easily modify the model in order to accommodate this case; we believe, however, that our modified model would not entirely describe the situation faced by the United Shoe Machinery corporation. In fact the latter faces both an entry-prevention problem and a "durable goods Monopoly" problem. Thus, both our model and the model by, for example,

(8) Shavell (1980) and Hart-Moore (1985) among others have argued that contracts that are completely contingent on every state of nature may be too costly to write. In the case of contract-offers one would expect that transaction costs would be even greater since a priori there can be an even larger number of different contract offers than there are states of nature. This is not to say that the contract may not be contingent on some (potential) offers by the entrant. We make a simplifying assumption by excluding this possibility.

(9) Since the discount factor in our model is equal to one, both parties to the contract are indifferent to the timing of the payments; so long as the aggregate payments remain the same nothing is changed if we allowed for transfers in period 1.

(10) Lewis Kornhauser pointed out to us that as the problem is set up here the possibility arises that the entrant will enter the market and simply sell an additional unit to the downstream firm at a negative price; the latter will then freely dispose of the additional unit. The reason why this is possible is because we have assumed a zero marginal cost for both the incumbent and the entrant. We can, however, easily eliminate this possibility by assuming that marginal cost, c, is positive and that $\theta - c < 0$. We can then return to our previous formulation by adding c to all prices.

(11) The following question arises about the optimal compensation price $P_o^{e*} = 1 - \theta/2$. Consider two probabilities of entry $\theta_1$, $\theta_2$, where $\theta_1 > \theta_2$ then the optimal price $P_{oi}^e$ corresponding to each probability is such that:

$$P_{o1}^e = 1 - \frac{\theta_1}{2}; \quad P_{02}^e = 1 - \frac{\theta_2}{2}$$

(11) continued

In other words, the higher the probability of entry the lower is the optimal compensation price. This seems counter-intuitive, since one would tend to believe that when an incumbent faces a higher threat of entry (that is the entrant's expected surplus is larger), the incumbent would want to raise $P_o^e$. In fact the reason why there is an inverse relation between $P_o^e$ and the probability of entry is that the price of the good in the contract ($P$ and $P^e$) is also inversely related with $\theta$.

One may wonder whether this property of the compensation price is robust. We have considered two changes which might have altered this result. First, we allowed for transfers between the incumbent and the downstream firm before the contract is performed and we found that it did not affect our result. Secondly, we verified whether we would still obtain the same inverse relation if we changed the formalisation of the probability of entry in the following way:

Define the support of the uniform distribution $f(\pi^A)$ by $[-\theta,t]$; where $t \geq -\theta$. Then $\psi = \frac{\theta}{t+\theta}$, and $P_o^{e*} = 1 - \psi + \frac{\theta}{\sqrt{3}}$. Clearly, the inverse relationship between $P_o^{e*}$ and $\psi$ is preserved.

(12) According to Proposition 2.1, the incumbent chooses the optimal compensation price in such a way that entry is not completely deterred. In proposing such a contract, the incumbent does <u>strictly</u> better than by "integrating vertically", i.e. by choosing the compensation price in such a way as to completely block entry.

Notice also that if the entrant's opportunity cost was observable and verifiable, the incumbent could choose the price $P_o^e$ to equal the opportunity cost of the entrant; the latter would be indifferent between entering and not entering and

(12) continued

under our assumption that the entrant would always enter in that case, we would obtain a social optimum.

(13) Even though it is clear that contracts between the incumbent and his customer are suboptimal from a Welfare point of view it is not clear how the courts can prevent such exclusionary practices. One possibility could be to set an upper bound on liquidated damages, but this will not work in general, for the parties to the contract can achieve the same outcome through side-payments when they cannot choose $P_o^e$ optimally. Another possibility could be to limit the length of the contract. This is indeed the step that has been taken by the Federal Trade Commission in the case, <u>F.T.C. vs. Motion Picture Advertising Service Company</u>, and the Supreme Court affirmed (see Bork (1978, pp.308)). The drawback with this procedure is that the nominal length of the contract is not a good measure of the lock-in effect of the contract.

(14) For a recent discussion on the issue of third party insurance see the survey by Hart and Homstrom (1985).

(15) Some situations may arise where the downstream firm is better informed about the probability of entry. Then we have a classic self-selection problem and all the results obtained in this section would also apply to this case.

(16) The two OSI contracts $\underline{c} = (\underline{P} = 1 - k\theta;\ P_o^e = 1 - \frac{k\theta}{2})$ and $\bar{c} = (\bar{P} = 1 - \theta;\ \bar{P}_o^e = 1 - \frac{\theta}{2})$ do not form a separating equilibrium: For $(\underline{c},\bar{c})$ to form a separating equilibrium it is necessary to have:

(16) continued

$$V^*(\bar{c},\bar{\psi},\ \beta=1) \geq V^*(\underline{c},\ \underline{\psi},\ \beta=0)$$

or $$\theta^2/4 + 1 - \theta \geq (\theta - \frac{k\theta}{2})\ \frac{k\theta}{2} + 1 - k\theta$$

Simplifying we obtain that: $\theta/4(1 - k) \geq 1$, which is not possible since $\theta < 1$.

(17) He could also have a mixture of both negotiation procedures.

(18) The proof of Propositions 4.1 and 4.2 can also be modified to incorporate the case where there is competition in the product market between all the different downstream firms: in order to have "accept" as a dominant strategy for all firms the incumbent must only modify his contract offer to each firm in the following way: (in the case where there are only two firms) he promises to sell to the firm who accepts the contract offer two units of input (instead of just one) at zero price, if the other firm rejects its contract offer. Then, if either of the firms rejects its contract offer it cannot make positive profits. Thus it is willing to accept the optimal contract for the incumbent defined in the proof of Proposition 4.1.

## APPENDIX TO SECTION 2

### 2.A: MORE GENERAL DISTRIBUTION

The main qualitative results of this paper rely on the following property P, already established in section 2 in the particular case where the distribution function f(x) characterising the stochastic entry is uniform on $[-\theta, 1 - \theta]$. (Proposition 2.1).

---

P

The incumbent's maximisation program:

$$\text{(I)} \qquad \max_{u} \; R(u) = \psi'(u).u + 1 - \psi'(u)$$

$$\text{s.t.} \quad \psi'(u) \geq 0$$

where $\psi = \int_{-\theta}^{0} f(x)dx$ is the probability of entry when no contract has been signed and $\psi'(u) = \int_{-\theta}^{-u} f(x)dx$ is the probability of entry after a contract $c = (P_o^e, P, P^e)$ such that $P_o^e - P^e = u$, has been accepted.

(I) has a solution $u^*$ such that:

(i): $0 < \psi'(u^*) < \psi$ .

(ii): $R(u^*) > 1 - \psi$.

---

(In Proposition 2.1 we obtain: $u^* = \frac{\theta}{2}$, $\psi'(u^*) = \theta - u^* = \frac{\theta}{2} \in \,]0,\theta[$, and $R(u^*) = \frac{\theta^2}{4} + 1 - \theta > 1 - \theta = 1 - \psi$.)

This property P turns out to be quite general as we can easily show:

<u>Proposition:</u> Property P is satisfied by any continuous distribution function f(x) such that the support of f has its lower bound $(-\theta)$ finite and strictly negative and contains the segment $[-\theta,0]$.

<u>Proof:</u> Let $\underline{l}(u) = \psi'(u).u = u.\int_{-\theta}^{-u} f(x)dx$.

We have: $\underline{l}(0) = 0$; $\underline{l}(\theta) = 0$; $\underline{l}(u) > 0$ for $u > 0$, small enough; $\underline{l}(u) \leq 0$ for $u \notin [0,\theta]$; $\underline{l}(u)$ is continuous on $[0,\theta]$. Therefore, $\underline{l}(u)$ has a global maximum $u^* \in ]0,\theta[$, such that $\underline{l}(u^*) > 0$.

We clearly have:

1) $\psi'(u^*) = \int_{-\theta}^{-u^*} f(x)dx > 0$, because $u^* < \theta$.

2) $\psi'(u^*) < \psi = \int_{-\theta}^{0} f(x)dx$, because $u^* > 0$ and $[-\theta,0] \subset$ supp. f.

3) $R(u) \leq R(u^*)$ for all u, and $R(u^*) > 1 - \psi$.

The proposition is established.

## <u>2.B: PROOF OF PROPOSITION 2.2</u>

The incumbent faces the following maximisation problem:

$$\max_{(P_o^e,P,P^e)} \quad (P^e - P_o^e + \theta)(P_o^e - P) + P$$

$$\text{s.t.} \quad (P^e - P_o^e + \theta)(1 - P^e) + (1 - P^e + P_o^e - \theta)(1 - P) - r_{12} = \theta - r \quad (*)$$

Setting $u = P_o^e - P^e$, the above program is equivalent to:

$$\max_{(u,P,P^e)} \quad (\theta - u)u + 1 - \theta - r_{12} + r \equiv R(u, P, P^e)$$

Clearly, for all u, $R(u,P,P^e)$ is maximised for $P = P^e$, i.e. for $r_{12} = 0$. From (*), we then have: $P^* = P^{e*} = 1 - \theta + r$.
Finally, $R(u,P^*,P^{e*})$ is maximised for $u^* = \frac{\theta}{2} = P_o^{e*} - P^{e*}$. Then, necessarily: $P_o^{e*} = 1 - \frac{\theta}{2} + r$.

## 2.C: PROOF OF PROPOSITION 2.3

Here the incumbent faces the following maximisation program:

$$\max_{(P_o^e,P,P^e)} \quad (P^e - P_o^e + \theta)V(P_o^e) + (1 - P^e + P_o^e - \theta)V(P)$$

$$\text{s.t.} \quad (P^e - P_o^e + \theta)(P - P^e) + 1 - P = \theta \quad (**)$$

The first order conditions are:

$$\text{(a)} \qquad V'(P) = \lambda \quad (\lambda \equiv \text{Lagrange multiplier})$$

$$\text{(b)} \qquad V(P_o^e) - V(P) + \lambda(P - 2P^e + P_o^e - \theta) = 0$$

$$\text{(c)} \qquad V(P) - V(P_o^e) + (P^e - P_o^e + \theta)V'(P_o^e) + \lambda(P^e - P) = 0$$

(a),(b),(c) => $[V'(P_o^e) - V'(P)].(P^e - P_o^e + \theta) = 0$. If $P^e - P_o^e + \theta = \psi'' = 0$, then (**) implies: $P = 1 - \theta$, so that the incumbent's payoff is: $V(1 - \theta)$. If $P^e - P_o^e + \theta = \psi'' \neq 0$ then necessarily: $V'(P_o^e) = V'(P)$, which implies by strict concavity of V, that $P_o^e = P$. The f.o.c. (b) then becomes: $P - 2P^e + P^e - \theta = 0$, i.e. $P = P_o^e = P^e + \frac{\theta}{2}$. Then $\psi'' = P^e - P_o^e + \theta = \frac{\theta}{2} > 0$ and, by (**) we obtain: $P^* = 1 - \theta + \frac{\theta^2}{4}$, which yields the payoff $V(1 - \theta + \frac{\theta^2}{4})$ $> V(1 - \theta)$ to the incumbent. Therefore, the optimal contract is uniquely defined by: $P_o^{e*} = P^* = 1 - \theta + \frac{\theta^2}{4}$, $P^{e*} = P^* - \frac{\theta}{2}$, and $V(1 - \theta + \frac{\theta^2}{4})$ is the incumbent's optimal payoff.

## APPENDIX TO SECTION 4

We will show that when the incumbent negotiates sequentially with n downstream firms he can achieve the same payoff as when he negotiates simultaneously with them; when $D \geq \theta$, his payoff is $n + \left(\frac{n - D}{2}\right)^2$.

(Notation: let $P^j$ = the price firm $F_j$ has to pay if it trades with the incumbent. Let $P_o^j$ = the price firm $F_j$ has to pay if it does not trade with the incumbent.)

Step 1: Suppose that the first (n-1) downstream firms $F_1, F_2, \ldots, F_{n-1}$ have rejected the incumbent's offer. Then it is optimal for the incumbent to offer the contract, $c = (P^n = \min(1 - n\theta + D, 1);\ P_o^n = P^n + \theta)$ to the $n^{th}$ firm, $F_n$.

Proof: Without loss of generality suppose that $1 - n\theta + D \leq 1$. If the incumbent offers any contract c where $P^n = 1 - n\theta + D$, then it will be accepted by $F_n$. We have to check that it is optimal for the incumbent to choose $P_o^n = P^n + \theta$. Consider the incumbent's program below:

$$
\text{(I)} \quad \left\{ \begin{array}{ll} \max\limits_{P_o^n} & \hat{\psi}.P_o^n + (1 - \hat{\psi})(n - n\theta + D) \\ & \\ \text{s.t.} & \hat{\psi} \geq 0 \end{array} \right.
$$

$$
\text{where} \quad \hat{\psi} = (n - 1)\theta - D + (P^n + \theta - P_o^n). \qquad (*)
$$

The first term on the RHS of (*) represents the entrant's gain from entering and trading with the (n-1) firms who have not signed a contract; the second term represents the entrant's marginal gain from trading with the $n^{th}$ firm who has signed a contract. The entrant will only trade with the $n^{th}$ firm if $P^n + \theta - P_o^n \geq 0$.

Since $n - n\theta + D > P^n + \theta = 1 - n\theta + D + \theta$, it is clearly in the best interest of the incumbent to minimise $\hat{\psi}$ by setting $P_o^n = P^n + \theta$. It is not in his interest to set $P_o^n > P^n + \theta$, for when $P_o^n = P^n + \theta$, the entrant is indifferent between trading and not trading with the $n^{th}$ firm, whenever his opportunity cost is equal to $-\theta$; being indifferent, the entrant will trade with the $n^{th}$ firm (by assumption) and the incumbent can increase his payoff by $P_o^n$ with positive probability.

<u>Step 2</u>: Suppose that firms $F_1, F_2, \ldots, F_{k-1}$ have rejected offers made by the incumbent and that firms $F_{k+1}, F_{k+2}, \ldots, F_n$ will accept contracts of the form:

$$c_j = (P^j, P_o^j = P^j + \theta) \qquad j = k+1, \ldots, n$$

Then it is optimal for the incumbent to have $F_k$ accept a contract: $c_k = (P^k, P_o^k = P^k + \theta)$.

<u>Proof</u>: Suppose that $P_o^k < P^k + \theta$. Then the probability of entry faced by firm $F_j$, $(j = k+1, \ldots, n)$ when $F_j$ rejects and all firms $F_i$ $(i = k+1, \ldots, n;\ i \neq j)$ accept a contract, is given by:

$$\hat{\psi} = k\theta - D + (P^k + \theta - P_o^k)$$

Now all firms $F_j$ will insist on a price $P^j$ such that:

$$P^j \leq 1 - k\theta + D - (P^k + \theta - P_o^k)$$

The best the incumbent can do in that case is to set $P^j = 1 - k\theta + D - (P^k + \theta - P_o^k)$; $j = k+1, \ldots, n$. Then his payoff is:

$\hat{\psi}.P_o^k + (1 - \hat{\psi})((k - 1) + (n - k + 1)(1 - P^j))$. Again, since $(k - 1) + (n - k + 1)(1 - P^j) > P^j + \theta$, for all k,j it is in the best interest of the incumbent to minimise $\hat{\psi}$, by setting $P_o^k = P^k + \theta$.

Steps 1 and 2 imply that if $F_1$ rejects the incumbent's offer then the Nash-equilibrium outcome of the game is such that every firm $F^j$, $j > 1$, accepts a contract where $P_o^j = P^j + \theta$, so that if $F_1$ rejects a contract offer, the probability of entry will be zero in equilibrium and $F_1$ can only hope for an expected payoff of zero (since $D > \theta$). Therefore $F_1$ will accept a contract $c_1 = (P^1 = 1, P_o^1)$.

<u>Step 3</u>: Given that $F_1$ accepts $c_1 = (P^1 = 1; P_o^1)$, if the next (n-2) firms reject their contract offer, $F_n$ will be offered a contract of the form $(P^n, P_o^n = P^n + \theta)$ that it will accept.

The proof of step 3 is identical to the proof of step 1.

<u>Step 4</u>: Given that $F_1$ accepts $c_1 = (1, P_o^1)$, that firms $F_2, \ldots, F_{k-1}$ reject their contract offer and that firms $F_j$, $j = k+1, \ldots, n$, accept contracts where $P_o^j = P^j + \theta$, it is optimal for the incumbent to have firm $F_k$ accept a contract where $P_o^k = P^k + \theta$.

The proof of step 4 is identical to the proof of step 2.

Again, steps 3 and 4 imply that if $F_2$ rejects the incumbent's offer all other firms will accept contracts where $P_o^j = P^j + \theta$, so that $F_2$ can hope to get at most:

$$\psi_2 = \int_{-2\theta}^{1-P_o^1-D} f(x)dx = 2\theta + 1 - P_o^1 - D.$$

(provided $P_o^1 \leqslant 1 + \theta$, otherwise $\psi_2 = 0$).

Therefore $F_2$ will accept a contract $c_2$ of the form:

$$c_2 = (1 - \psi_2, P_o^2).$$

We can establish the analogue of steps 3 and 4 when both $F_1$ and $F_2$ have respectively accepted contracts $(1, P_o^1)$ and $(1 - \psi_2, P_o^2)$. Hence firm $F_3$, if it rejects its contract offer can hope to get at most:

$$\psi_3 = \int_{-3\theta}^{1-P_o^1+1-\psi_2-P_o^2-D} f(x)dx = 1 - P_o^2 + \theta.$$

More generally, in equilibrium, $F_k$ can hope to get at most:

$$\psi_k = 1 - P_o^{k-1} + \theta .$$

Hence, given that all firms accept contracts of the form, $c_k = (1 - \psi_k, P_o^k)$ the incumbent will choose $P_o^1, P_o^2, \ldots, P_o^n$ so as to maximise his total payoff:

$$R = \hat{\psi}.\left(\sum_{j=1}^{n} P_o^j\right) + (1 - \hat{\psi}).\left(n - \sum_{j=2}^{n} \psi_j\right)$$

where $\hat{\psi} = \int_{-n\theta}^{1-P_o^1+1-\psi_2-P_o^2+\ldots+1-\psi_n-P_o^n-D} f(x)dx$. Now, setting all $\psi_j$ equal to zero we have:

$$R = \hat{\psi}.\left(\sum_{j=1}^{n} P_o^j - n\right) + n .$$

where $P_o^j = 1 + \theta$, for all $j = 1,\ldots,n-1$, since $\psi_j = 0 \Leftrightarrow P_o^j = 1 + \theta$, also, $P_o^j = 1 + \theta \Rightarrow \hat{\psi} = (1 + \theta - P_o^n - D)$ so that:

$$R = (1 + \theta - P_o^n - D)((n - 1)\theta - 1 + P_o^n) + n .$$

Solving for the optimal $P_o^n$ we get:

$$R = \left(\frac{n-D}{2}\right)^2 + n .$$

# CHAPTER 3: VERTICAL RESTRAINTS IN A MODEL OF VERTICAL DIFFERENTIATION

## SECTION 1: INTRODUCTION

There has recently been renewed interest in the study of vertical restraints in contractual relationships between manufacturers and distributors. A number of theoretical studies have attempted to explain why manufacturers may want to impose vertical restraints on retailers and what the welfare implications of these practices are (for recent surveys see Caves (1984) and Rey-Tirole (1985a)). At the same time several economists at the FTC have published two exhaustive empirical studies on resale-price-maintenance and other vertical restraints (see Overstreet (1983) and Lafferty et al. (1984)).

This recent literature concentrates almost exclusively on efficiency explanations for vertical restraints and the older view that these restraints may be devices for enforcing retailer- or manufacturer-cartels has not received much new attention. The present paper does not depart from this trend; it presents an alternative efficiency explanation for vertical restraints.

Essentially, three different efficiency explanations have been given so far for vertical restraints.[1] One of the earliest is due to Telser (1960) and is based on an externality argument. He argued that if retailers provide pre-sales services, such as informing the consumer about the characteristics of a product, then too much price competition among retailers may hurt the manufacturer. An individual retailer may be tempted to cut his retail costs, by not providing any pre-sales services, and reduce his price accordingly; however, if all retailers follow this policy, aggregate demand for the manufacturer's product will suffer. When there is such a free-rider problem in the provision of costly pre-sales services, it may be optimal for the manufacturer, and from a welfare point of view, to set up vertical restraints that limit competition among retailers. A minimum retail price or exclusive territories, for example, would eliminate this free-rider problem. More recently, several other authors have developed similar externality arguments (see Mathewson-Winter (1984) and Marvel-McCafferty (1984)).

The second explanation for vertical restraints is based on the double-marginalization problem (Spengler (1950) and Dixit (1983)). The idea is that if retailers have a local monopoly, then when manufacturers offer a linear-price contract the retail price (wholesale price plus retailer's margin) will be above the price charged by a vertically integrated monopolist, unless the manufacturer sets the wholesale price equal to his unit cost of production. This problem can be overcome through either a maximum-retail-price or a franchise-fee contract with exclusive territories.

The third explanation is concerned with the optimal density of retail outlets (see Gould-Preston (1965), Gallini-Winter (1983) and Dixit (1983)). If aggregate demand for the manufacturer's product is increasing in the number of outlets, then the manufacturer may want to reduce competition in the retail market so as to guarantee higher margins to retailers and thereby encourage entry of new retailers.

All these explanations have one feature in common: vertical restraints, such as resale-price-maintenance, exclusive territories, quantity fixing or franchise fees - whether used in combination or individually - are shown to be equivalent to vertical integration. In other words, they are efficient. (One of the main purposes of Mathewson and Winter (1984) is to find the minimum set of restraints that guarantee efficiency). Rey and Tirole (1985b), however, have shown that if there is uncertainty about demand or retail costs, then none of these standard vertical restraints are always efficient.

In this paper we develop a fourth explanation for vertical restraints based on efficiency considerations. We consider the case where retailers provide cum-sales or post-sales services. (This case does not fit into Telser's explanation, since here services are not a public good: the consumer only benefits from a given retailer's services if he purchases the good from him). The provision of cum-sales services gives rise to a situation of vertical differentiation - characterized by the fact that if two distinct products are offered at the same price, then all consumers will choose the higher quality one (all consumers agree on what constitutes a higher- versus a lower-quality product). It is a well-known result (see Mussa-Rosen (1978)) that a vertically integrated monopolist would offer products of different qualities in order to price-discriminate among consumers with different incomes and different willingness to pay for quality. It is also well-known (Shaked-Sutton (1982)) that if the manufacturer sells the good to independent retailers, then the latter will choose different qualities in order to "relax price competition through product differentiation".

We show, however, that, despite this apparent coincidence of behaviour, there is a substantial conflict of interests between manufacturer and retailers, which makes it desirable for the manufacturer to resort to vertical restraints. Furthermore, simple vertical restraints, like resale-price-maintenance (from now on RPM) or franchise fees, are not sufficient to enforce the efficient outcome. The reason is the following. Essentially, the manufacturer who sells to independent retailers faces two conflicting aims: (i) the extraction of the retailer's rent and (ii) the maximum extraction of consumer surplus by means of quality-differentiation and high prices. With a linear-price contract (which only specifies the wholesale price) retailers relax price competition through product differentiation and as a consequence the second objective is partially achieved, but not the first (since retailer's profits will be positive). The imposition of a franchise fee will not bring about efficiency, the main reason being the fact that competition between retailers keeps prices "too low". If, on the other hand, retailers are prevented from making positive profits by differentiating their products - by means, for example, of RPM - then the first objective is achieved, but not the second.

Therefore - unlike all the existing models of vertical restraints with no uncertainty - in our model the simplest forms of vertical restraints, although they represent an improvement on linear pricing, do not restore efficiency, that is, they do not approximate the outcome of vertical integration. In order to achieve the first-best the manufacturer has to resort to more sophisticated contracts and we describe one of them.

A further feature of our model, which is worth noting here, is that there is a clear conflict between manufacturer and retailers as to the choice of differentiation. As noted by Rey and Tirole (1985a), this conflict was absent from all models of vertical restraints with horizontal differentiation in the retail market presented so far in the literature.[2]

One FTC case which comes close to our analysis is that of Coors beer. Coors beer defended its use of RPM as a means of encouraging competition in services by its distributors. The latter were supposed to offer refrigeration and product rotation services, which increased the quality of the beer (see McLaughlin (1979)). Clearly, the consumer could only benefit from these services if he purchased the beer. Moreover, Coors specifically complained about distributors who would offer poor refrigeration services and sell the beer at a discount. By imposing RPM Coors was encouraging distributors to compete in services. It could thereby guarantee a minimum quality for its beer.[3]

Other examples of sources of vertical differentiation at the retail level are: waiting time (the time between the ordering and delivery of the good); the average ratio between the number of customers and the number of sales-assistants; the provision of facilities such as (free) credit, (free) installation, (free) delivery, (free) repairs; the location of the retail outlet[4]; etc.

The chapter is organized as follows. Section 2 presents the model. Section 3 characterizes the efficient solution, which we label "vertical integration". Section 4 compares the optimal linear-price contract with vertical integration. Section 5 analyses two standard vertical restraints - franchise fee and RPM - and describes an optimal contract. Section 6 offers some concluding comments.

## SECTION 2: THE MODEL

### a) A Model of Vertical Differentiation in the Retail Market

The model of consumer choice which we use was first introduced by Gabszewicz and Thisse (1979) and subsequently used in a number of papers by the same authors and Shaked and Sutton (1982,1983).

There is a continuum of consumers represented by the unit interval [0,1]. Consumers have identical tastes, but different incomes. The income of consumer $t \in [0,1]$ is given by $E(t)$ where

$$E(t) = E.t, \qquad E > 0. \tag{1}$$

For our purposes there is no loss of generality in assuming that

$$E = 1 . \tag{2}$$

Thus income is uniformly distributed on [0,1]. Consumers are assumed to buy at most one indivisible unit of the good sold by retailers. The quality of the good is denoted by k, where $k \in [c,d]$ with $c < d$. All consumers have the same utility function $V(k,e)$, where e is the income remaining after the purchase (or non-purchase) of one unit of the good. Finally, every consumer is able to perfectly observe the quality of the good he wants to purchase. We assume that

$$V(k,e) = U(k).e \tag{3}$$

When a consumer does not purchase the commodity, his utility is given by

$$V(0,e) = U_o e, \qquad U_o > 0. \tag{4}$$

The following assumptions are made about the function U(k):

1) U is continuously differentiable.

2) $U(k) > U_o$, for all $k \in [c,d]$. That is, consumers like the good.

3) $U'(k) > 0$ for all k, that is, consumer's utility is increasing in quality.

There are only two outlets in the retail market, $R_1$ and $R_2$. We restrict the number of retailers to two, only to have a simple model and all the arguments developed below can be generalized to the case of n outlets $(n > 2)$.

We introduce another simplification by assuming that the two retailers can only offer two different levels of "quality": a low level, $k_l$ and a high level, $k_h$, with $0 < k_l < k_h$.[5] "Quality" here can be thought of as all the services that retailers can provide with the sale of the manufacturer's product. Examples of such services were given in Section 1.

It is standard in all models of pure vertical differentiation to assume that any quality level, k, can be produced at zero cost. We will make the same assumption here. The justification for this assumption is that, if in the presence of price competition one retailer will refrain from increasing the quality of his product, even though the higher quality could be produced at zero cost, then <u>a fortiori</u> he will refrain from increasing the quality of his product if the higher quality is more expensive to produce. All the essential features are present in the zero-cost case and nothing is added by introducing costs of production.

We assume that, given the wholesale price charged by the manufacturer, the retailers play a two-stage game as follows. First, $R_1$ and $R_2$ simultaneously decide which quality they want to produce. Then, having observed each other's quality, they compete in prices. That is, in the last stage of the game they simultaneously decide what mark-up to choose above the wholesale price, w, charged by the manufacturer. Thus, if $R_1$ and $R_2$ choose the same quality, they are in the classic Bertrand situation and the only equilibrium is one where they both charge a zero mark-up. The game is illustrated in Figure 1. The dashed lines connecting the decision nodes of retailer 2 express the fact that the two nodes lie in the same information set. Since both retailers choose quality simultaneously, retailer 2 when he chooses his quality level does not know whether retailer 1 chose $k_l$ or $k_h$. Furthermore, $m_l$ and $m_h$ denote, respectively, a low and a high mark-up. Both retailers choose their mark-ups simultaneously, given their observations of each other's quality choices.

An implicit assumption behind our two-stage formulation of the game is that, typically, prices can be changed much more quickly and easily than qualities. One can imagine, for example, that the quality produced by a retailer depends on what type of store he sets up. If a retailer wants to change the quality of his product, he must "set up a new store" (e.g. change the location of his store (see footnote 4); hire/lay off sales assistants; hire/lay off extra personnel for delivery, installation, repairs, increase his storing capacity in order to reduce waiting time for customers, etc.).

In order to define the payoffs of manufacturer and retailers we need to derive the demand functions faced by the retailers. Before

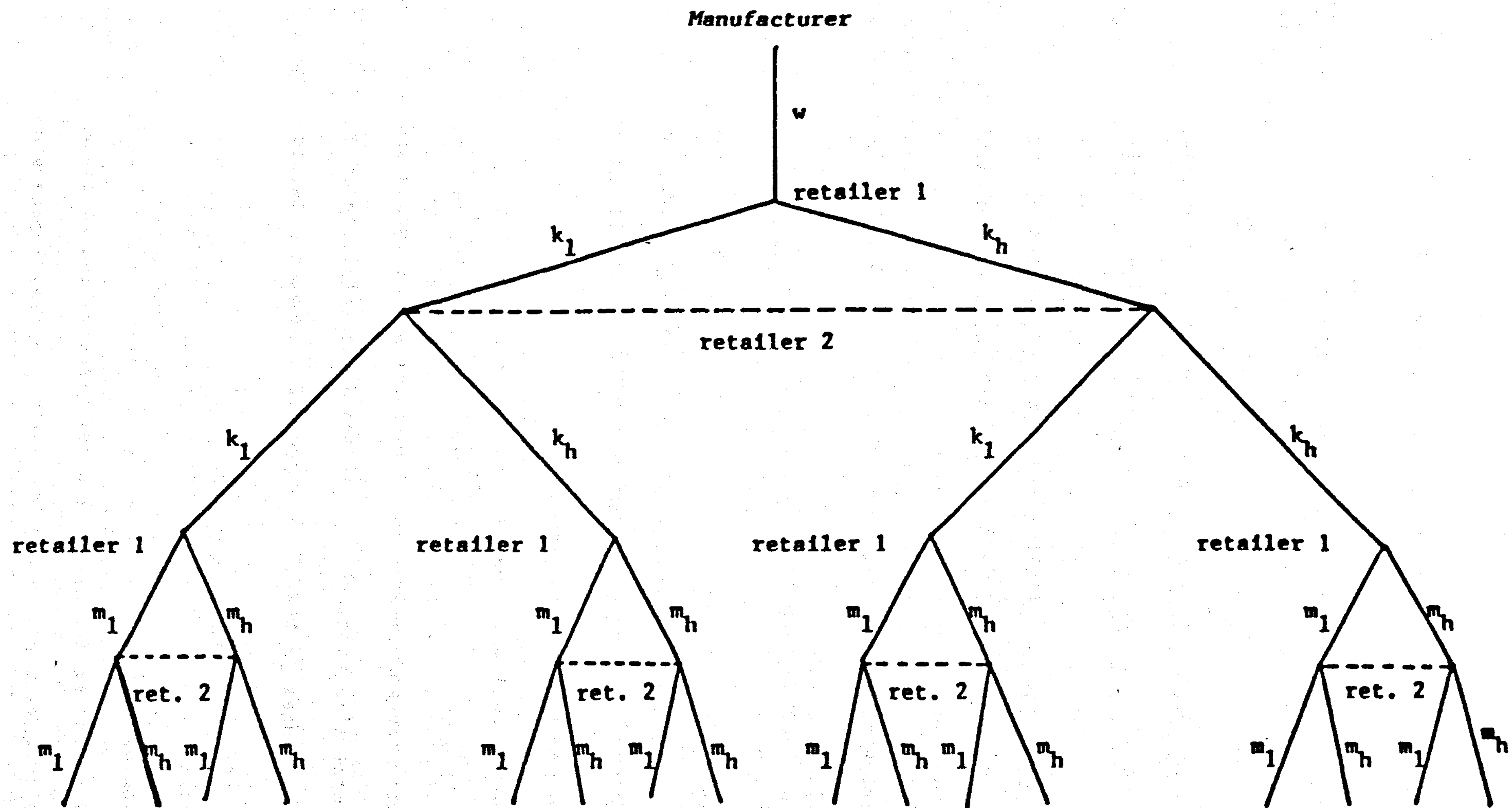

Manufacturer
w
retailer 1
$k_l$
$k_h$
retailer 2
$k_l$
$k_h$
$k_l$
$k_h$
retailer 1
retailer 1
retailer 1
retailer 1
$m_l$
$m_h$
ret. 2
$m_l$
$m_h$
$m_l$
$m_h$

FIGURE 1

doing this, however, we shall describe the wholesale market.

b) <u>The Wholesale Market</u>

There is a single firm in the upstream market (cf. footnote 3). It produces a homogeneous good at zero unit cost, and it distributes its output to the two retailers $R_1$ and $R_2$. We assume that the contract between the manufacturer and a retailer cannot be made contingent on the quality (services) chosen by the latter. This will be the case, for example, if quality is not verifiable by a court. Alternatively, we could argue that it may be prohibitively costly to fully describe the quality that a retailer is supposed to supply (of course, in our model where there are only two quality levels, this is a very strong assumption; however, in a more general model where quality is chosen from an interval [c,d] - or where, possibly, quality is multidimensional - this is a much weaker assumption).

The fact that the manufacturer cannot sign contracts which are contingent on quality would be of no consequence if he only distributed his output to one retailer. We assume, however, that this form of market-foreclosure is not in the manufacturer's interest, for the following reason. Suppose that $R_1$ and $R_2$ each have a captive market, but they also share a substantial common market (we have only formalized the common market, above). If the manufacturer only supplies one retailer he loses one of the captive markets and if this market is sufficiently large it would not be in his interest to supply only that retailer. This would not be a serious justification if the monopolist could monitor in which market each retailer sold the commodity, for then he could give the whole common market to one firm and let the other firm only supply its captive market. We assume, however, that the monopolist cannot monitor where each retailer sells the commodity.

Finally, we assume that the manufacturer has all the bargaining power; thus, he sets the contract and retailers will accept any contract which gives them non-negative profits (this assumption is common to all the existing literature).

We will now turn to the derivation of the demand functions faced by the retailers.

Suppose first that $R_1$ and $R_2$ choose the same quality level, k. Then a consumer $t \in [0,1]$ is indifferent between purchasing the commodity at price p and not purchasing it if

$$V(0, E(t)) = V(k, E(t) - p) \qquad (5)$$

or, using (1)-(4),

$$U_o t = U(k)(t - p) \qquad (6)$$

Given the retail price p and the quality k, we can define the indifferent consumer to be

$$t' = U(k)p/(U(k) - U_o) \qquad (7)$$

All consumers $t > t'$ will strictly prefer to buy the commodity and all consumers $t < t'$ will prefer not to buy it. Therefore total demand on the retail market is given by $1 - t'$ or

$$D(p,k) = 1 - \frac{U(k)}{U(k) - U_o} p \qquad (8)$$

Now suppose that $R_1$ and $R_2$ choose different quality-levels. Without loss of generality, let $R_1$ choose quality $k_h$ and $R_2$ quality $k_l$. Furthermore, let $p_1$ and $p_2$ be $R_1$ and $R_2$'s price, respectively. Also let $x = U(k_h)$ and $y = U(k_l)$.

First of all, consumer t will be indifferent between quality $k_l$ and $k_h$ if

$$V(k_l, E(t) - p_2) = V(k_h, E(t) - p_1) \qquad (9)$$

or, equivalently, if

$$y(t - p_2) = x(t - p_1) \qquad (10)$$

Since $x > y$, (10) requires $p_2 < p_1$. Let $\bar{t}$ be the indifferent consumer; then solving (10) with respect to t we obtain

$$\bar{t} = \frac{x}{x - y} p_1 - \frac{y}{x - y} p_2 \qquad (11)$$

(Note that $x > y$ and $p_2 < p_1$ imply $\bar{t} > 0$.) All $t < \bar{t}$ will prefer $k_l$ and all $t > \bar{t}$ will prefer $k_h$.

Next, define $t_o$ to be the consumer who is indifferent between buying nothing and buying the low-quality good: solving

$$V(k_l, E(t) - p_2) = V(0, E(t)) \qquad (12)$$

for t we obtain

$$t_o = \frac{y}{y - U_o} p_2 \qquad (13)$$

Then, over the relevant range[6], demand for the high-quality good and the low-quality good, denoted by $D_1(p_1,p_2)$ and $D_2(p_1,p_2)$ respectively, is given by

$$\begin{cases} D_1(p_1,p_2) = 1 - \bar{t} \\ D_2(p_1,p_2) = \bar{t} - t_o \end{cases} \tag{14}$$

Substituting (11) and (13) in (14) we obtain

$$\left[\begin{array}{l} D_1(p_1,p_2) = 1 - \dfrac{x}{x - y} p_1 + \dfrac{y}{x - y} p_2 \\ D_2(p_1,p_2) = \dfrac{x}{x - y} p_1 - \dfrac{y(x-U_o)}{(x-y)(y-U_o)} p_2 \end{array}\right. \tag{15}$$

where, as said before,

$$x = U(k_h) \quad \text{and} \quad y = U(k_1) \tag{16}$$

In the next Section we shall look at the case of vertical integration.

## SECTION 3: VERTICALLY INTEGRATED MONOPOLIST

Not surprisingly, in view of the work by Blair and Kaserman (1983), and more recently Grossman and Hart (1985), we are unable to give a satisfactory definition of vertical integration. We define a vertically integrated monopolist to be a firm which maximizes profits by choice of prices and qualities. That is, a firm which, of the two following alternatives, chooses the one which gives greater profit:

$$\max_{p} pD(p) \tag{17}$$

where D(p) is given by (8) (one quality), and

$$\max_{p_1, p_2} \; p_1 D_1(p_1, p_2) + p_2 D_2(p_1, p_2) \tag{18}$$

where $D_1$ and $D_2$ are given by (15) (distinct qualities).

Thus, vertical integration is equivalent to Pareto-efficiency among contracting parties. In most of the literature on vertical restraints, vertical integration is also defined to be equivalent to Pareto-efficiency (see Rey-Tirole (1985a)).

Lemma 1. A vertically integrated monopolist always chooses to produce both qualities at prices

$$\hat{p}_1 = \frac{2y(x - U_o)}{3xy + xU_o + y^2 - yU_o} \tag{19}$$

for the high-quality product and

$$\hat{p}_2 = \frac{(x+y)(y-U_o)}{3xy + xU_o + y^2 - yU_o} \tag{20}$$

for the low-quality product.

Proof. See Appendix.

By producing both qualities, the monopolist can discriminate between consumers with a high willingness to pay for services (that is, consumers with high income) and consumers with a low willingness to pay. The idea that quality may be used to discrinminate among consumers was first developed by Mussa and Rosen (1978).

## SECTION 4: LINEAR PRICING

The manufacturer faces many different contractual possibilities. The simplest possible contract is one where he fixes the wholesale price and sells whatever amount is demanded by retailers at that price. We will show in this Section that such a linear-price contract is strictly dominated by vertical integration. This suggests that there is a role for vertical restraints in this model. These are considered in the following Section.

Suppose that the manufacturer sets a linear-price contract for both retailers, where the wholesale price is denoted by w.[7] Let $m_1$ and $m_2$ be the mark-up chosen by retailer $R_1$ and $R_2$ respectively. Then if the two retailers choose the same quality, there will be a unique Bertrand-Nash equilibrium given by $m_1 = m_2 = 0$ (and therefore both retailers make zero profits). On the other hand, if the two retailers differentiate their products then their demand functions will be given by $D_1(m_1,m_2;w)$ and $D_2(m_1,m_2;w)$, obtained from (15) by replacing $p_1$ with $m_1 + w$ and $p_2$ with $m_2 + w$, and their profit functions will be

$$\pi_1 = m_1 D_1(m_1,m_2;w) \qquad (21)$$

and

$$\pi_2 = m_2 D_2(m_1,m_2;w) \qquad (22)$$

We can now determine the subgame-perfect equilibria of the two-stage game of Figure 1 for every possible wholesale price charged by the manufacturer.

Lemma 2. If $w < \bar{w} = (y - U_o)/(y + U_o)$, there exists a unique[8] (pure-strategy) perfect equilibrium of the two-stage game at which one retailer chooses the high-quality level, $k_h$, and the other retailer chooses the low-quality level, $k_l$, and the manufacturer's profits are given by

$$\pi_M(w) = \frac{3(x-U_o)}{4x-y-3U_o} w - \frac{3xy-xU_o-2yU_o}{(4x-y-3U_o)(y-U_o)} w^2 \qquad (23)$$

Proof. See Appendix.

Lemma 2 states the by now well-known result that (if the wholesale price is not too high) retailers want to "relax price competition through product differentiation".

When the manufacturer charges a very high wholesale price ($w \geq \bar{w}$), one of the retailers ($R_2$) cannot produce $k_l$ and at the same time earn positive profits. Then both retailers will produce the same quality-level, $k_h$, and make zero profits. This is established in the lemma below.

Lemma 3. If $w \geq \bar{w}$ there exists a unique perfect equilibrium of the two-stage game, where both retailers choose the high-quality level, their profits are zero and the manufacturer's profits are given by

$$\hat{\pi}_M(w) = w\left(1 - \frac{x}{x-U_o} w\right) \qquad (24)$$

Proof. See Appendix.

The question now arises of what the optimal wholesale price is for the manufacturer. The next lemma shows that depending on the parameter values, x, y and $U_o$, the optimal wholesale price may or may not induce product differentiation.

Lemma 4. The optimal wholesale price is less than $\bar{w}$ - and, therefore, the optimal linear-price contract induces product differentiation - if and only if y is sufficiently close to x; more precisely, if and only if the following inequalities are simultaneously satisfied:

$$\begin{cases} (x - U_o)(y + U_o) < 2x(y - U_o) \\ 9(x - U_o)^3(y + U_o)^2 > 4U_o(2x - y - U_o) \\ \qquad (4x - y - 3U_o)(3xy - xU_o - 2yU_o) \end{cases} \tag{25}$$

Proof. See Appendix.

Given Lemmas 1 to 4,[9] it is straightforward to establish the following Proposition.

Proposition 1. A linear-price contract is strictly dominated by vertical integration.

Proof. We have to distinguish two cases:

(i) If (25) is satisfied, then retailers differentiate and make strictly positive profits; it follows that vertical integration dominates the optimal linear-price contract;

(ii) If (25) is not satisfied then retailers do not differentiate and make zero profits. By Lemma 1 we know that it is not optimal for a vertically integrated monopolist to produce only one quality.

Proposition 1 suggests that there is a role for vertical restraints in this model. We will show in the next Section that, indeed, both resale-price-maintenance and franchise-fee contracts dominate the optimal linear-price contract.

## SECTION 5: RESALE-PRICE-MAINTENANCE AND FRANCHISE-FEES

Usually, resale-price-maintenance (RPM) is defined to be a provision in the contract, restricting the retailer's choice of the final price. The most commonly observed restriction is a retail-price-floor, but some manufacturers also use retail-price-ceilings. In our model, since there is no uncertainty, there is no need for the manufacturer to set a retail-price-floor which is different from a retail-price-ceiling; thus, when he signs an RPM contract, the manufacturer is in fact setting the retail price.

In most of the literature on vertical restraints, a franchise fee is defined to be simply a fixed payment from the retailer to the manufacturer. Then a combination of a franchise fee and a linear price gives rise to the simplest form of non-linear pricing.

We do not consider two other standard forms of vertical restraints: exclusive territories and quantity-fixing. To use exclusive territories amounts to giving the whole common market to one retailer. We have assumed, however, that this is not feasible since the manufacturer cannot monitor on which market (captive or common) the retailer sells his commodity.[10] Quantity-fixing restraints are also not feasible for two reasons: first, demand on each market may be stochastic; secondly, arbitrage between retailers may limit the possibility of quantity-fixing (especially if there are several retail markets such as this one).

Having defined the various forms of vertical restraints considered in this paper, we will briefly discuss how these restraints are enforced. The problem of enforceability arises mainly with RPM. How can the manufacturer be certain that a given final price has been charged by a retailer to his customer? In practice, mainly two ways have been used by manufacturers to find out about the retail price (these methods have been used, for example, in the audio-components-industry, see McEachern-Romeo (1984)). The first method was to use warranty cards to verify the resale price. The second method was to hire private investigators (so called "shoppers") who would visit the different outlets and verify whether the retailer was setting the prescribed price.

We now show that RPM and franchise-fee contracts dominate the optimal linear-price contract, but do not achieve the efficient outcome.

Proposition 2. RPM and franchise-fee contracts dominate the optimal linear-price contract.

The proof of this proposition is straightforward:

(i) By setting the resale-price, the manufacturer prevents retailers from "relaxing price competition through product differentiation"; retailers are forced to choose the same resale-price, so that they can only compete in quality. Consequently, both retailers will choose the high quality, $k_h$, in equilibrium. The manufacturer can then set the retail price equal to the wholesale price and extract all the retailers' profits. By setting an RPM contract, the manufacturer cannot discriminate among consumers (since both retailers produce the same quality); he can, however, extract all of the retailers' rent. It is shown in the Appendix (see proof of Lemma 4, in particular (A.20)) and in Footnote 11, that this second effect always dominates the first effect, so that RPM dominates the optimal linear-price contract.[11]

(ii) To add a franchise fee to a linear-price contract can only make the manufacturer better off, since whenever retailers make positive profits with a linear-price contract, the manufacturer can use a franchise fee to extract part or all of these profits.

In most of the existing models of vertical restraints where the number of retailers is fixed exogenously, RPM or franchise fees achieve the same outcome as vertical integration. This is not the case in our model. We estbalish this claim in the next Proposition.

Proposition 3. Franchise-fee contracts and RPM contracts are strictly dominated by the vertical integration outcome.

The proof that vertical integration dominates RPM is straightforward: with RPM retailers do not differentiate and thus the manufacturer cannot discriminate among consumers. It is less obvious why a franchise-fee contract is dominated by vertical integration. If the manufacturer charged the same franchise fee to both retailers, he would be unable to extract all the high-quality retailer's profits, since the latter makes strictly greater profits than the low-quality retailer. On the other hand, the franchise fee cannot be a function of quality, because, by assumption, quality is not verifiable by the courts. One could argue, however, that prices could be used to infer quality and, as a consequence, a double franchise fee could be used to extract all of the retailers' profits:

the retailer with the higher price is obviously producing a higher quality and can therefore be charged a higher franchise fee. However, even though this contract would represent an improvement on the simple franchise fee, it would not bring about efficiency. The reason is that the manufacturer must circumvent two problems here: double-marginalization and price-competition among retailers. It is well-known that in order to eliminate double-marginalization with a franchise-fee contract, the manufacturer must set the wholesale price equal to marginal cost (i.e. set $w = 0$) and then set the franchise fee equal to the retailer's profit. In our model, however, when $w = 0$ the price-competition effect outweighs the double-marginalization effect. In other words, when $w = 0$ retailers set their respective mark-ups, $m_1$ and $m_2$, below the monopoly prices. To see this, it suffices to compare the optimal prices chosen by retailers when the wholesale price is $w = 0$ with the prices chosen by the vertically integrated monopolist (given by (19) and (20)). Retailers choose $m_1$ and $m_2$ to maximize (21) and (22) respectively. The optimal mark-ups $m_1^*(w)$ and $m_2^*(w)$ when $w = 0$ are given by

$$m_1^*(0) = \frac{2(x - y)(x - U_o)}{x(4x - y - 3U_o)} \qquad (26)$$

$$m_2^*(0) = \frac{(x - y)(y - U_o)}{y(4x - y - 3U_o)} \qquad (27)$$

It is easy to check that (26) is strictly less than (19) and (27) is strictly less than (20). Furthermore, it can be shown that

$$\frac{d}{dw}\,(m_1^*(w) + w) > 0 \quad \text{and} \quad \frac{d}{dw}\,(m_2^*(w) + w) > 0.$$

Thus, the manufacturer can force retail prices up by increasing $w$ from zero and the optimal franchise-fee contract will have $w > 0$. It can be shown, however, that there does not exist any $w \geq 0$ such that both (28) and (29) below are satisfied (where $\hat{p}_1$ and $\hat{p}_2$ are given by (19) and (20), respectively)

$$m_1^*(w) + w = \hat{p}_1 \tag{28}$$

$$m_2^*(w) + w = \hat{p}_2 \tag{29}$$

Thus this form of franchise-fee contract cannot achieve full efficiency.

The above discussion suggests that an optimal contract ought to restrict the set of prices which can be chosen by retailers. This can be done by means of a price-dependent franchise fee, as shown in the following Proposition.

<u>Proposition 4</u>. The manufacturer can approximate the outcome of vertical integration arbitrarily closely by fixing the wholesale price

$$w = \hat{p}_2 \tag{30}$$

(where $\hat{p}_2$ is given by (20)) and the following franchise fee:

$$A = \begin{cases} +\infty & \text{if retail price} \neq \hat{p}_1 \text{ or } w \\ 0 & \text{if both retail prices are equal to } w \end{cases} \tag{31}$$

and otherwise

$$A = \begin{cases} -\varepsilon & \text{if retail price} = w \\ (\hat{p}_1 - w)D_1(\hat{p}_1,\hat{p}_2) - \varepsilon & \text{if retail price} = \hat{p}_1 \end{cases} \tag{32}$$

where $\hat{p}_1$ is given by (19), $D_1$ is given by (15) and $\varepsilon > 0$ is arbitrarily small.

Proof. See Appendix.

The intuition behind Proposition 4 is as follows. The clause "$A = +\infty$ if the retail price is different from $\hat{p}_1$ or w" has the purpose of restricting the set of retail prices which can be chosen by retailers to $\{\hat{p}_1, \hat{p}_2\}$, that is, to the price which would be chosen by a vertically integrated monopolist (thereby eliminating the problem discussed above, namely that price competition between retailers leads to retail prices which are "too low"). Secondly, the contract must ensure that retailers prefer to differentiate rather than produce the same quality. This is achieved by (32) which ensures them an equal profit of $\varepsilon$ if they differentiate and negative profits if they both choose high quality. Finally, the clause "$A = 0$ if both retail prices are equal to w" ensures that retailers will not make positive profits if they both choose low quality.

The optimal contract described in Proposition 4 - and any contract in which <u>all</u> retail prices are taken into account - may be extremely costly to enforce, especially when the number of retailers is large (which is usually the case). In order for such a contract to be enforceable, all final prices must be observable to all parties and this may be impossible or prohibitively costly. Therefore a simpler - although suboptimal- contract may be preferable. For instance, there are situations in which RPM is a good substitute for an optimal contract: when y is close to x, resale price maintenance strictly dominates the optimal linear-price contract and yields an outcome which is close to that of vertical integration. In fact, by Lemma 4 (see also footnote 11), we know that when y is close to x the optimal linear-price contract induces quality-differentiation and is strictly dominated by RPM (cf. (A.20) in the Appendix). Furthermore, using (A.3) and (A.6) in the Appendix, the difference between the profits of the vertically integrated monopolist and the profits of the monopolist who uses RPM is given by

$$p_1q_1 + p_2q_2 - p^*q^* = \frac{(x - U_o)(y - U_0)(x - y)}{4x(3xy + xU_o + y^2 - yU_o)} \qquad (33)$$

which tends to zero as y tends to x (it also tends to zero as y tends to $U_o$, but we know from Footnote 11 that when y is close to $U_o$ the optimal linear-price contract and RPM are equivalent). Furthermore, a contract like RPM can be enforced at low cost: all that is required for enforceability is for the manufacturer to observe prices with positive probability. By increasing the penalty for not selling at the required retail price, the manufacturer can reduce his probability of inspecting prices (i.e. reduce the number of "shoppers") and thus reduce his enforcement costs.

As a final remark we shall stress the fact that we have only considered the case where there is no uncertainty. As Rey-Tirole (1985b) pointed out, uncertainty about demand or retail costs may significantly affect the desirability and optimality of the various standard forms of vertical restraints.

## SECTION 6: CONCLUSION

We have shown that when the "quality" of the product distributed by retailers depends on some effort or services provided - together with the product - by the retailers themselves, the manufacturer will find it profitable to resort to vertical restraints in order to bring the retailers' choices in line with his interest. Simple forms of vertical restraints, such as resale price maintenance and simple franchise fees, are not sufficient to fully compensate for the "distortionary effects" of price competition between retailers. In order to achieve the first-best, that is, the outcome of vertical integration, the manufacturer has to resort to more sophisticated contracts which include lump-sum payments from retailers to the manufacturer or vice versa, depending on the prices charged by all retailers.

The range of applicability of our model could be extended considerably if we allowed for competition in the manufacturers' market (cf. footnote 3). The analysis of the audio components industry by McEachern and Romeo (1984) suggests that such a model might be applicable to this industry. Also the Magnavox case with the FTC would fit into such a model: Magnavox was the third largest seller of colour televisions and was using RPM to reach the consumers interested in expensive sets; RPM was necessary since the sale of expensive sets required a lot of pre-sale and post-sale effort on the part of retailers (see Goldberg (1982)).

FOOTNOTES

(1) The three main explanations reviewed below are not exhaustive: more explanations can be found in the literature (see the survey by Rey-Tirole (1985a)).

(2) On page 21 the authors observe that "The existing models are usually location models in a homogeneous space (e.g., a circle with a uniform density of consumers). The "principle of maximum differentiation" holds for both competing retailers and a vertically integrated structure. Thus there is no conflict between the manufacturer and a fixed number of retailers as to the latters' locations."

(3) The Coors beer case is complicated by the fact that the manufacturer was competing with brands manufactured by other firms. In this paper we have followed the existing literature on vertical restraints and restricted ourselves to the case of one manufacturer dealing with many retailers. Allowing for competition among manufacturers complicates the analysis considerably. Some of the issues arising in this context have been analysed by Bonanno-Vickers (1986). The most general case with many manufacturers and many retailers, each of which is allowed to carry the goods of several manufacturers, can be analysed as a multi-principal, multi-agent situation (see Bernheim-Whinston (1986)) and would enable one to analyse several new issues.

(4) As an example consider the following case, in the same spirit as Hotelling's (1929) model. A homogeneous good can be sold at

(4) continued

any point on a line segment, which we take to be the unit interval [0,1]. There are many consumers, all located at the right extreme of the line segment. Consumers have different incomes but identical preferences: in particular, they all dislike travelling. The utility a consumer derives from not consuming the good and keeping his income e is given by (4) of Section 2. On the other hand, if he consumes one unit of the good and an income e and has to travel a distance d to obtain the good, his utility is $V(1,e,d) = U_1 e f(d)$, where $-f(d)$ is disutility of travelling and therefore $f'(d) < 0$. In this example, $d = 1 - k$, where $k \in [0,1]$ denotes the location of the shop. We can then write $V(k,e) = U_1 e f(1-k)$ and define $U(k) = U_1 f(1-k)$. Clearly, $U'(k) > 0$. The parameter k (location) can now be interpreted as an index of quality and, provided $U_1 f(1) > U_o$, we have an example of the general model of Section 2. Alternatively, one could imagine the consumers being spread out over the segment (rather than bunched at one extremity) while the firms are established on the same side of the segment, outside of the market.

(5) In the model by Shaked and Sutton (1982) firms are allowed to choose any quality level $k \in [c,d]$ $(0 < c < d)$. They show that in equilibrium one firm always chooses $k = d$ and they do not explicitly derive the quality chosen by the other firm. Here, the retailers' choice is restricted to $k_l$ or $k_h$. We will explain, in Section 4, that this simplification of the model eliminates an important aspect of the conflict between manufacturer and retailers, namely that retailers would not choose the same quality levels as a vertically integrated manufacturer. We were

(5) continued

forced to simplify the model in this way in order to keep the calculations tractable!

(6) More precisely, the demand functions are given as follows. By (10), $D_2 = 0$ if $p_2 \geq p_1$. Also, $D_2 = 0$ if $p_2 \geq (y - U_o)/y$, where the RHS is the reservation price of the richest consumer for the low-quality good. Thus (15) requires $p_2 < \min\{p_1, (y-U_o)/y\}$.

Similarly, $D_1 = 0$ if $p_1 \geq (x-U_o)/x$, where the RHS is the reservation price of the richest consumer for the high-quality good. Also, $D_1 = 0$ if $p_2 < p_1$ and $\bar{t} \geq 1$, where $\bar{t}$ is given by (11); this is equivalent to $p_1 \geq (x-y)/x + yp_2/x$. If $p_2 \geq p_1$ and $p_1 < (x-U_o)/x$, then $D_1 = 1 - xp_1/(x-U_o)$. Thus (15) requires $p_2 < p_1$ and $p_1 < \min\{(x-y)/x + yp_1/x,\ (x - U_o)/x\}$. It can be shown that there is no loss of generality in restricting oneself to the price range for which (15) is satisfied.

(7) The manufacturer sets the same wholesale price to both retailers, since at the time of contracting he does not know what quality level will be chosen by each retailer.

(8) The pure-strategy equilibrium is unique from the point of view of the manufacturer (which is the point of view we are interested in), since it does not matter to him which retailer chooses the high quality and which chooses the low quality.

We ought to add that there is also a symmetric mixed-strategy equilibrium at which each retailer chooses the high-quality level with probability $\mu = \pi_1^*/(\pi_1^* + \pi_2^*)$, where $\pi_1^*$ and $\pi_2^*$ are the profits of the high-quality and low-quality retailer at the pure-strategy equilibrium, respectively. In the mixed-strategy equilibrium the manufacturer's profits are given by

(8) continued

$$2\mu(1 - \mu)\pi_M(w) + \mu^2(w - xw^2/(x - U_o)) +$$

$$+ (1 - \mu^2)(w - yw^2/(y - U_o)).$$

We shall follow the literature (see, for example, Shaked-Sutton (1982)) and restrict our analysis to the pure-strategy equilibrium.

(9) A quick comparison of Lemmas 1 and 4 suggests that there is a potential conflict between manufacturer and retailers as to the choice of quality differentiation. This turns out to be indeed true. In a more general model where retailers can choose quality in an interval [c,d], they would, in general, choose different quality levels from the ones chosen by a vertically integrated monopolist.

To choose two quality levels in [c,d] is equivalent to choosing x and y in [U(c),U(d)]. Shaked and Sutton (1982) show that one retailer would choose x = U(d), while the other retailer would choose y in [U(c),U(d)] to maximise

$$\pi_2^* = \frac{y(x-U_o)}{(y-U_o)(x-y)}\left[\frac{(x-y)(y-U_o)}{y(4x-y-3U_o)} - x\,\frac{(y+U_o)(x-y)}{y(4x-y-3U_o)}\right]^2$$

The integrated monopolist, on the other hand, chooses x and y in [U(c),U(d)] to maximise $\hat{p}_1\hat{q}_1 + \hat{p}_2\hat{q}_2$ given by

$$\frac{y(x - U_o)}{3xy + xU_o + y^2 - yU_o}$$

It is tedious, but straightforward, to check that the two problems do not yield the same first-order conditions: thus the optimal quality choices of a vertically integrated monopolist will be different from the quality choices of the two retailers.

(10) If the manufacturer could monitor where the retailer sold his commodity, then a combination of exclusive territories and franchise fees would lead to an efficient outcome.

(11) We know from Lemma 4 that the optimal linear-price contract induces quality-differentiation if and only if (25) is satisfied. This does not imply that if (25) is <u>not</u> satisfied then the optimal linear price contract is equivalent to RPM. In fact, if (25) is not satisfied but

$$(x - U_o)(y + U_o) < 2x(y - U_o)$$

(cf. (A.27) in the Appendix) then the optimal linear-price contract is $w = \bar{w} > \hat{w}^*$ and, by definition of $\hat{w}^*$, the manufacturer is strictly better off if he resorts to RPM and fixes the retail price equal to the wholesale price equal to $\hat{w}^*$.

Therefore, RPM does not <u>strictly</u> dominate the optimal linear-price contract if and only if

$$(x - U_o)(y + U_o) \geq 2x(y - U_o)$$

(which is the case if y is close to $U_o$).

## APPENDIX

Proof of Lemma 1. First consider the profit function given by (18). It is straightforward to show that it is strictly concave in $(p_1, p_2)$ and solving the first-order conditions we obtain (19) and (20). The corresponding high-quality and low-quality outputs are given by

$$\hat{q}_1 = D_1(\hat{p}_1, \hat{p}_2) = \frac{x(y + U_o)}{3xy + xU_o + y^2 - yU_o} \tag{A.1}$$

and

$$\hat{q}_2 = D_2(\hat{p}_1, \hat{p}_2) = \frac{y(x - U_o)}{3xy + xU_o + y^2 - yU_o} \tag{A.2}$$

and the corresponding profits are given by

$$\hat{p}_1\hat{q}_1 + \hat{p}_2\hat{q}_2 = \frac{y(x - U_o)}{3xy + xU_o + y^2 - yU_o} \tag{A.3}$$

Now consider the profit function given by (17) with $k = k_h$ (it is obvious that the monopolist would not choose the low quality). Solving the first-order condition we obtain that the optimal price is given by

$$p^* = \frac{x - U_o}{2x} \tag{A.4}$$

and the corresponding output and profits are

$$q^* = 1/2 \tag{A.5}$$

$$p^*q^* = \frac{x - U_o}{4x} \tag{A.6}$$

Now, (A.3) > (A.6) if and only if

$$4xy > 3xy + xU_o + y^2 - yU_o \tag{A.7}$$

which is equivalent to

$$x(y - U_o) > y(y - U_o) \tag{A.8}$$

which is obviously true.

Proof of Lemma 2. Fix an arbitrary $w < \bar{w}$. By Bertrand's theorem, if both retailers choose $k = k_h$ or $k = k_1$, there is a unique Bertrand-Nash equilibrium of the second-stage game given by $m_1 = m_2 = 0$. On the other hand, if retailer $R_1$ chooses $k_h$ and retailer $R_2$ chooses $k_1$, then the profit functions of the two retailers are given by (21) and (22), respectively. $\pi_i$ is strictly concave in $m_i$ $(i = 1,2)$ and from the first-order conditions we obtain a unique Nash equilibrium in prices at which the two retailers' mark-ups are given by

$$m_1^* = \frac{2(x-U_o)(x-y)}{x(4x-y-3U_o)} - w\,\frac{(2x-U_o)(x-y)}{x(4x-y-3U_o)} \tag{A.9}$$

and

$$m_2^* = \frac{(x-y)(y-U_o)}{y(4x-y-3U_o)} - w\,\frac{(y+U_o)(x-y)}{y(4x-y-3U_o)} \tag{A.10}$$

respectively. The corresponding equilibrium profits are given by

$$\pi_1^* = \frac{x}{x-y}\left[\frac{2(x-y)(x-U_o)}{x(4x-y-3U_o)} - w\,\frac{(x-y)(2x-U_o)}{x(4x-y-3U_o)}\right]^2 \tag{A.11}$$

and

$$\pi_2^* = \frac{y(x-U_o)}{(y-U_o)(x-y)}\left[\frac{(x-y)(y-U_o)}{y(4x-y-3U_o)} - w\,\frac{(y+U_o)(x-y)}{y(4x-y-3U_o)}\right]^2 \tag{A.12}$$

respectively. It can be checked that $m_2^* > 0$ (and thus $\pi_2^* > 0$) if and only if

$$w < \bar{w} = \frac{y - U_o}{y + U_o} \tag{A.13}$$

Also, $\pi_2^* \geq 0$ implies $\pi_1^* > \pi_2^*$. Thus, as long as $\pi_2^* > 0$, the two retailers will differentiate. It follows that if $w < \bar{w}$, there are two pure-strategy equilibria, one where $R_1$ chooses $k_h$ and $R_2$ $k_l$ and the other where $R_1$ chooses $k_l$ while $R_2$ chooses $k_h$. It is not surprising, then, that there is also a symmetric mixed-strategy equilibrium, which is described in footnote 8 (in the mixed-strategy equilibrium both retailers choose the same quality with probability $\mu^2$ and $(1 - \mu)^2$, then the demand function they face is given by (8); this explains the expression for the manufacturer's profits given in footnote 8).

Next, if $w < \bar{w}$ and if retailers are at a pure-strategy equilibrium, the manufacturer's profits are given by

$$\pi_M(w) = (q_1^* + q_2^*)w \tag{A.14}$$

where $q_1^*$ and $q_2^*$ are the equilibrium outputs of the retailers. $\pi_M(w)$ is then obtained through straightforward computation.

<u>Proof of Lemma 3</u>. By (A.13), $w \geq \bar{w}$ implies $m_2^* \leq 0$. Thus $R_2$ will choose high quality and the unique perfect equilibrium is where both retailers produce the high-quality good and $m_1 = m_2 = 0$. The profit function of the manufacturer is then obtained from (8) with $k = k_h$.

<u>Proof of Lemma 4</u>. Let $w^*$ be the (unique) maximum of the function $\pi_M(w)$ (given by (23)). Then

$$w^* = \frac{3(x - U_o)(y - U_o)}{2(3xy - xU_o - 2yU_o)} \tag{A.15}$$

Also

$$\pi_M(w^*) = \frac{9(x - U_o)^2(y - U_o)}{4(4x - y - 3U_o)(3xy - xU_o - 2yU_o)} \qquad (A.16)$$

Let $\hat{w}^*$ be the (unique) maximum of the function $\hat{\pi}_M(w)$ (given by (24)). Then

$$\hat{w}^* = (x - U_o)/(2x) \qquad (A.17)$$

Also

$$\hat{\pi}_M(\hat{w}^*) = (x - U_o)/(4x) \qquad (A.18)$$

The following facts can be checked easily: (A.15) is less than (A.17) or

$$w^* < \hat{w}^* \qquad (A.19)$$

and (A.16) is less than (A.18) or

$$\pi_M(w^*) < \pi_M(\hat{w}^*). \qquad (A.20)$$

It follows that as soon as $\hat{w}^* \geq \bar{w}$, the manufacturer will want to set $w = \hat{w}^*$, thereby inducing retailers <u>not</u> to differentiate. Now, $\hat{w}^* \geq \bar{w}$ if and only if $(x-U_o)(y+U_o) \geq 2x(y-U_o)$ (A.21)

for which it is necessary that y be <u>not</u> close to x. On the other hand, if $\hat{w}^* < \bar{w}$, i.e. if

$$(x - U_o)(y + U_o) < 2x(y - U_o) \qquad (A.22)$$

(which is the case if y is close to x, since when $y = x$ (A.22) becomes $2x > x + U_o$, which is obviously true), then by (A.19) also $w^* < \bar{w}$ and the manufacturer's choice will depend on whether

$$\pi_M(w^*) \gtrless \hat{\pi}_M(\bar{w}) \qquad (A.23)$$

(recall that $\hat{\pi}_M(w)$ is strictly concave and therefore, given that $\hat{w}^* < \bar{w}$, the best wholesale price in $[\bar{w}, +\infty)$ is $\bar{w}$). Now,

$$\hat{\pi}_M(\bar{w}) = \frac{U_o(y - U_o)(2x - y - U_o)}{(x - U_o)(y + U_o)^2} \tag{A.24}$$

Thus

$$\pi_M(w^*) > \hat{\pi}_M(\bar{w}) \tag{A.25}$$

if and only if

$$\frac{9(x-U_o)^2}{4(4x-y-3U_o)(3xy-xU_o-2yU_o)} > \frac{U_o(2x-y-U_o)}{(x-U_o)(y+U_o)^2} \tag{A.26}$$

which is the case if y is close to x, since when y = x (A.26) becomes $(x-U_o)^2 > 0$, which is obviously true. (A.22) and (A.26) prove (25).

To sum up, if (25) is satisfied, the optimal wholesale price is $w^*$ (given by (A.16)) which is less than $\bar{w}$ and therefore, by Lemma 2, induces product differentiation. If (25) is not satisfied, then the optimal price will be

$$\begin{cases} \hat{w}^* = (x-U_o)/(2x) & \text{if} \quad (x-U_o)(y+U_o) \geq 2x(y-U_o) \\ \bar{w} = (y-U_o)/(y+U_o) & \text{if} \quad (x-U_o)(y+U_o) < 2x(y-U_o) \end{cases} \tag{A.27}$$

and since in both cases $w \geq \bar{w}$, retailers will not differentiate.

Proof of Proposition 4. We first determine the Bertrand-Nash equilibria of the second stage of the game of Figure 1. First of all, it is clear that the franchise fee given by (31) restricts the set of prices which can be chosen by retailers to $\{\hat{p}_1, \hat{p}_2\}$.

Lemma (i). If the two retailers choose different qualities, there is a unique Nash equilibrium (N.E.) at which the high-quality retailer's price is $\hat{p}_1$ and the low-quality retailer's price is $\hat{p}_2 = w$ and both retailers make positive profits (given by $\varepsilon$).

Proof. First we show that $(\hat{p}_1,\hat{p}_2)$ is a N.E.: the low-quality retailer's profits are $\varepsilon > 0$, while if he switched to $\hat{p}_1$ he would face zero demand and pay a positive franchise fee (if $\varepsilon$ is sufficiently small). The high-quality retailer's profits are $\varepsilon > 0$, while if he switched to $\hat{p}_2 = w$ his profits would be zero. Next we show that $(\hat{p}_1,\hat{p}_1)$ is not a N.E.: the low-quality retailer faces zero demand and pays a positive franchise fee, while he can increase his profits to $\varepsilon > 0$ by switching to $\hat{p}_2$. Similarly, $(\hat{p}_2,\hat{p}_2)$ is not a N.E. because the high-quality retailer can increase his profits from zero to $\varepsilon > 0$ by switching to $\hat{p}_1$.

Lemma (ii). If both retailers choose high quality, $k_h$, there is a unique N.E. where they both charge the same retail price, $\hat{p}_2 = w$, and make zero profits.

Proof. $(\hat{p}_2,\hat{p}_2)$ is a N.E. because both retailers make zero profits and if either retailer switched to $\hat{p}_1$ he would face zero demand and pay a positive franchise fee (if $\varepsilon$ is sufficiently small). $(\hat{p}_1,\hat{p}_2)$ is not a N.E. because the retailer with the higher price pays a positive franchise fee and faces zero demand, while he can make zero profits by switching to $\hat{p}_2$. $(\hat{p}_1,\hat{p}_1)$ is not a N.E. because each retailer's profits are

$$(1/2)(\hat{p}_1 - w)(\hat{p}_1,k_h) - A = (1/2)(\hat{p}_1 - w)(\hat{p}_1,k_h) - - (\hat{p}_1 - w)D_1(\hat{p}_1,\hat{p}_2) + \varepsilon \qquad (A.28)$$

Now, using (8) and (19), we have

$$\frac{1}{2} D(\hat{p}_1,k_h) = \frac{1}{2} - \frac{1}{2(x-U_o)} \frac{2y(x-U_o)}{(3xy+xU_o+y^2-yU_o)} \qquad (A.29)$$

while $D_1(\hat{p}_1,\hat{p}_2)$ is given by (A.1). It is easy to show that (A.1) > (A.29) and therefore if $\varepsilon$ is sufficiently small both retailers are making negative profits while either retailer could make zero profits by switching to $\hat{p}_2$.

<u>Lemma (iii)</u>. If both retailers choose low quality, $k_1$, there is a unique N.E. where they both charge the same retail price, $\hat{p}_2 = w$, and make zero profits.

<u>Proof</u>. $(\hat{p}_2,\hat{p}_2)$ is a N.E. because if either retailer switched to $\hat{p}_1$ he would make negative profits. $(\hat{p}_1,\hat{p}_2)$ is not a N.E. because the retailer with the higher price makes negative profits. Finally, $(\hat{p}_1,\hat{p}_1)$ is not a N.E. because each retailer's profit is given by

$$(1/2)(\hat{p}_1 - w)D(\hat{p}_1,k_1) - A = (1/2)(\hat{p}_1 - w)D(\hat{p}_1,k_1) - (\hat{p}_1 - w)D_1(\hat{p}_1,\hat{p}_2) + \varepsilon \tag{A.30}$$

Now, using (8) and (19) we have

$$\frac{1}{2} D(\hat{p}_1,k_1) = \frac{1}{2} - \frac{y}{2(y-U_o)} \frac{2y(x-U_o)}{(3xy+xU_o+y^2-yU_o)} \tag{A.31}$$

while $D_1(\hat{p}_1,\hat{p}_2)$ is given by (A.1). Again, it is easy to show that (A.1) > (A.31) and therefore if $\varepsilon$ is sufficiently small each retailer makes negative profits.

By Lemmas (i)-(iii) we can conclude that there is a unique perfect equilibrium of the two-stage game at which one retailer chooses high quality and charges $\hat{p}_1$ and the other retailer chooses low quality and charges $\hat{p}_2$ and each retailer's profits are $\varepsilon > 0$. This is the choice of qualities and prices of the vertically integrated monopolist and therefore by choosing $\varepsilon$ arbitrarily small the manufacturer can approximate the outcome of vertical integration.

# REFERENCES

ALCHIAN, A. and H. DEMSETZ (1972) "Production, information costs, and economic organisation", American Economic Reivew, 62 (December): 777-95.

BAIMAN, S. (1982) "Agency research in managerial accounting: a survey", Journal of Accounting Literature, 1: 154-213.

BAIMAN, S. and J. DEMSKI (1980a) "Variance Analysis Procedures as Motivation Devices", Management Science, Vol. 26, No. 8: 840-848.

BAIMAN, S. and J. DEMSKI (1980b) "Economically Optimal Performance Evaluation and Control Systems", Journal of Accounting Research, 18: 184-220.

BAIN, J. (1956) "Barriers to New Competition", Cambridge, Harvard University Press.

BARON, D. and D. BESANKO (1984) "Regulation, asymmetric information and auditing", Rand Journal of Economics, Vol. 15, No. 4: 447-470.

BAUMOL, W.J. (1959) "Business Behaviour, Value and Growth", Macmillan, New York.

BECKER, G.S.(1968) "Crime and Punishment: An Economic Approach", Journal of Political Economy, Vol. 66: 169-217.

BERNHEIM, B.D. and M.D. WHINSTON (1986) "Common Agency", Econometrica, forthcoming.

BLAIR, R.D. and D.L. KASERMAN (1983) "Law and Economics of Vertical Integration and Control", Academic Press, New York.

BONANNO, G. and J.S. VICKERS (1986) "Vertical separation", mimeo, Nuffield College, Oxford.

BORK, R. (1978) "The Antitrust Paradox", New York, Basic Books.

BULOW, J. (1982) "Durable goods monopolists", Journal of Political Economy, 314-332.

CAILLAUD, B., R. GUESNERIE, P. REY and J. TIROLE (1984) "The Normative Economics of Government Intervention in Production in the light of Incentives Theory: A Review of Recent Contributions", mimeo, IMSSS, Stanford University.

CALVO, G. and S. WELLICZ (1978) "Supervision, Loss of Control, and the Optimal Size of the firm", Journal of Political Economy, 943-952.

CAVES, R. (1984) "Vertical Restraints in Manufacturer-Distribution Relations: Incidence and Economic Effects", mimeo, Harvard University.

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

CYERT, R.M. and J.G. MARCH (1963) "A Behavioural Theory of the firm", Prentice-Hall, Englewood Cliffs, N.J.

DASGUPTA, P. and J. STIGLITZ (1984) "SUnk costs and Competition", mimeo, Princeton University.

DEMSKI, J. and D. KREPS (1982) "Models in Managerial Accounting", Journal of Accounting Research, 20: 117-148.

DIAMOND, P.A. and E. MASKIN (1979) "An Equilibrium ANalysis of Search and Breach of Contract, I: Steady States", Bell Journal of Economics, Vol. 10, No. 1: 282-316.

DIXIT, A. (1979) "A model of duopoly suggesting a Theory of Entry Barriers", Bell Journal of Economics, Vol. 10, No. 1: 20-32.

DIXIT, A. (1980) "The Role of Investment in Entry-deterrence", Economic Journal, Vol. 90: 95-106.

DIXIT, A. (1983) "Vertical integration in a monopolistically competitive industry", International Journal of Industrial Organisation, 1: 63-78.

FAMA, E. (1980) "Agency Problems and the Theory of the firm", Journal of Political Economy, 88: 288-306.

FRIEDMAN, M. (1953) "The Methodology of positive Economics", in Essays in Positive Economics, Chicago: University of Chicago Press.

FUDENBERG, D. and J. TIROLE (1984) "Dynamic Models of Oligopoly", mimeo, MIT.

GABSZEWICZ, J.J. and J.F. THISSE (1979) "Price competition, quality and income disparities", Journal of Economic Theory, 22: 340-359.

GALE, D. and M. HELLWIG (1985) "Incentive Compatible Debt Contracts: The One Period Problem", Review of Economic Studies, Vol. LII (4), 171, 647-664.

GALLINI, N. and R. WINTER (1983) "Vertical price restraints in monopolistic competition", International Journal of Industrial Organisation, 1: 275-286.

GJESDAL, F. (1981) "Accounting for Stewardship", Journal of Accounting Research, 19 (1): 208-231.

GOLDBERG, V. (1982) "Resale price maintenance and the FTC: the Magnavox investigation", William & Mary Law Review, 23: 440-500.

GOULD, J.R. and L.E. PRESTON (1965) "Resale price maintenance and retail outlets", Economica, XXXII: 302-312.

GROSSMAN, S. and O. HART (1980) "Takeover Bids, the Free-Rider Problem, and the Theory of the Corporation", Bell Journal of Economics, Vol. 11: 42-64.

GROSSMAN, S. and O. HART (1982) "Corporate Financial Structure and Managerial Incentives", in J. McCall, ed. "The Economics of Information and Uncertainty", Chicago: University of Chicago Press.

GROSSMAN, S. and O. HART (1983) "An Analysis of the Principal-Agent Problem", Econometrica, 7-46.

GROSSMAN, S. and O. HART (1985) "The costs and benefits of ownership: a theory of vertical and lateral integration", mimeo, MIT.

GUASCH, J.L. and A. WEISS (1980) "Wages as Sorting Mechanisms in Competitive Markets with Asymmetric Information: A Theory of Testing", Review of Economic Studies, Vol. 47: 653-664.

GUASCH, J.L. and A. WEISS (1981) "Self-Selection in the Labour Market", American Economic Review, Vol. 71: 275-284.

GUASCH, J.L. and A. WEISS (1982) "An Equilibrium Analysis of Wage-Productivity Gaps", Review of Economic Studies, Vol. 49: 485-497.

HARRIS, J.R. (1970) "On the Economics of Law and Order", Journal of Political Economy, 78: 165-174.

HARRIS, M. and B. HOLMSTROM (1983) "On the duration of Agreements", mimeo, IMSSS, Stanford University.

HARRIS, M. and A. RAVIV (1986) "Corporate Control Contests and Capital Structure", mimeo, J.L. KELLOGG, Graduate School of Management, Northwestern University.

HART. O.D. (1983) "The Market Mechanism as an Incentive Scheme", Bell Journal of Economics, Vol. 14: 366-382.

HART, O.D. and B. HOLMSTROM (1986) "The Theory of Contracts", in T. BEWLEY, ed. "Advances in Economic Theory, 1985", Cambridge: Cambridge University Press.

HART, O.D. and J. MOORE (1985) "Incomplete Contracts and Renegotiation", mimeo, London School of Economics (STICERD).

HOLMSTROM, B. (1979) "Moral Hazard and Observability", Bell Journal of Economics, 10: 74-91.

HOLMSTROM, B. (1982a) "Moral Hazard in Teams", Bell Journal of Economics, 13: 324-340.

HOLMSTROM, B. (1982b) "Managerial Incentive Problems - A Dynamic Perspective", in Essays in Economics and Management in Honor of Lars Wahlbeck. Helsinki: Swedish School of Economics.

HOLMSTROM, B. and P. MILGROM (1985) "Aggregation and Linearity in the Provision of Intertemporal Incentives", Cowles Discussion Paper No. 742.

HOLMSTROM, B. and L. WEISS (1985) "Managerial Incentives,Investment and Aggregate Implications: Scale Effects", Review of Economic Studies, LIC (3), 170: 403-426.

HOTELLING, H. (1929) "Stability in competition", Economic Journal, 39: 41-57.

JENSEN, M. and W. MECKLING (1976) "Theory of the firm: Managerial Behaviour, Agency Costs, and Capital Structure", Journal of Financial Economics, 3: 305-360.

JONES, R.A. and J.A. OSTROY (1984) "Flexibility and Uncertainty", Review of Economic Studies, LI, 13-32.

KLEIN, B., R. CRAWFORD and A. ALCHIAN (1978) "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", Journal of Law and Economics, Vol. 21: 297-326.

D. KREPS (1984) "Signaling Games and Stable Equilibrium", mimeo, Stanford University.

LAFFERTY, R.N., R.H. LANDE and J.B. KIRKWOOD (eds.) (1984) "Impact evaluations of Federal Trade Commission vertical restraints cases",Bureau of Competition, Bureau of Economics, Washington: Federal Trade Commission.

LELAND, H. and D. PYLE (1977) "Informational Asymmetries, Financial Structure, and Financial Intermediation", Journal of Finance, 32, No. 2.

MACHLUP, F. (1967) "Theories of the Firm: Marginalist, Behavioural, Managerial", American Economic Review, Vol. 57: 1-33.

MANNE, H. (1965) "Mergers and the Market for Corporate Control", Journal of Political Economy, 73: 110-120.

MARCUS, A. (1982) "Risk-Sharing and the Theory of the Firm", Bell Journal of Economics, Vol. 13: 369-378.

MARRIS, R. (1964) "The Economic Theory of Managerial Capitalism", New York: Free Press.

MARVEL, H.P. and S. McCAFFERTY (1984) "Resale price maintenance and quality certification", Rand Journal of Economics, 15: 346-359.

MASKIN, E. and J. TIROLE (1985) "Principals with Private Information, II: Dependent Values", notes.

MATHEWSON, F. and R. WINTER (1984) "An economic theory of vertical restraints", Rand Journal of Economics, 15 (1): 27-38.

MATHEWSON, F. and R. WINTER (1986) "The Economics of vertical restraints in Distribution", in J. STIGLITZ and F. MATHEWSON (eds.): "New Developments in the Analysis of Market Structure", Cambridge: MIT Press.

McEACHERN, W. and A. ROMEO (1984) "Vertical restraints in the audio-components industry: an economic analysis of FTC intervention", in LAFFERTY et al. (1984).

McLAUGHLIN, A. (1979) "An economic analysis of resale price maintenance", unpublished dissertation, UCLA.

MILGROM, P. and J. ROBERTS (1982) "Limit Pricing and Entry Under Incomplete Information", Econometrica, 50: 280-312.

MIRRLEES, J. (1974) "Notes on Welfare Economics, Information and Uncertainty", in M. BALCH, D. McFADDEN and S. WU (eds.), "Essays in Economic Behaviour Under Uncertainty", Amsterdam: North-Holland, 243-258.

MIRRLEES, J. (1975) "The Theory of Moral Hazard and Unobservable Behaviour - Part I", mimeo, Nuffield College, Oxford.

MIRRLEES, J. (1976) "The Optimal Structure of Authority and Incentives within an Organisation", Bell Journal of Economics, 7: 105-131.

MIRRLEES, J. (1979) "The Implications of Moral Hazard for Optimal Insurance", Seminar given at Conference held in honour of Karl Borch, Bergen, Norway, mimeo.

MODIGLIANI, F. (1958) "New Developments on the Oligopoly Front", Journal of Political Economy, LXVI (3): 215-232.

MOOKHERJEE, D. (1982) "Optimal Incentive Schemes in Multiagent Situations", mimeo, London School of Economics (STICERD).

MUSSA, M. and S. ROSEN (1978) "Monopoly and Product quality", Journal of Economic Theory, 18: 301-317.

MYERSON, R. (1983) "Mechanism Design by an Informed Principal", Econometrica, 51 (6): 1769-97.

NALEBUFF, B. and D. SCHARFSTEIN (1985) "Self-selection and Testing", mimeo, forthcoming in The Review of Economic Studies.

NELSON, R. and S. WINTER (1982) "An Evolutionary Theory of Economic Change", Cambridge: Harvard University Press.

OVERSTREET, Th. R. Jr. (1983) "Resale price maintenance: economic theories and empirical relevance", Bureau of Economics Staff Report, Washington: Federal Trade Commission.

POLINSKI, A. and S. SHAVELL (1979) "The Optimal Trade-off between the Probability and Magnitude of Fines", American Economic Review, 69: 880-891.

POSNER, R. (1976) "Antitrust Law: An Economic Perspective", The University of Chicago Press, Chicago.

REY, P. and J. TIROLE (1985a) "Vertical Restraints from a Principal Agent Viewpoint", mimeo, MIT.

REY, P. and J. TIROLE (1985b) "The Logic of Vertical Restraints", mimeo, MIT.

ROGERSON, W. (1985) "Repeated Moral Hazard", Econometrica, 53: 69-76.

ROSS, S. (1977) "The determination of financial structure: The incentive signaling approach", Bell Journal of Economics, 8 (Spring): 23-40.

SALOP, S. (1986) "Practices that (Credibly) Facilitate Oligopoly Co-ordination", in J. STIGLITZ and F. MATHEWSON: New Developments in the Analysis of Market Structure. Cambridge: MIT Press.

SAPPINGTON, D. (1983) "Limited Liability Contracts between Principal and Agent", Journal of Economic Theory, vol. 29: 1-21.

SCHARFSTEIN, D. (1985a) "The Disciplinary Role of Takeovers", mimeo, MIT.

SCHARFSTEIN, D. (1985b) "Product Market Competition and Managerial Slack", mimeo, MIT.

SHAKED, A. and J. SUTTON (1982) "Relaxing price competition through product differentiation", Review of Economic Studies, 49: 3-14.

SHAKED, A. and J. SUTTON (1983) "Natural oligopolies", Econometrica, 51: 1469-1484.

SHAVELL, S. (1980) "Damage Measures for Breach of Contract", Bell Journal of Economics, Vol. 11 (Autumn): 466-483.

SHLEIFER, A. and R. VISHNY (1984) "Greenmail, White Knights, and Shareholders' interests", mimeo, MIT.

SIMON, H. (1959) "Theories of decision making in economics and behavioural science", American Economic Review, 49: 253-58.

SOLOW, R. (1967) "The New Industrial State or Son of Affluence", Public Interest , 9: 100-108.

SPENCE, A.M. (1977) "Entry, Capacity, Investment and Oligopolistic Pricing", Bell Journal of Economics, Vol. 8: 534-544.

SPENGLER, J. (1950) "Vertical Integration and anti-trust policy", Journal of Political Economy, LVIII (4): 347-352.

STIGLER, G. (1970) "The optimum Enforcement of Laws", Journal of Political Economy, 78: 526-536.

STIGLITZ, J. (1975) "Information and Economic Analysis", in M. PARKIN and A.R. NOBAY (eds.): Current Economic Problems, Cambridge: Cambridge University Press.

STIGLITZ, J. (1984) "Price Rigidities and Market Structure", American Economic Review, Vol. 74 (2): 350-355.

STIGLITZ, J. and A. WEISS (1981) "Credit-Rationing in Markets with Imperfect Information", American Economic Review, 71 (3): 393-410.

TELSER, L. (1960) "Why should manufacturers want fair trade?", Journal of Law and Economics, III, 86-105.

TIROLE, J. (1986) "Hierarchies and Bureaucracies", mimeo, MIT.

TOWNSEND, R. (1979) "Optimal Contracts and Competitive Markets with Costly State verification", Journal of Economic Theory, 21: 265-293.

WILLIAMSON, O. (1963) "Managerial Discretion and Business Behaviour", American Economic Review, LIII (5): 1032-1057.

WILLIAMSON, O. (1964) "The Economics of Discretionary Behaviour: Managerial Objectives in a Theory of the Firm", Prentice-Hall, Englewood Cliffs, W.S.

WILLIAMSON, O. (1967) "Hierarchical control and optimum firm size", Journal of Political Economy, 75: 123-38.

WILLIAMSON, O. (1971) "The vertical integration of production: Market failure considerations", American Economic Review, 61: 112-23.

WILLIAMSON, O. (1975) "Markets and Hierarchies: Analysis and Antitrust Implications", New York: The Free Press.

WILLIAMSON, O. (1979) "Assessing Vertical Market Restrictions: Antitrust ramifications of the transaction-cost-approach", University of Pennsylvania Law Review, Vol. 127: 953-93.

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