Contestable Markets and the Theory of the Multiproduct Firm

by

Dhanayshar Mahabir

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Arts.

Department of Economics
McGill University, Montreal
July 1985

Preface

Nearlly all of chapter five plus parts of chapter four can be considered to be an original contribution. In particular the case for endogenous potential competition (internal conditions), the introduction of latent excess capacity and the origins of scope economies are basically original ideas.

DEDICATION

I believe it is most appropriate that I dedicate this work to my parents. For being unable to read and write, they will never be able to study it or comprehend even the simplest expressions, yet it was they more than any other who offered total encouragement, support and assistance in every possible way. Their emphasis on learning only reflected how much they themselves felt powerless without it and thus how necessary it was to acquire the requisite training. For all those years of nurturing and moral support, I wish now to return the compliment.

To my parents...and to the peasantry from whence I came and to whom I will return.

ACKNOWLEDGEMENTS

I owe an immense debt to Professors Robert D Cairns and Christopher Green who encouraged me to be original, to develop new ideas and to explore the unknown.

I am also grateful to my collegues Jean-Francois Baladi and Hans Peter Von Sicard for providing the translation of the abstract.

FOREWORD

"It more frequently happens that a business, or even an industry finds its advantage in using a good deal of the same plant technical skill, and business organisation for several classes of products. In such cases the cost of anything used for several purposes has to be defrayed by its fruits in all of them. But there is seldom any rule of nature to determine either the relative importance of these uses, or the proportions in which the total cost should be distributed among them."

Alfred Marshall
Principles of Economics

"The theory of joint production is an important part of the study of economics and it is rather complicated."

Heinrich Von Stackelburg
Grundlagen der Theoretischen Volkswirtschaftslehre
The Theory of the Market Economy

ABSTRACT

This study examines the theory of the multiproduct firm that has emerged in the decade 1970-1980 with some reference to earlier contributions. The bulk of the analytical material presented draws heavily on the works of W.J. Baumol, J. Panzar and R.D. Willig, the pioneers in the area.

In addition, due consideration is paid to the theory of Natural Monopoly and Contestability since they developed in conjunction with and form an integral part of the theory under consideration. Upon exploring and assessing the theory, the writer has concluded that although the theory of the multiproduct firm is acceptable, that of contestability is not. The requirements of contestable markets are extremely stringent and it was argued that if some of these restrictions are relaxed by considering the existence of latent excess capacity and its relation to economies of scope, the same ideal results could be obtained in a more realistic environment.

It was also argued that the conclusions emerging from the new approach, though conflicting with orthodox opinion, arise primarily from the specific assumptions

employed. This applies to the unsustainability of the Natural Monopoly, the endogenous explanation of industry structure, intertemporal mis-allocations of the weak invisible hand and the absence of games in an oligoplistic structure.

Nevertheless this study credits "the uprising in industrial organization" for its explicit focus on the micro economics of the mega corporations, the powers of potential competition, costless exit and absence of sunk costs, and the influence of technology in modifying industry structure. It is suggested too that the array of concepts and tools introduced into the literature by the new theory will undoubtedly assist in the construction of improved models of the industrial economy.

PRECIS

Ce mémoire examine la théorie de la firme à multi-produits parue, pendant la dernière décennie (1970-1980), et souligne quelque peu les premières contributions dans ce domaine. La majeure partie de l'analyse presentée ici est inspirée du travail de Baumol, Panzar et Willig, les précuseurs dans ce domaine.

De plus, la théorie du monopole naturel et celle de la contestabilité sont particulièrement analysées puisque celles-ci se sont développées conjointment et forment une partie intégrale de la théorie considerée. Après avoir expliqué et evalué la théorie de la firme à multi-produits, l'auteur conclu que, bien que la théorie de la firme à multi-produits soit acceptable celle de la contestabilité ne l'est pas. Les restrictions placées sur les marchés contestables sont extrêmement astreignantes et l'auteur soutiens que si quelques-unes de ces restrictions sont relachées en prenant en consideration l'existence d'une capacité potentielle excessive -- et le rapport entre celle-ci et les 'économies de scope' -- les mêmes résultats idéaux peuvent être obtenus dans un environnement plus réaliste.

L'auteur trouve aussi que les conclusions tirées de

cette nouvelle théorie, quoique étant en conflit avec l'opinion orthodoxe, proviennent des hypothèses particulières employées. Ceci s'applique à la difficulté de défendre le monopole naturel, l'explication endogène de la structure industrielle, la mal-allocation intertemporelle de la faible main invisible, et l'absense des jeux dans une structure oligopolistique.

Néanmoins, ce mémoire trouve approprié 'le renouveau dans la théorie de l'organisation industrielle' pour avoir mis au point la micro-économie de la mega-corporation, les pouvoirs de la competition potentielle, les bénéfices de la sortie sans frais, et l'absense des coûts coulés et l'influence de la technologie sur la structure industrielle. En outre, le nombre immense de nouveaux outils qui ont été introduit dans la littérature contribueront, sans doute, à la construction de modèles ameliorés de l'économie industrielle.

TABLE OF CONTENTS

INTRODUCTION

The more industrialized economies are populated by firms, which produce a variety of outputs. Yet received theory has continued to focus on the small firm producing a single output, implicitly assuming that such analysis can easily be transferred to multi-output production.

Dissatisfaction with orthodox theory in this respect is not new. The 'managerial school' has tried to introduce new concepts and a different framework of analysis for the study of the large corporation characterised by a divorce of ownership from management. But still they did not deal with the micro economics of multi-output production as a separate and potentially powerful area of research.

Precursors in the field of micro theory did at times casually consider the complexities involved in multiproduct firms. Marshall[1], Dewsnup[2] and Carlson[3] all considered, to some extent, the mechanics of such activity. However, in an early paper by Weldon[4], the economics of the multiproduct firm was examined and a few of the novel elements of recent theory were in fact established in this article. Recognition of economies of

scope, the public good nature of fixed inputs and the impossibility of measuring average cost in this type of firm were clearly outlined.

These valuable ideas and insights, though seminal in their time, never received the momentum that they should have (with the exception of articles by Ralph W. Pfouts)[5]. But the theme was revived in the mid-1970's when a team of researchers headed by Baumol began to produce a series of articles on this topic. The re-emergence of multiproduct analysis arose no doubt as a result of the debate to open the american telecommunications market to competition. The researchers in question associated with Bell Laboratories and Princeton and New York Universities first investigated the concepts of the natural monopoly, the multiproduct natural monopoly, and then the multiproduct firm in general.

After the 1970 paper by Baumol and Bradford[6] on optimal departures from marginal cost pricing (restating the Ramsey Rule of different pricing according to varying demand elasticities), research focused on the concept of natural monopoly. This topic, worthy of an investigation by itself (see Sharkey)[7] had proved to be a bothersome issue to economists. Disagreements lay on what its

definition was, what the underlying characteristics were and why there existed the need for departures from marginal cost pricing.

Hotelling[8], building on the insights of the French engineer J. Dupuit[9], had proposed in a classic paper that the services of industries characterised by increasing returns to scale should be priced at their marginal cost (in order to maximise consumers surplus) with the deficit being made up by the state through general taxation.

Later Demsetz[10] proposed the alternative solution of opening natural monoplies to competitive bidding, thereby preventing any monopoly pricing by the unregulated monopoly, preempting the need for regulation or government interference and ensuring least cost production.

This still did not adequately characterise the natural monopoly and grave ambiguities lurked in the background. Furthermore, following Demsetz' solution, a major problem still remained. Having granted a natural monopoly, the contract for which it bid, then during the tenure of such a contract the firm was legally protected. Such protection attenuated incentives for innovative activities since it effectively prevented entrants with

superior products from participating in the market during the time of the contract.

Thus the topic was later researched by Faulhaber[11] and for the first time one could observe the makings of a complete theory of natural monopoly. A natural monopoly existed if it could produce a product or set of products at a cost less than that of a series of other firms producing one product each or producing in many different combinations. Thus the term subadditivity appeared; a cost function exhibiting such characteristics lent itself to the emergence of monopoly. The implication of this analysis was that a single firm could be the cheapest supplier of a vector of outputs (considering explicitly now the multiproduct firm) but might find no vector of prices at which to at least break even and yet be sustainable against entry. That the natural monopolist cannot always sustain himself from wasteful entry was the subject of the paper in 1977 by J.C. Panzar and R. Willig.[12] This idea, we shall argue, is the single most important contribution of the new theory of industrial organisation.

That a cost function could be subadditive, and that a monopoly could be the most efficient supplier of its output vector and yet not prevent wasteful and profitable

entry is an extremely powerful proposition. But this, we shall later argue, results not so much from a failure of orthodox theory, but in the implicit assumption of the new approach. Whereas traditionally a natural monopoly was characterised by both declining average and marginal cost -- now a cost function could be subadditive but marginal and average costs could be rising.[13] When the monopoly is forced to produce on the rising portion of its subadditive average cost curve -- then it becomes vulnerable to wasteful entry. Though the cost function is subadditive, it does not mean that the monopolist has an absolute advantage in each and every subsector of his output vector. Being forced to supply total market demand, he then becomes vulnerable by leaving open a superior ray vector that an entrant could immediately adopt.

The possibility of the the unsustainability of a natural monopoly led to research into the nature of the multiproduct firm. Why was the firm involved in producing a vector of outputs rather than one product? Clearly there existed advantages to multi-output production. These advantages were discussed under the terms economies of scope, joint production, cost complementarities and the existence of trans-ray convexity.

All these new concepts reflect the cost savings due to multi-output production and enable subadditivity in one's cost function. As such the reader will immediately see why the existence of an unsustainable natural monopoly is so crucial to the overall social welfare. An entrant invading the unsustainable (but efficient) natural monopolist will product a subset (ye) (where ye refers to the output of the entrant) and price at a lower average cost than the incumbent. However, such entry is wasteful if the remaining output of the monopolist increases in cost as a result of the loss of scope economies, economies which were reaped when the entire vector was producted but are now diminished by entry.

Once the theoretically possible case of an unsustainable natural monopoly and the need for protection from entry had been unequivocally established, there arose the problem of how to simultaneously prevent entry yet force the monopolist to produce efficiently and innovate rapidly.

Further research by Baumol et al[14] provided some so-called weak invisible hand theorems giving conditions under which the monopolist may sustain himself and yet serve the public good. The adoption of Ramsey Optimal Prices were (under a set of stringent conditions)

guaranteed to be sustainable and yet limit monopoly exploitation to the level of natural entry barriers. However, there still existed situations where a monopolist could find no subsidy free prices that will prevent entry.

The weak invisible hand theorems and the requirements of sustainability are (as we shall see later) extremely rigid conditions with applications only in theoretical models. In reality if a natural monopolist became vulnerable to entry then he might be invaded for he may lack the precise information required for calculating sustainable prices.

Nonetheless the introduction of new concepts for analysing joint production -- least ray average cost, trans ray convexities, cost complementarities, scope economies, product specific economies of scale (and such bizarre constructs as the "floating hyperbagel and transylvanian cost functions") paved the way for finally incorporating the large firm firmly into the general body of micro-theory.

A definite attempt was now being made to study multiproduct behaviour qua multiproduct behaviour, and many new and rigorous ideas emerged. Here scope

economies play a major role in understanding these characteristics indicating specifically that multiproduct firms evolve precisely because of the characteristics of their cost functions. Thus to produce output x requires some fixed (or sunk) input Y. But Y being available to product x, now costlessly becomes available for the production of z. The total cost of one firm producing both x and z is thus lower than the total cost of two separate firms producing these goods individually. However, no attempt was made to differentiate and identify the sources of economies of scope as opposed to scale and, as this writer will argue later, it is scale not scope that invariably is crucial. We see that scale is the precursor of scope and that the latter is frequently an advantage of operating at a certain scale of output.

But the foundations had firmly been laid for a rigorous study of multiproduct activity, not only in natural monopoly but in oligopolies, duopolies and all other market structures that exist.

Now one encounters the concept of contestability.[18] A contestable market is one in which entry is free and easy -- entrants possess no legal disadvantages or setbacks of any sort. They also possess identical

technology or access to such technology, and can enter whenever a profitable opportunity presents itself. When profits are dissipated or when incumbents re-adjust their pre-entry prices, such entrants can escape with the minimum amount of loss since costless exit is also an essential condition in such a market.

The importance of contestable markets is identified in a multiproduct setting as the key towards the generation of an endogenous industry structure that is efficient (cost minimizing) and sustainable.

A multiproduct firm, then, in a contestable market must price and produce efficiently taking advantage of all cost savings accruing to it via multiproduction. It must also produce each good in just the right quantity so that when combined in the vector of outputs the firm produces at least ray average cost. An exogenously determined demand for product x, y, z will then determine how many firms (with equal access to technology) must engage in such production, which will not only be the most efficient configuration but satisfy total demand as well. A contestable market also ensures that no firm produces at a ray average cost (to be defined) that is not minimal. Otherwise the firm would be vulnerable to a profit seeking entrant waiting in the wings ready to

supply at the least cost price.

Nonetheless contestable markets like competitive markets (the former is but a generalisation of the latter) suffer from the consequence of non-existence of an equilibrium. In contestability the equilibrium is the sustainable price-output vector but there is no guarantee of the existence of such equilibria (unsustainable natural monopoly being one example). Furthermore perfect contestability does not take into consideration such traditional market failures as externalities, and public goods (non exclusion principle) or social welfare. But it does attack the problem of natural monopoly squarely and, as suggested before, this may be its single great contribution.

A market deemed to be contestable and characterized by the absence of sunk costs guarantees an outcome superior to that of protection. It is the threat of competition, existence of potential entry and entrants hovering waiting for a profitable opportunity that is necessary to discipline the monopolist. Such threat of entry is the potent force that will induce firms to eliminate x-inefficiency, satisficing behaviour and organisational slack. Cross subsidy is impossible since an entrant could produce an overpriced product line,

price at a lower level and eliminate the incumbent's excess profits. Marginal cost pricing where the number of firms exceeds one is guaranteed, again dictated to by the free entry and costless exit characterisation. Only in natural monopolies is a Ramsey vector of prices sustainable resulting from special characteristics of the cost function, the nature of demand and the existence of non-dominated prices.

Though contestable markets are unlikely to exist in reality, it has been suggested by the theorists that it could be used as a valuable benchmark of comparison of how actual performance compares with ideal performance. It is now postulated that in the regime of contestability and potential competition, there is no need for a large number of firms to yield optimal results, only one single firm is necessary in the case of natural monopoly, and even an oligopoly produces results consistent with perfect competition. Elimination of large numbers mean that the benchmark for comparison is more closely aligned to the realities of the more developed economies.

To obtain the social benefits of multioutput production a contestable market is required -- one characterised by the absence of sunk costs. Such a market eliminates the need for imposing price taking

behaviour (firms are constrained here by potential entry) and guarantees the efficient production of outputs in the optimal number of firms. At a few strokes the new theory of industrial organisation is established, and the multioutput unit is effectively incorporated into the body of knowledge.

This thesis is an attempt to succinctly outline this new theory of industrial organisation as put forward by Baumol et al[16] and which developed over the decade 1970-1980. But this is a critical summary -- one designed to examine the salient elements of the theory and to modify and explore key aspects of this theory that will not only enhance its potency but render it more applicable to contemporary industrial economies.

Some of the modifications attempted here are, first, an integration of consumers demand in a contestable market with the preliminary result being that even in the absence of sunk costs and complete freedom of entry and exit an incumbent firm will be able to protect itself by 'capturing consumers'. There is thus the possible situation that an entrepreneur can price his products above marginal costs but because of brand preferences and product loyalty on the part of captured consumers due to advertising and reputation, etc., the incumbent is

partially protected from entry. Furthermore, it is argued that the results of contestability will only be achieved if entrepreneurs all produce identical nondifferentiated products (e.g. telecommunications) but even here consumers must be convinced of the credibility of an entrant in continuing his production rather than just offering a lower price in the initial period. Under changing conditions of demand and alterations in consumers tastes the theory loses some of its robustness.

In what follows the assumptions of the theory of contestability are left basically intact but the results of the theory are criticised. One of the underlying themes of this paper is that even a perfectly contestable market is not guaranteed to produce optimal results. Beside examining this possibility by introducing consumer demand, an introduction of time into the analysis renders it unstable. It is easily seen that if production takes time to plan and complete and if such time is crucial to achieve minimum efficient scale, an entrant will be reluctant to invade a market when his production lag is long enough for an incumbent to lower his price and render production by the entrant unprofitable.

Another important point of disagreement is in the treatment of incumbent vs. entrant advantages. The

theory treats the entrant as a guerilla lurking in the background and the incumbent(s) is seen as unwieldy and bureaucratic. However, this writer has tried to identify advantages that belong to the incumbent as such and to determine the extent to which this established position may protect entry. The important notion of ex-ante and ex-post manifests itself. Under the new theory an incumbent who prices at a surplus profit level will become vulnerable to entry -- but the analysis suffers from a major flaw. If the incumbent produces 1000x per week at prices p^* (Y_I = output of incumbent) so inviting entry and the entrant produces 1000x per week (point of least average cost) and prices at $p' < p^*$ then the problem is immediately seen. The market must now clear 2000x at p' and depending upon the elasticity of demand both parties are prone to destruction. An incumbent thus could use his incumbent status to threaten that <u>his</u> level of output will remain constant ex-post (the Bain-Sylos postulate) even if the entrant thinks (ex-ante) of entering. There are clear advantages then of being the established firm and this is a most credible threat to entry.

Another opinion (rather than a critique as such) stressed continuously in this paper is that some of the conditions of contestability are excessively harsh and

the identical results could be obtained by removing some of the restrictions while at the same time bringing the theory closer to reality. The need for a large number of hovering entrants with nothing to do but wait for potential possibilities can easily be dispensed with. Competition among the few on the other hand is more realistic and just as potent. Thus if firms possess latent excess capacity and quasi-fungible plant, this will act as a more credible threat to suboptimal behaviour than a pool of entrants in the wings. This kind of competition where the actors are already in possession of their labour skills, reputation, sunk facilities and latent excess capacity is now a more appropriate threat to adverse behaviour than a potential entrant who is unknown on the one hand and who must of necessity possess some idle capacity (e.g. labour) anticipating any entry profitabilities. Furthermore, the reality of industrial economies is not one of myriads of potential entrants on the wings but that of corporate entry, mergers and hostile takeovers by already established firms.

In linking the multiproduct firm to contestability, an investigation of economies of scope and economies of scale was undertaken. Having identified the source of scope economies as basically indivisibilities in plant

and the existence of costless latent excess capacity, it was then argued that economies of scope and scale were inseparable. While the scope economies supplied the dynamics of multi-output production they first of all emerged due to size, i.e. operating with indivisibilities in the production process and/or in utilizing plants of large sizes in order to achieve minimum average cost. Thus it is scale not scope that arguably proves to be the ultimate barrier to entry. Building upon this investigation, the notion of the unsustainability of a natural monopoly in an environment where consumers are aware of the scope economies of such a structure, eliminates the need for intervention. The endless series of games that invariably follow unsustainability is diminished when every player becomes aware of his losses with a partial co-alition. This argument is presented more fully in chapter three.

These are the major observations of this writer concerning the theory of contestability and the multiproduct firm as it now stands. The limitations are recognised but the strengths are also stressed. Potential entry and ease of exit have finally been recognised as constraints to behaviour (not only business but social, political and organisational as well, in political and social life the threat of violence with

immunity may be enough to discipline dissidents); the large numbers requirement of perfect competition is no longer necessary to obtain efficiency and thus the theory comes close (in this sense) to the realities of the day. The multiproduct firm is analysed fully by this theory and the set of new tools that are introduced will prove to be of immense use in further research; natural monopoly is not seen as a market failure but rather the result of a perfectly functioning market, and the ambiguities surrounding increasing returns industries are dispersed with by identifying the characteristics and sources of such increasing returns and of the constraints to behaviour that potential entry poses. Natural monopolies and industries experiencing increasing returns (though not identical phenomena) emerge out of competition and are able to survive precisely because they are the most efficient structure for input use.

The internal explanation of industrial structure emphasizing technology and the cost function is another novel modification and though this writer believes that such internal determination is a long period result, the idea of technology leading structure is clearly a valuable insight in the study of the evolution of industry. Certainly it is the study of the cost function (reflecting technology) that has facilitated research

into the economies of multiproduct operations.

In the final analysis then the "new theory of industrial organisation" will initiate a major debate on the nature of industry in the advanced economy. Once the theories have been studied by the general profession, new and superior critiques will appear that could only assist in refining and reformulating these new ideas which are still in their evolutionary stages. Although this writer is critical on many occasions -- this is not meant to indicate total disagreement with the ideas. On the contrary, one cannot help but be stimulated by the sheer power of the innovations and to respect the pathbreaking and novel results of the research. Because the analysis is so crucial -- criticism is necessary, if only to eliminate some of the more obvious flaws that inevitably persist in any masterpiece.

NOTES & REFERENCES TO INTRODUCTION

1. Marshall, A. (1925). Principles of Economics, London and New York, Macmillan, 1925.

2. Devsnup, E.R. (1914). "Railway Rate Making," American Economic Review, 4 (Supp.), 1914, pp. 81-100.

3. Carlson, S. (1937). "A Study on the Pure Theory of Production", London 1937

4. Weldon, J. (1948). "The Multiproduct Firm," Canadian Journal of Economics and Political Science, vol. 14, May 1948, pp. 176-190.

5. Pfouts, R. (1961). "Theory of Cost and Production in the Multiproduct Firm," Econometrica 24 (4), October 1961, 650-659.

5b. Pfouts, R. (1964). "Multiproduct Firms vs. Single Product Firms: The Theory of Costs and Production," Metro-economica, Vol.16, pp. 51-66.

6. Baumol, W.J. and Bradford, D.E. (1970). "Optimal Departures from Marginal Cost Pricing," American Economic Review, 60, June 1970, pp. 265-283.

7. Sharkey, W.W. (1982). The Theory of Natural Monopoly, Cambridge, U.K.: Cambridge University Press, 1982.

8. Hotelling, H. (1938). "The General Welfare in Relation to Problems of Taxation and Railway and Utility Rates," Econometrica 6, 1938, pp. 242-269.

9. Dupuit, J. (1844). "On the Measurement of the Utility of Public Works", in Arrow and Scitovsky (eds.), Readings in Welfare Economics, "American Economic Association", vol. XIII, R.D. Irwin Inc., Illinois, 1969

10. Demsetz, H. (1968). "Why Regulate Utilities?" Journal of Law and Economics, 11, April, 1968, pp. 55-65.

11. Faulhaber, G. (1975). "Cross Subsidization: Pricing in Public Enterprise," American Economic Review, 65, December 1975, pp. 966-977.

12. Panzar, J.C. and Willig, R. (1977). "Free Entry and the Sustainability of Natural Monopoly," Bell Journal of

Economics, 8, Spring, 1977, pp. 1-22.

13. Sharkey, W.W., op. cit.

14. Baumol, W.J. et al. (1977). "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct Monopoly," American Economic Review, 67, June 1977, pp. 350-365.

15. Baumol, W.J. et al. (1982). Contestable Markets and the Theory of Industry Structure, Harcourt, Brace, Jovanovich, New York, 1982.

16. Ibid.

CHAPTER 1

COST CONCEPTS

We mentioned in the introduction that the cost function plays the pivotal role in the analysis of the multiproduct firm. A knowledge of its characteristics is essential to fully analyse the complexities that multi-output production introduces. The world of reality is the world of the multiproduct firm and thus there must be a thorough understanding of the cost of producing outputs in such a firm.

Why are such costs so important in the analysis? The answer is clearly given by the fact that the characteristics of the cost function provide information on economies of scale and of joint production. While multiproduct scale economies are difficult to measure -- new constructs that explain the nature of multiproduct activity must be invented. These will play perhaps the most crucial role in the entire exercise.

"The cost function relates to the firms input decisions given its output levels and input prices."[1] The theory of duality indicates that under certain restrictions, one can derive a cost function from a given

production function, and by the applications of Sheppard's Lemma, the input demand function will be obtained.[2]

It will be useful then to review briefly the conditions under which one can secure information about costs given only the production function.

1.2 Duality - A Review

Assumptions: In order for the duality results to hold certain restrictions on the cost function must be imposed.

1. It must be a linearly homogenous function in prices for producible output bundles and strictly positive input prices.
2. It must be strictly monotonically increasing in outputs.
3. It must exhibit concavity in input prices, i.e. $c(tw + 1 - t)\, w', y) \geq tc\,(w,y) + (1-t)\, c(w'y)$, with t representing an arbitrary weight
4. It must be continuous with respect to outputs.

The assumption of concavity implies of course that the firm substitutes one factor for another depending on changes in factor prices.

Given these assumptions and the implicit notion of cost minimising behaviour, the optimal input demand schedules are obtained by Sheppards lemma,

i.e. $x(w,y) = \frac{\partial c(w,y)}{\partial w_i}$, $i = 1...n$

Here c(w,y) represents the cost function with w being the vector of input prices and y the vector of outputs. "The derivative property of the cost function tells us that optimal combinations of input and outputs can be found by looking at the derivatives of the cost function with respect to the factor prices.....(i.e.) we can recover information about a technology by investigating its cost function.[4]

We see thus that the cost function of the firm under some restrictions summarises all the relevant aspects of its technology.

It is important to note, however, that there is an implicit neglect of Averch Johnson biases and non price taking behaviour. Regarding the latter, any firm that finds itself in the position of influencing its input prices will exhibit a cost function and a corresponding technology that is far different than that which will be derived from the duality theorems.

Nevertheless these new developments have paved a way

for the study of multi-output production. Innovations such as the translog cost function[5] have been used by researchers to study the problems and idiosyncracies of multi-output production. It is recognised, however, that the use of such techniques is still in its infancy and as yet the results have been mainly inconclusive.[6] The use of the translog or 'generalised Diewert' itself poses problems in estimation since the number of parameters to be estimated frequently exceeds the available data points thus leaving no degrees of freedom. Furthermore, certain economists have been completely dissatisfied with the approach of gathering information on technology from cost functions.[7]

The end result thus is that the techniques are still in their embryonic stages but this does not detract from the importance of the use of the cost function in understanding peculiarities of multiproduct outputs.

The following section will explain some cost concepts applicable both to the multiproduct case and the single output case.

1.3 Applicable Cost Concepts

The familiar single product case establishes the relationship between changes in cost and changes in output. This is the situation most frequently analysed.[8] Here one encounters the total cost concept: average cost, marginal costs, fixed cost and variable costs. A 'U' shaped average cost curve representing increasing, constant and diminishing returns is usually drawn with a marginal cost curve intersecting average cost at the point of minimum average cost. Other cases where average cost everywhere declines as in the case of 'natural monopoly" are also sometimes encountered.

We shall not elaborate on these well-established concepts here but rather some new terms will be introduced. These are the ray average cost, trans ray convexity, subadditivity and Incremental cost. While the single product concepts enable one to derive unambiguous measures of economies of scale, in the multiproduct case such a derivation is not as clear-cut. Instead one encounters diminishing ray Average Cost, cost complementarities, economies of scope and trans ray convexity. It will immediately become apparent why these new concepts are needed.

A firm engaging in joint production makes decisions which do not have simple analogies in the single product environment'. The production of one commodity means that it becomes cheaper to produce another, using the same common facilities. Whether it is feasible to add a production line depends on the nature of the good (substitute or complement), the incremental cost of this good, its price on the market and the additional revenues and cost that it entails. There clearly is a limit to adding the marginal line of output and it is only when the firm loses by adding another product line that it will contract its operations.

The cost concepts which are employed throughout the analysis will now be outlined.

1.3.1 Ray Average Costs

Since average costs play the primary role in our analysis, it is important to explain this concept clearly and to observe at once its various limitations. Calculating the average cost of n products in a multiproduct firm (MPF) is nearly impossible where these products all use the same common productive facilities.

To overcome the aggregation problem of adding apples and oranges, use is made of a ray drawn from the origin into output space.

Along the Ray, output bundles are constant while the scale of production is varied. As such, any output bundle (termed the composite commodity y') 10x + 4y increased by 50% to 15x + 6y shows how cost behaves along this ray. As this scale is continuously altered, the average cost curve of the composite commodity y' is traced and the point of minimum average cost is observed. While this is a neat construct, there are difficulties associated with it. The most obvious is that if the output mix is held constant, no cost advantage can be gained by changing its composition. It can be argued that changing the composition of outputs may result in cost savings through the reaping of greater scope economies.

While average cost calculation along a ray is necessarily arbitrary -- in a contestable market (to be defined) the weak invisible hand inevitably leads to the production of the most efficient vector. A contestable market thus ensures that not only will production be undertaken at the point of minimum ray average cost but that the most efficient bundle will be produced as well.

Marginal cost, it must be noted, can be calculated in the multiproduct firms by

$$\frac{\partial c(y_1, y_2)}{\partial y_1}$$

1.3.2 Average Incremental Cost

This is the change in total cost that results from dropping output y_1 from the production set divided by the y_1 formerly produced

$$AIC(Y_1) = \frac{c(y_1, y_2) - c(0, y_2)}{y_1}$$

This concept does, however, neglect a significant fact. In the presence of cost complementarities and economies of scope, the elimination of y_1 may significantly cause the cost of y_2 to rise. The loss of Y_1 thus ensures that all the benefits of joint production are not reaped.

Nonetheless this concept is used to indicate that "product-specific returns to scale to output 1 [in a multiproduct firm] are measured by

$$S_1 = \frac{AIC_1(Y_1)}{MC_1}$$

Thus, average incremental and marginal costs are interrelated in the same way as are single product average costs and marginal costs. Hence, $Si \gtrless 1$ as there

are increasing, constant, or decreasing returns to scale with respect to output type i. Consequently, if Si > 1, total incremental cost ..... will rise less than proportionately as the production of Y increases, with the quantities of all other output types held constant."[10]

Bailey and Friedlaender used the railroads to explain this particular concept. They argued that "it should be clear that the average cost of freight service or passenger service alone cannot be unambiguously defined, since the trackage is shared by both services, and there is no one correct way to allocate its costs between the two services. The incremental cost of either service, however, can be defined readily. This equals the total cost of both services less the total cost of providing one service alone."[11]

In this example the incremental cost of one service is just the variable cost of that service. It is possible therefore that marginal costs are constant implying no product-specific returns to scale while there do exist at the same time returns to joint production. To study this phenomenon one requires another concept -- trans ray convexity.

1.3.3 Trans Ray Convexity

If a cost function is of the form

$$C[ty' + (1-t)y''] \leq tc(y') + (1-t)c(y'')$$

where t represents some weights, then this cost function is trans ray convex. It thus requires that the production cost of a weighted average of a pair of output bundles y' and y" produced together be no greater than the weighted average of the cost of producing each of them in isolation. One can immediately infer from this characteristic that any such function will contribute savings due to common production, i.e. cost complementarities and economies of scope. If such scope economies (to be defined) exceed product specific scale economies, the function is trans ray convex and a firm will experience cost savings by altering the composition of its output vector keeping its scale of operation unchanged. This concept and its closely allied construct -- trans ray supportability[11] will come to play an important role in our study of the sustainability of a natural monopoly.

These constructs are enough to begin our study and as others are needed they will be introduced. Some more important developments such as economies of scope and

sustainability -- because they are of such crucial importance -- will be developed fully in subsequent chapters.

NOTES & REFERENCES TO CHAPTER1

1. W.J. Baumol, (1977) Economic Theory and Operations Analysis, Prentice Hall International, Inc., London, 4th ed.

2. Varian, H.R. (1978) Microeconomic Analysis, W.W. Norton and Company Inc., New York, pp. 32-35.

3. Ibid., p. 32.

4. Ibid., p. 35.

5. Melvin Fuss and L. Waverman, The regulation of telecommunications in Canada (1981). In "Studies in Public Regulation", edited by Gary Fromm (London: MIT Press, 1981).

6. Ibid.

7. Ralph Turvey. (1982) Book Review, "Studies in Public Regulation," Economic Journal, Vol. 92, pp. 731-732, September.

8. Even as advanced a text as Varian (op. cit.) deals only with the single product firm.

9. Ralph W. Pfouts. (1961) The Theory of Cost and Production in the Multi-Product Firm, Econometrica, Vol. 29, No. 4, October, pp. 650-658.

10. Elizabeth Bailey and Ann Friedlander. (1982) Market Structure and Multiproduct Industries, Journal of Economic Literature, Vol. XX, September, pp. 1030

11. Ibid., pp 1031

12. W.W. Sharkey and L.G. Telser. (1978) "Supportable Cost Functions for the Multiproduct Firm," Journal of Economic Theory, 18, No. 1, June, pp. 23-37.

CHAPTER 2

THE THEORY OF THE MULTIPRODUCT FIRM - EARLY WORKS

It was suggested before that nearly all firms engage in some multi-output activity yet theory still analyses firms' behaviour with one product in mind giving the impression that such analysis can easily be extended to the multi-product case.

In order to understand the mechanics of this process (multiproduct activity) and to see how it integrates into the new theory of natural monopoly, and the entire contestability argument it is import to obtain a clear picture not only of the causes of the multiproduct firm but on any idiosyncracies that such an analysis entails.

This concept of the multi-output character of firms is not new. Marshall[1] explicitly mentioned the peculiarities on both demand and supply of jointness in production and his citation of Dewsnup[2] indicated that older economists have been intrigued by the mechanics of the multi-output firms.

Not much seemed to have been written in the earlier period though, perhaps due to the absence of the large

corporation and the predominance of the family-owned companies. A later work by Carlson[3] did touch tangentially on the topic but he did not deal with the mechanics of the multiproduct firm as such.

It was only in 1948 that Weldon[4] explicitly treated the mechanics of the multiproduct firm and this seminal contribution, it can be argued, contained many of the salient points now put forward by newer researchers. Apart from the cursory treatment of Hicks, the theme was considered to some extent by Robertson and by Pfouts. But, apart from this, the early ideas of Weldon failed to gather momentum and the treatment of multiproduct firms had to wait until the mid-seventies.

We move now to study the mechanics of the multiproduct firm and to identify those characteristics that make it special. Recent theory[5] has taken the cost function as a starting point.

Given information on costs, assuming cost minimising behaviour (and the usual regularity conditions), one can by utilisation of Sheppard's Lemma[6] recover information on technology. Assuming no Averch Johnson biases or suboptimal internal behaviour (X-inefficiency, organisational slack, non-price motivations and

satisficing behaviour), the vector of efficient input quantities at fixed or varying input prices can be obtained. This would represent the technology of the firm where this "technology" consists of all the feasible combinations of inputs that are consistent with profit maximising behaviour.

Having thus obtained information on costs and assuming convexity, homogeneity, etc., the property of this cost function can be obtained. The new approach then concludes that if the cost function exhibits economies of scope, i.e.

$$\sum_{i=1}^{n} C(Y_i) < C[\sum_{i=1}^{n} Y_i]$$

in that it is cheaper for one firm to produce a vector of output at a cost less than a series of individual firms then this facilitates and encourages multi-output production.

A related concept is that of cost complementarity, in which the production of one good explicitly reduces the cost of another good so that the second's production cost is much lower than if it were produced in a separate firm.

These two characteristics are encompassed in the concept of a trans ray convex cost function. Such a

technology entails that a weighted average of two goods X and Y, when produced in the same firm, costs less than the same weighted average produced in separate firms.

All this amounts to is that there are cost complementarities and that the incremental cost of producing an additional line Y_1 in combination with Y_2 is less than if it were being produced alone. It is noteworthy at this point to observe that the newer literature has not investigated the sources of these economies of scope or the reasons for a sub-additive cost function. Such an investigation (which is attempted in this paper) is necessary to examine the extent to which it is scope as opposed to scale that in reality poses the greater barrier to entry.

Difficulties in the Multiproduct Firm

Quite early, writers recognised the difficulty of analysing costs in the firm where many outputs utilised common production facilities. Weldon for example clearly recognised the impossibility of accurately measuring average costs in such an environment, a task which later writers have also failed to accomplish. To overcome this

problem we focus our attention on the modified concept of average cost as constructed by pioneers in the field. Since there is no way of calculating the average costs of n commodities produced by a single set of common inputs (the classic aggregation problem), a weaker and, as we shall see, somewhat arbitrary construct has been developed.

This is the Ray Average Cost concept[7] that has been proposed by Baumol et al. Here an arbitrary combination of outputs is taken as the starting point, e.g. six shirts, twelve trousers and they are formed into a composite commodity Y°. The fixed combination is then scaled proportionally by a common factor t and being scaled on this ray (a ray being a line from the origin extended into output space reflecting fixed proportion but varying scales) one could subsequently observe the change in costs that result. The change in total cost divided by the common scale factor t will give the measure of ray average cost -- a measure it is hoped will capture the economies of scale (or ray economies of scale) that arise from multiple production.

The criticism that one will introduce on this concept is not so much its arbitrariness but on the determination of the output weights. Why is an initial

weight of 6x and 12y chosen and not 2x and 3y or 2y and 2x, etc? This becomes a problem because not only may average cost decline when output is increased (as a result of cost complementarities and economies of scope) but it may also decline when output bundles alter. Thus economies of joint production may be obtained when outputs vary not on any pre-determined path but on a haphazard probably non linear route. Furthermore, to observe the presence of cost savings due to changes in size, it is not required that all outputs be scaled in fixed proportions. Perhaps by altering output composition rather than scale the average cost may decline since the firm moves to a more efficient ray.

The general argument, however, is that the multiproduct firm arose because of the inherent advantages in terms of cost that accrue to such a firm. We now focus attention on the causes and origins of economies of scope.

The first formal utilisation of this concept appears to have been Weldon in 1948. The term common production costs was used in its place and for him these common production costs arose due to divisibility savings in management, labour and transportation. Unlike the new school, his analysis focussed more on the potentials of

the market than on savings due to costs as such. Thus it was seen that the "chief raison d'etre of the multiproduct firm is the supplying of products to a market close at hand."[8] While his analysis did not recognise the importance (or possibility) of cost reduction as a result of changing composition of outputs he did delve quite deeply into the sources of economies of scope.[9]

Such economies of scope arise as a result of efficient utilisation of fixed inputs which are needed to produce output x but that once in place are available like a public good for the production of output y. This public good nature of fixed inputs and the indivisibilities of factors meant that in each firm there lurked some "latent excess capacity" which with only a moderate amount of imagination (his mercenary entrepreneur) would be available at little or no extra cost for the production of another output line. An aggressive entrant not already established in a different line of activity and so possessing no latent excess capacity of his own will be outpriced in the market since his average cost of Y must now include that set of fixed inputs required to produce Y alone. Being already an incumbent firm possessing latent excess capacity may thus prove to be a considerable barrier to entry.

But this is not all; the paper explicitly recognises the various advantages of a multiproduct entrepreneur. "A single product entrepreneur cannot continuously price below average cost for any extended period." On the other hand a multi-output firm will find it in its interest to price one commodity below AC to stimulate demand for another. While the analysis restricts itself to the case of complementary goods one can immediately observe the possibilities of charges of predatory pricing so often levelled by potential competitions against established incumbents.

By placing emphasis on market possibilities as opposed to the technological cost function Weldon's paper could also serve as a useful beginning for a study of the rise of the multinational corporation. By being established in one line of business and by possessing its fixed inputs, the addition of extra production lines becomes a matter of time. With new lines and with the exhaustion of plant subadditivity, such firms already with capital in place may find it easy to extend operations. Not only does product x make the cost of product y cheaper in plant A, but now the existence of plant A enables plant B in a different locale to be established at lower cost. In addition to reaping and exhausting all internal scope economies -- managers,

workers, fixed plant, etc., the decision to move to another region may be prompted both by transport savings and also by the ability to react instantaneously to changes in demand. Having already reached such a size, the onus on reaping the benefits of manipulating the market rests entirely on managers. This internal explanation of transnational corporations is admittedly very limited. But the hypotheses that the technology which accompanies concentration is of primary importance in analysing current structure, is one that cannot be easily discarded.

The mechanism thus moved from indivisibilities of fixed input to multiproduction -- then to increase in size -- then to input specialisation as a result of size -- managers, R&D departments, etc. -- and then the decision to go abroad when economies of co-ordination with specialised manpower and equipment became feasible at very large sizes. With the current debate and upheaval in industrial organisation this early exploratory paper may finally be recognised. But Weldon, as previously pointed out, was not the only writer to explore the multiproduct firm. Robertson in a series of lectures[19] discussed the possibility of firms producing goods that were inherently related. As such the increased production of one usually depresses the price

of the other; so much so that it may be desirable to curtail production of the first so as not to spoil the market for the second.

A more technical treatment of the multiproduct firm was provided by Pfouts. After examining the optimal condition of production for the single product firms (as developed by Samuelson, 1947) and comparing them to the condition of productive efficiency in the multiproduct firm, he concluded that it was inaccurate to treat this firm as a series of single product entities. The difficulty for him lay in the presence of fixed factors.

> "The importance of fixed factors in the theory of the multiproduct firm arises in part from the possibility of transferring units of a fixed factor from use in producing one product to use in producing a different product. This serves to bind the production of different products together. Thus within the firm, each product is competing with all the firm's other products for use of the available fixed factors. Therefore the multiproduct firm cannot legitimately be regarded as a collection of single product firms."[11]

One can of course argue with Pfouts that the products do not necessarily compete for the fixed inputs but rather complement each other through scope. If indeed such products competed with each other for the fixed inputs, one would find that the vector of outputs would be a function of the market price since this would

determine how much of x, y and z should be produced to maximise profits (given the constraints of fixed costs).

The writer also suggests that there is some cost of retooling and refitting, etc. whenever another product must be manufactured. But is it not possible that these costs are more than overcome by the complementarities in production?

Extending Pfouts' logic further, one finds that where products compete for fixed inputs, then it is the market that determines how many varieties of output are produced. For, if the alteration costs are large, a firm might well prefer to produce only good x and reap economies of specialisation. But it chooses to produce y and z as well, foregoing scale economies which may exist in x. Why? Because the market for x is too small and the fixed inputs can only be effectively utilised by producing y and z. Of course, the revenues from y and z must at least cover their usage of the fixed cost. The analysis of Pfouts thus rests greatly on the constraints set by the market and not on any technical incentives that allows for multiproduct activity.

The use of the technique of linear programming constrained the author to look only at minimising the

cost involved in multiproduct activity without considering the more interesting and invariably rewarding dual of maximising gains to entrepreneurs. If, by minimising the cost of transfer, one arrives at a vector of 10x, 5y and 4z (given the fixed costs F), it is crucial to determine whether this level will maximise profits. It may well be that the entrepreneur should produce 20x instead of 10x due to the favourable price for this commodity and a smaller amount of the other two. Only in a perfectly competitive economy will the dual of cost minimisation be profit maximisation.

Furthermore Pfouts argues that

> "It is only in the case in which there is excess capacity in all fixed factors...(that) it can be claimed in any meaningful sense that the multiproduct firm is simply a collection of single product firms."[12]

This is an observation that one can safely challenge. In the presence of excess capacity one finds the conditions amenable to multi-output activity and since such excess capacity facilitates the production of other lines, it places the incumbent at a decided advantage.

If firm A produces x, y and z (due to the existence of excess capacity) and entrant B produces y alone, it

means that the average cost of y in B is greater than in A. For while B must purchase new fixed equipment just to produce y, Firm A, already in possession of such capacity to produce y, can in any short period sell this good at a price that covers only the variable cost of production. The ownership of excess capacity provides an incentive to become a multiproduct outfit and effectively protects this firm from entry. In opposition to Pfouts, then, one may well argue that it is excess capacity that distinguishes the single product firm from the multiproduct firm. It is this excess capacity that allows all the benefits of joint production to be reaped.

From this brief discussion of early explanations for the multiproduct firm we can now focus on the contemporary contributions. The theory of the Natural Monopoly will be discussed at some length in order to highlight the basic thrust of the new approach to the multiproduct firm.

NOTES & REFERENCES TO CHAPTER2

1. Alfred Marshall. (1961) Principles of Economics, 9th ed., p. 360, Book V, Chapter VI, McMillan & Co. (LTU).

2. E.R. Dewsnup. (1914) "Railway Rate Making," American Economic Review, 4 (Suppl.), 1914, pp. 81-100.

3. Sune Carlson. "A Study on the Pure Theory of Production", Loc. cit.

4. J.C. Weldon. (1948) "The Multriproduct Firm" Canadian Journal of Economics and Political Science, pp. 176-190.

5. W.J. Baumol, J.C. Panzar, R.D. Willig. Contestable Markets and the Theory of Industry Structure, Harcourt, Brace, Jovanovich, Inc., N.Y.

6. H.R. Varian. Loc. cit., p. 32.

7. W.J. Baumol. (1982) "Contestable Markets: An Uprising in the Theory of Industry Structure," American Economic Review, March, 72, 1-15.

8. J.C. Weldon. Op. cit., p. 184.

9. Elizabeth Bailey and Ann Friedlander in their review article (op. cit.) stress the need for further research into the sources of economies of scope.

10. D.H. Robertson. (1957) Lectures on Economic Principles, London, Staples Press, Vol. 1, Chapter XIII.

11. R.W. Pfouts. (1964) "Multiproduct firms vs. single product firms: The theory of costs and production," Metro-economica, Vol. 16, p. 54.

12. R.W. Pfouts. Loc. cit., p. 657.

CHAPTER 3

THE THEORY OF NATURAL MONOPOLY

"The Beginnings of the New Approach to the Multiproduct Firm"

Natural monopoly has always attracted the attention of economists interested in market failure. Its study is associated with such names as Cournot, Dupuit, Hotelling, Coase and Baumol.

However, as Coase[1] pointed out, natural monopoly is a very elusive phenomenon, lacking any clear definition. A declining average cost curve which seemed a likely defining characteristic lost some of its potency when it was observed that it is possible for average cost to be falling while marginal costs are rising over particular output ranges. Because of this possibility, not only must the cost curves be known, but a level of output must be specified to correctly identify the natural monopoly. In the single output case this is a difficult task. In the multiproduct situation the task becomes impossible.

The view presented by the authors Baumol, Bailey, Panzar and Willig claims to have provided a clear unequivocal definition of this concept and to indicate that even though a natural monopoly exists, it may be vulnerable to wasteful entry.

3.1 The Unsustainable Natural Monopolist - Overview

Whenever the cost function is strictly subadditive a natural monopoly is said to exist, i.e. when one firm can produce the entire vector of output or outputs at lower cost than any combination of other firms each producing the same or a different level of output then the cost function is subadditive. (This concept of subadditivity applies both to the single and multiproduct case. The writers in question however focus attention on the multiproduct outfit which they believe prevails in the advanced economy.)

If, furthermore, the monopolist could find a subsidy free vector of prices that deters entry and yet allow him to break-even then he is not only feasible but also a sustainable natural monopolist.

The multiproduct analysis is thus integrated into the structure of thought. Even if the monopolist produces one commodity -- transport -- he still might be considered a multiproduct firm in that he engages in grain, manufactures, livestock and human cargo all with varying elasticities of demand and different characteristics. Though he may experience constant returns to scale, by altering his scope or diversifying his output (i.e. same one hundred cars per train but more grain and less manufactures), he could experience cost savings or revenue increases.

If the monopolist is the single least cost supplier of outputs, then it would appear that he is invulnerable to entry and so is able to devise subsidy free prices to protect his position. But this is not necessarily the case, as is demonstrated by Faulhaber[2] who showed that sub-additivity does not necessarily imply sustainability.

Although the cost function may be subadditive in the sense that one supplier is the most efficient producer of the vector, that supplier may become vulnerable to entry if he is producing on the rising portion of his average cost curve or ray average cost curve. The rising portion of his ray average cost curve implies that the producer does not operate at least ray average cost thus leaving

him vulnerable to a potential entrant who becomes aware of this opportunity.

As such, entrants not constrained to serve the entire market demand will, by operating on a lower point on the ray, be able to enter, price lower and steal the market of the efficient monopolist. Had the latter been operating inefficiently (eventhough the cost function is subadditive), the entrant may very well produce on a different ray altogether; i.e. a lower ray, thus benefiting from the slack of the incumbent as well as the technological benefits of least ray production.

This result presents serious policy implications for the regulation of multi-output firms. For even if the internally efficient monopolist's cost function is subadditive but he does not operate at the point of minimum ray average cost then an entrant can invade, produce Ye (entrants output) leaving the remainder for the incumbent. The problem of wasteful entry emerges thus when the monopoly is left to produce only this remainder. In the absence of the full slate, Ym (monoplist pre-entry output) scope economies are not exhausted and the benefits of simultaneous production, cost complimentarities and the efficient use of scarce inputs are not obtained. The cost of [Ym-Ye]inevitably

rises so that the net effect may be an overall increase in costs over and above what would have been incurred had the multiproduct monopolist been protected.

This unsustainability of the multiproduct natural monopolist and the failure of the competitive market to yield an optimal result with free entry is, to this writer, the chief contribution of the entire literature.

We proceed next to a discussion of conditions necessary for both feasibility and sustainability, i.e. how a natural monopolist may prevent entry and yet remain feasible. This involves a discussion of the weak invisible hand theorem and the various restrictions on behaviour that it imposes. Furthermore, it is also seen that even this weak invisible hand fails on occasions and, in the event of unsustainability, the need for some form of intervention may arise. But first an example demonstrating unsustainability will be provided to highlight some common problems and a possible solution.

3.2 Unsustainability -- A Game Theoretic Approach

We have just seen how a natural monopolist, through being the single most efficient producer of a vector of outputs, may yet be vulnerable to entry by non-innovative firms. This observation was first demonstrated by Panzar and Willig[3].

Earlier writings on the subject of natural monopoly did not focus on this anomalous possibility because their concept of a natural monopoly, i.e. MC being everywhere lower than AC, is different from the definition used by the new approach.

Here a natural monopoly is defined as one in which the cost function is subadditive, i.e. a single producer can, by individual operations, produce his output vector at a lower total cost than the sum of his his separate rivals can, in each and every firm and output configuration. But even if the monopolist's cost function is subadditive and he enjoys the benefit of economies of scope, scale, cost complimentarities and trans ray convexities, he may be forced to supply total market demand by producing on the rising portion of his ray average cost curve which is still subadditive but

which is not analogous to least ray average cost production. Herein lies the essence of unsustainability.

To provide an intuitive explanation of this unsustainability concept, use will be made of a simple game which lacks a core.

We devise a game between an incumbent monopolist and a series of potential entrants. (Following Faulhaber)[4]. In this example, the monopolist produces only one output (electricity). But the reasoning and logic indicates how unsustainability could apply to the multiproduct case as well.

Suppose that to serve three towns with electricity costs $660 in total, $400 for any two and $300 individually. This cost function is subadditive since single firm production is least costly.

	Average Cost
i.e. C[A + B + C] = $660	220 = AC_3
C (A+C), C (A+B) or C (B+C) = $400	200 = AC_2
C (A) or C (B) or C (C) = $300 alone	300 = AC_1
C (A+C) + C (B) =400+300=700	233 = AC_4

Since the average cost of serving any two alone is less

than the average cost of serving three simultaneously and sice an entrant can with the same plant and equipment enter the market, there are no prices that will prevent entry.

For if A and B contract with a entrant to supply them at $200 each, then there is the incentive for C to seek another supplier and make unprofitable the existing agreement of A+B. Since C must now pay a price of $300 (the stand alone cost of supplying himself) then he has the incentive to contract with another supplier to produce at $400 while he secretly negociates with either B or A. If C is willing to pay more than $200 (the price that A+B pays together) then the entire scenario becomes one of unstable reactions.

By C deciding to pay $220 per supply, B or A can now, by contracting with C, pay $180 each which is less than they ordinarily will pay if all were supplied by the same plant. The game of collusion, strategic interactions and unstable promises, therefore, leads to a situation that is worse for all. No player will contract with a supplier if he knows that the left out party (C) will offer a better deal to him. As long as there is thus an excluded partner in the game who is willing to pay an amount $AC_2 + E > AC_2$ up to his stand alone cost of

serving himself, the natural monopoly will be unstable,

i.e. $C(A + B) = \$200$ each $C(C) = \$300$

now with <u>entrants</u> and C contracting

$C(C) = \$201 < 220 < 300$ $C(B)$ or $C(A) = \$199 < 200$

$\$201 < AC_3 < AC_1$ average cost

(stand alone cost) with two players

with unsustainability arising from

$AC_2 < AC_3 < AC_1$

Free entry here in a perfectly contestable environment leads, therefore, to an unsustainable solution. However, one must investigate the causes of this dilemma -- why can a natural monopolist not sustain himself against free entry. What prevents him from at least price discriminating and keeping the market all to himself? The answer lies in the configuration of his cost curve. If he charges a higher price to one group of consumers, entry becomes feasible in that market. If he produces at least ray average cost, the possibility of entrants contracting with excluded parties becomes a possibility. By producing the entire vector he invites an entrant to produce at the point of least ray average cost. The entrant can now produce the smaller vector which keeps supply on the dominated revenue curve.

Here both ray average cost and price is thus lower than the corresponding point of production of the monopolist. Perfect information about costs on the part of consumers will lead to a series of circumstances detrimental to all. But this perfect information on the unsustainability of the natural monopoly should precipitate a solution acceptable to all. Demsetz style contracting (where production rights are obtained by competitive bidding)[8] ensure our two consumers a low price and the other a higher price with the option of taking it or refusing the service. The failure of all players to reach any agreement individually does not, however, rule out any form of collective agreement. All participants stand to lose by aiming to get the most out of the contract. If they each know that their dual contracts are inferior at each stage to another possible contract with the neglected party and if such knowledge is freely available, then a likely economic solution is to undertake a collective long term contract where each player wins. In the absence of a common agreement the service may not be provided and even if it is provided by an incumbent monoplist -- the service (because the producer is unsustainable) will be discontinuous. Final optimal outcomes in the presence of common information on costs is not an impossible result. Where there are large numbers of consumers and bargaining costs are high,

however, the solution may call for intervention in the form of the regulation of entry into the market.

From this brief exposition of the unsustainability of the single product natural monopolist we now focus our attention on the necessary and sufficient conditions for sustainability in the multiproduct environment. The concept of subadditivity will also be explored.

3.3.1 Sustainability and Subadditivity -- The Multiproduct Natural Monopolist

We have seen that unsustainability arises when the monopolist produces on the rising portion of his average cost curve (which is subadditive) and as such he does not operate at a point of least ray average cost.

Another cause of this phenomenon is the existence of strong product specific economies of scale coupled with weak cost complementarities yielding a cost function that is not strictly trans ray convex. The problem is further aggrevated by the existence of positive cross elasticities of demand for one of the monopolist's products. If, therefore, the multiproduct monopolist raises the price of a particular product to subsidise

another, an entrant can reap a profit if his product is a close substitute for the first. Free entry, therefore, causes the monopolist to lose this product and the cost complementarities that go with it.

The true losses from free entry thus occur when the monopolist no longer produces his entire set ($y_1 \ldots y_n$) and so cannot reap scope savings. His remaining set [$y_1 \ldots y_n - 1$] therefore increases in cost with the possibility that overall costs in the economy of $c[y_1 \ldots y_{n-1}] + c[y_n]$ are greater than if the entire set were produced by the monopolist alone. The natural monopolist is thus not only unsustainable but subjected to wasteful entry as well.

The natural monopoly, it is useful to note, emerges in the competitive market precisely because its cost function is subadditive. It is not seen as a market failure but as an example of market success. The process of competition ensures that technical efficiency prevails so that if one firm is cheaper than a multitude then it will emerge.

But the market in turn causes the monopolist to be vulnerable to wasteful entry (after a certain time period) for though this firm may possess an absolute

advantage in producing a set of products it does not necessarily possess this advantage in producing each and every one of its outputs. The whole is definitely cheaper than the sum of the parts. But then a few of the parts may prove to be quite enticing for not all these parts are of equal value. It is the existence of these choice parts that can be profitably marketed in isolation that causes the monolith to become vulnerable.

The market succeeds on the one hand by allowing the monopoly to emerge but fails on the other by sometimes allowing it to be invaded after it has become established. Baumol et al, however have shown that under a set of restrictive conditions the natural monopolist could by the second best welfare rule utilise the Ramsey prices which maximise welfare subject to a profit constraint, and that this set of prices for the multiproduct firm producing outputs with varying elasticities can, if adopted, yield the vector of sustainable prices which enables the monopolist to at least break even and yet prevent entry.[4] In a contestable market, moreover, such an outcome is analogous to competitive market behaviour where with the 'weak invisible hand', society obtains the blissful results of natural monopoly efficient production as well as the benefits of maximum consumers' and producers' surplus.

But, in instances where no entry deterring Ramsey optimum exists the requirements for government protection emerge. From this brief overview we turn to a fuller exposition of the underlying theory with the hope of demonstrating the use and power of the analysis of a multiproduct firm.

3.3.2 The Subadditive Cost Function -- An Investigation

> "A cost function is subadditive for a particular output vector y when y can be produced more cheaply by a single firm than by any combination of smaller firms."

This notion is at the heart of the theory of natural monopoly.

$c(y) \leq c(x) + c(y-x)$ for all $x \leq y$

i.e. $c(y) \leq \sum_{i=1}^{n} c(x_i)$

where $\sum_{i=1}^{n} x_i = y$

This is of course the single product concept of subadditivity and to extend it to the multiproduct case, one requires the definition of ray subadditivity. "A cost function C(.) is said to be strictly ray-subadditive at y if, for any set of two or more positive numbers Vi that sum to one,

$\sum C(Vi\ y) > C(y \sum Vi) = C(y)$"

where $y = y_1, y_2, \ldots y_n$.

At this point on the ray, it is clearly cheaper to

produce the full state $Vi = 1$ rather than producing only portions of it.

It is now useful to examine the causes of subadditivity and to determine the extent to which they could assist in the proper identification of a natural monopoly. But before this is done it must be observed that "subadditivity is a global not a local concept. Specifically, (i) to determine that a cost function c(.) is subadditive at some output vector y it is necessary to have information (explicit or implicit) about the value of c(.) for every possible output vector smaller than y, that is for all $y^* \neq y$, $y^* \leq y$."[9] Unlike the property of scale economies at output y it is impossible to assess subaddivity only in this range of output y. "The cost function must be scrutinized not merely in the neighbourhood of that point, but also all the way to the axes and the origin...thus to prove subadditivity we must have information on the costs of every potential small or intermediate producer...."[10]

As a theoretical device, however, it is novel in that for the first time it is postulated (according to the authors) that the internal cost function of a firm will lead to the associated industrial structure.

Here then are some sufficient conditions for subadditivity as outlined by Baumol et al in their new book.[11]

(1) Existence of economies of scope up to the level of output y_i i.e. $c(\sum_i^n y_i) > \sum_i^n c(y_i)$ This is the only necessary condition.

(2) Decreasing Ray Average Cost --
$c(y) \equiv \sum_{i=1}^n U_i\, c(y) < \sum_{i=1}^n c(U_i y);\ \sum_{i=1}^n U_i = 1,\ U_i \geq 0.$
i.e. no proportional division of output between firms will reduce total cost.

(3) Decreasing average incremental cost up to y and economies of scope at y imply multiproduct economies of scale and hence subadditivity at y. Hence, since there is decreasing average incremental cost in one product line, this must be monopolised to save costs.

(4) Trans ray convexity along any one hyperplane together with declining Ray Average Costs provide another condition.[12]

(5) If c(.) is supportable at y^0, then c(.) is subadditive there.

This notion of supportability implies that there exists a

price output vector which just allows the firm to cover its costs. "Consider any non zero pair of output vectors, y^a and y^b, such that

$y^a + y^b = y^o$. Then by supportability, $c(y^a) > y^a$. $h(y^a)$ and

$c(y^b) > y^b . h(y^a)$.

Adding ..... we have $c(y^a) + c(y^b) > y^a + y^b . h(y^o) = y^o . h(y^a) = c(y^a)$[13].

These five conditions are each sufficient to guarantee that a cost function will be subadditive: such conditions as these contribute to the birth, growth and maturity of natural monopolies. The question which invariably arises at this point is, given these conditions on the cost function, does a price-output vector automatically exist that will enable the monopolist to be feasible and yet deter entry? We see that this requirement of sustainability is akin to the notion of equilibrium (a state of rest) when no motives exist for disturbing the status quo and when all actors are satisfied. A sustainable price output vector as we shall see in a contestable market is an efficient configuration that minimises cost and only allows prices consistent with the first best welfare maximising level.

3.3.3 Sustainability

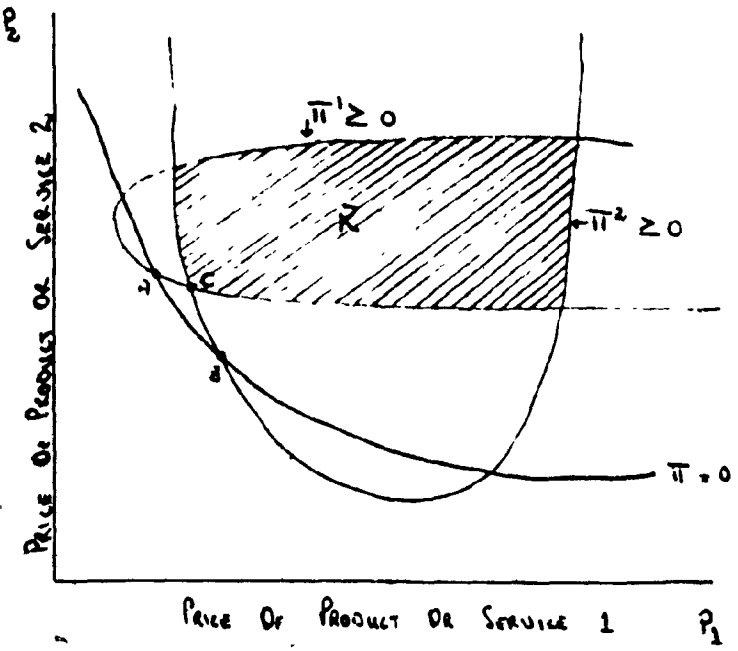
> "A vector of prices p^* is sustainable for a set of incumbent firms in an industry if the incumbent firms are financially viable at those prices and if no potential entrant can find a marketing plan for which the anticipated economic profits $p^e y^e - c(y^e)$ exceed the costs of entry $E(y^e)$."[14]

One can immediately see that a subadditive cost function is not always a sustainable one. Whereas the old notion of a declining AC[13] as the characteristic of monopoly basically ruled out marginal cost pricing when MC lay everywhere below AC, the new concept in its definition implicitly permits the anomalous and aberrant behaviour. Any firm which is cheaper in a set of productions in total may not be cheaper in each individual item and some firms may thus possess the ability to service one particular line of output at lower cost than the monopolist. A product may not experience cost complementarities when produced with another but the cost of the other will rise if the first is removed. That is to say, though the second needs the first the first does not necessarily require the second to lower its cost. The possibility of entry can thus be discerned in the stand alone product with the dependent commodities rising in cost if the monopolist loses production of that which is vulnerable. It is important now that "we investigate

whether there exist any prices which if chosen by a natural monopolist, will permit the monopoly firm to operate profitably in contestable markets and yet discourage any and all entrants who offer no innovation."[16] These conditions can be listed as follows:

1. Existence of Undominated Prices

There exists a price p^* such that there is no price $p < p^*$ for which $\pi(P) > \bar{\pi}(P^*)$, i.e. there must be no lower price than P^* that will yield a profit to an entrant. This condition implicitly imposes the condition of production at least ray average cost, violation of which, we saw before, leads to unsustainability. This can be demonstrated by considering a graph of conditional profitability given prices P_1 and P_2.



The figure provides a representation of sustainability in price space and represents an industry producing two substitute products.

> The curve $\pi = 0$ is the locus of undominated P P pairs which just permit a firm marketing both products as a monopoly to break even.
>
> The set $\pi' \geq 0$ represents all those pairs of prices which permit a firm marketing only good 1 to make non-negative profits. That is, for any given value of P charged by some other firm supplying good 2, this set includes the values of P_2 which offer non-negative profits to a firm supplying only good 1. The $\pi^2 \geq 0$ set is defined analogously.
>
> The intersection, R, of these two non-negative profit sets consists of all price pairs at which the two single product firms can simultaneously avoid losses. At point E both such firms earn exactly zero profits. It follows from the hypothesis that the industry is a natural monopoly that the $\pi = 0$ locus must pass to the 'south west' of R. That is, it tells us that a multiproduct firm can break even by marketing the pair of outputs at prices lower than those which independent producers of the items would have to charge. This is all the natural monopoly condition tells us ..... It is easy to see (however) that an undominated zero-profit point such as C cannot be sustainable, since a firm marketing only good 1 can charge a price slightly less than the monopolist and earn positive profits.[17]

Only if the incumbent monopolist restricts itself to a price vector on the arc AB will he be able to deter entry. Any other point on the locus $\pi = 0$ will invite entry into one product line causing the least cost producer to lose his cost advantages.

Having eliminated the possibility of dominated price vectors and non least ray average costs, it is important to note that the monopolist must be efficient, i.e. his cost function must be subadditive and he must take all

advantages of this subadditivity.

On the pricing side there are also some rules to be followed and by assuming a contestable market the first of these will be that

1) $P \nless MC$

If the multiproduct monopolist is already pricing at the limit where he effectively deters entry, then any $P < MC$ means that an entrant can reduce from his production plan all unprofitable units, produce where $P = MC$, and steal demand from the incumbent. Our monopolist therefore becomes unfeasible by inviting efficient entry. We focus now on our second primary condition which favours sustainability.

2. Undominated Prices and Trans Ray Convexity

As before the cost function must be trans ray convex and the price vector undominated. This, therefore, implies that the cost function is supportable at Y (the level of output in question). An entrant cannot hope to market the same product by producing a smaller quantity and hope to break even. Of course, this does not prevent an aggressive entrant from modifying slightly a product and earning positive profit.

Finally in a perfectly contestable market price must not be less than MC; there can be, then, no form of cross subsidy. In the case of natural monopoly with decreasing ray average cost and cost complementarities, however, a Ramsey-optimal vector that satisfy the conditions for pareto optimality under a profit constraint will by the force of the "invisible hand" allocate fixed costs so that they are shared among all the products sold by the firms (thus violating the first best principle).

> Here with prices held fixed at their market levels, the derivatives of profit with respect to quantities are proportional to the corresponding derivatives of consumers' plus producers' surpluses..... The monopolist knows that his profits are limited to E (the value of entry barriers) .... and that the profit of an entrant will be no greater than pseudo profit (pseudo revenues less costs, including the cost of entry) calculated at the monopoly market prices held fixed. Hence if he chooses an output vector at which profit is equal to the entry cost (so that pseudo profit is zero) and which at his fixed prices happens to maximize pseudo profit over T (the potentially profitable region) then pseudo profit must be less than zero everywhere else over T. That is profits of all potential entrants must be negative and the monopolist unique market position is guaranteed to be sustainable.
>
> The monopolist who seeks stationary prices that can protect him from entry, like the perfect competitor, has an incentive to choose outputs that maximize profits calculated at those parametrically fixed market prices. But by doing so, the firm inadvertantly maximizes net social welfare. Thus, the same invisible hand that guarantees welfare-optimal pricing under perfect competition, may guide the far sighted monopolist, seeking protection from entry to the Ramsey welfare optimum .... It should be noted however that even sustainability of Ramsey pricing does depend on global properties of the cost function.[18]

These properties are notably declining ray average costs together with trans ray convexity.

One, therefore, observed in an intuitive fashion four major effects impact and become essential for sustainability of a natural monopoly. These are its own price elasticity and cost complementarities in production together with cross elasticities (existence of substitutes) and product specific economies of scale in one product line. Cost complementarities and independent demands (non existence of substitutes produced by entrants) favour sustainability but decreasing average incremental cost and strong product specific returns to scale, weak cost complementarities, strong substitution effects in demand and non least ray average cost do not.[19] Where there is a supportable price output vector with no product specific fixed costs the adoption of the Ramsey optimum prices will guarantee that the feasible output is not only sold but will allow for both efficiency in resource use and pricing. In the absence of these stringent conditions no sustainable prices may exist and free entry may result in waste of resources. A policy option here is one of total protection of the natural monopolist so that he alone utilises all the economies of scale and scope while eliminating latent excess capacity in production. But immediately one

encounters a major problem.

In the absence of competition (actual or potential) the monopolist may have little incentive to become as efficient as he would be if the threat of entry persists. The threat of entry on the other hand hinders his scope of operations and renders intertemporal allocation of resources sub-optimal. Clearly a dilemma confronts the policy maker.

To this researcher the problem is not as grave as it may seen at first sight. Many factors are involved that must now be explicitly introduced. In a regime of open entry the products of the unsustainable monopolist will rise in cost. Questions then must be asked regarding the availability of substitutes for these high cost commodities, their place in terms of overall welfare (however defined -- basic need?), their importance in the national economy and the losses sustained in other parts of the economy with a possible loss in consumption. The latter indirect effects could very well be the most important, especially where the (n-1) products are used as inputs into other industries that now will experience a rise in cost. Knowledge of the overall total effect of entry is essential for policy purposes and only a cost benefit analysis accounting for all repercussions will

provide any meaningful guide to decision making. There may then be a role for government in the industrial economy. Even the weak invisible hand fails on occasions. That it is theoretically proven (via unsustainability) that the market could fail with free entry, potential competition, contestable markets and varying industrial structures is one of the most innovative contributions to have emerged from the overall uprising in industrial organisation. (Other economists have seen the matter differently. For Spence[20] the main contributions lay in conceptualising technical and meaningful terms for the multiproduct firm and for Brock[21] the overall contribution lays in the systematic and formal characterisation of industry structure.)

The fact remains that there is a case when the market fails and a reduction in total cost can only emerge with some form of policy directive. We focus attention now on the weak invisible hand theorem.

3.4 Natural Monopoly and the Weak Invisible Hand Theorem

The weak invisible hand theorem utilizes the optimality of Ramsey Prices subject to a profit

constraint (where this constraint may be zero). As first postulated by Frank P. Ramsey[22] and considerably modified, these prices, which represent the relationship between the elasticity of demand and the marginal cost of a set of products, are the second-best optimum prices in a multiproduct monopoly constrained to cover at least its total cost. Baumol has proven that under a set of restrictive conditions these prices which are the optimal prices for a natural monopoly as far as welfare maxima is concerned, will be selected by a monopolist who is guided by no such altruistic goals but only by the force of his own self interest. By so doing the monopolist also will find himself sustainable against entry.[23]

One could accept this result as is but it still would be worthwhile to examine its implications. In terms of our model the Ramsey prices will allocate common fixed costs among elastic and inelastic goods and it is this commonality of fixed inputs that causes such a discriminatory pricing scheme. We shall endeavour to show the sub-optimality of these Ramsey prices.

In order to obtain the result of the weak invisible hand, both the cost function and the revenue hyperplane must be tangent at one point and one point only. At output levels up to this point the vector of outputs

displays trans ray convexity and least ray average cost. The absence of product specific economies of scale in any one line of output as well as the absence of close entrant substitutes are crucial assumptions that are made. For example, the presence of trans ray convexity (and supportability) and the absence of least ray average cost production means that an entrant can invade, produce at least ray average cost and earn a profit at a lower price.

Furthermore the price vector must be undominated. This new concept (as we saw) entails the fact that at no lower price can the entrant produce and hope to make a profit. He is thus constrained to produce the entire set of goods of the monopolist or else see if there is any chance of trans ray convexity and scope economies at a lower output level. In order to be sustainable then the natural monopolist must reap all economies of joint production (scope, scale, cost complimentarity). He must, furthermore, earn zero profit after rent. "That is if E(y) is considered as a rent and is included in the monopolists' costs, sustainability requires that the monopolists's total revenue not exceed those total costs."[24]

These conditions accrue at one point of production

where different quantities of outputs are produced. Since he must sell these products in order to break even, reliance is made on the elasticities of demand for the various products. Pricing must be such that the outputs are sold to provide revenues sufficient to cover total cost. We see then that those goods whose demands are elastic will be priced at a lower level than those facing inelastic demands. The final result is one where fixed costs are allocated, profits are limited to the extent of natural entry barriers, entry is deterred and the sum of consumers and producers surplus maximised.

This result, as will become apparent, is highly dependent on the existence of a series of outputs with varying elasticities using common production facilities. In the presence of product specific fixed costs, those inelastic goods that bear a greater burden than their own fixed cost requires will become vulnerable to entry by any firm willing to enter. There may thus exist a Ramsey price in the presence of product-specific fixed cost that is vulnerable to entry.

As Boadway[28] has further observed, all cross elasticities of demands must be zero. A positive cross elasticity denotes a substitute. As the price of the inelastic good goes up, quantity demanded of substitute

products increases leading to a loss of sales of the monopolist and so the unsustainability of his revenues.

The firm must also be a price taker for the inputs it purchases, there must be no consumption externalities between the goods in question and the rest of the economy must be perfectly competitive.

One also observes that such a price is the optimum optimorum in the sense that natural monopoly will reduce resource use while maximising the various surpluses. But the price vector itself ignores the effect on equity of such a policy. The inelastic good will bear a high price whereas the consumers of such a good may not be in a position to retaliate. If domestic telephone service, for example, is inelastically demanded, then its price will rise higher than the cost of its production. Such a rise may reduce the disposable incomes of consumers in the lower stratum of the income ladder. In such a case consumers of domestic service will subsidise long distance users (where demand is assumed to be price elastic). The Ramsey schedule of prices can thus be seen as one which places heavy emphasis on drawing welfare conclusions from the notion of consumers surplus and elasticity of demand without investigating the welfare implications in terms of equity. Thus according to this

rule those who cannot escape will pay.

NOTES & REFERENCES TO CHAPTER3

1. Coase, R. (1970). "The Theory of Public Utility Pricing and its Application" Bell Journal of Economics 1, Spring 1970, pp. 113-129.

2. Faulhaber, G. (1975). "Gross Subsidization: Pricing in Public Enterprise," American Economic Review, 65, December 1975, pp.966-977.

3. Panzar J.C. and Willig R.D. (1977). "Free Entry and the Sustainability of Natural Monopoly," Bell Journal of Economics, 8, Spring 1977, pp. 1-22.

4. Faulhaber op. cit.

5. Demsetz loc. cit.

6. W. J. Baumol, E.E. Bailey and R.D. Willig. (1977) "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct Monopoly," American Economic Review, 67, June, pp. 350-365.

7. Contestable Markets Loc. cit. p.170

8. Ibid., p.175

9. Ibid, pp 171

10. Ibid, pp 171

11. "The reader should be forewarned, however that all these conditions are excessively strong in that they contain many element that are clearly not necessary for subadditivity. In fact of the cost concepts presented so far, only the condition of economies of scope is necessary for subadditivity since it is required for subadditivity over orthogonal output sectors and therefore is a prerequisite for overall subadditivity." Ibid., p. 174

12. "The argument relies upon the fact that given any subdivision of an output vector $y^0 = y^1 + y^2$ it is possible to express y^0 as a convex combination of vectors constituting proportional expansions $v_1 y^1$ and $v_2 y^2$ of the outputs in y^1 and y^2 i.e. $c(y^0)$ is no greater than a weighted average of these costs". Ibid. pp. 180.

13. Contestable Markets, Loc. cit., pp. 185

14. Contestable Markets, Loc. cit., p. 9.

15. Note that average cost may be falling while marginal cost is rising. Only where MC is below AC at all points will AC continue to fall.

16. Contestable Markets, Loc. cit., p. 192.

17. Ibid., pp. 196

18. Ibid., pp. 215

19. In symbolic language this becomes
$$\frac{\partial Q^2}{\partial P_1} c_1(q^1 q^2) - c_1(Q^1 0) + \frac{\partial Q^2}{\partial P_2}$$

$$(1-s_2\{Q^1, Q^2)\, c_2(Q^1 Q^2) \geq 0.$$

20. Spence, A.M. (1983). "Contestable Markets and the Theory of Industry Structure: A Review Article," Journal of Economic Literature, September 1983, 21, pp. 981-90.

21. Brock, William A. (1983). "Contestable Market and the Theory of Industry Structure: A Review Article," Journal of Political Economy, December 1983, 91, pp.1055-66.

22. F.P. Ramsey. Loc. cit.

23. Baumol et al (1977). Loc. cit.

24. Contestable Markets, Loc. cit., pp. 201

25. Boadway, Robin (1979). Public Sector Economics, Winthrop Publishers, Inc., Cambridge, Mass., 1979, pp 346.

CHAPTER 4

4.1 Economies of Scale and Economies of Scope

"Most important, the new literature has introduced as a complement to the old concept of 'economies of scale', the new concept of 'economies of scope' which measures the cost advantages to firms of providing a large number of diversified products as against specialising in the production of a single output."[1]

$$C(q) < \sum_{i=1}^{k} C(x_i) \quad \text{when} \quad \sum_{i=1}^{k} x_i = q$$

$$\text{and so} \quad C\left[\sum_{i}^{k} x_i\right] < \sum_{i=1}^{k} C(x_i)$$

The above representation of economies of scope implies that producing a vector together costs less than producing the same quantity in separate firms. There are thus advantages to multioutput production. This concept is closely allied to that of subadditivity where the cost of production of a vector is less if produced in one single firm than if this same vector is shared amongst any combination of competing firms producing in any

configuration that they desire. Economies of scope is thus a weaker concept than that of subadditivity. Scope ensures that the cost of simultaneous production is less than single output production but does not preclude the possibility of n firms existing each producing a share of total industry demand, but producing in such proportion that they are able to reap the economies of simultaneous production. The chief raison d'etre of the multiproduct firm is then the existence of a cost function exhibiting economies of scope. Sub-additivity on the other hand indicates that a single firm is the least cost producer of the vector of outputs.

Some of the contributing factors favouring subadditivity are as we have seen, the existence of economies of scope and economies of scale, cost complementarities and quasi-convexity (trans ray convexity). All these supporting causal effects have one thing in common -- they emphasise the overall benefits in terms of cost reduction of joint production as opposed to single output activity.

It is the objective of our exercise now to investigate the causes of economies of scope since, as pointed out by Bailey and Friedlaender, "it is clear that much more research is needed on how and in what

circumstances scope economies arise."[2] The literature so far takes economies of scope as given without investigating their causes. The following is an attempt to identify economies of scope and examine the extent to which this as opposed to economies of scale is the more important source of barriers to entry.

4.2 Economies of Scope in the Multiproduct Firm

Since scope economies play a fundamental role in our analysis and have been implicit in the work of Marshall, Boddy, Weldon and Dewsnup* it will be useful to carefully examine this concept to determine the extent to which it is a separate and new invention or whether it has always existed in the literature.

Scope economies arise when it is cheaper to produce a set of commodities in one firm than it is to produce individually in a multiplicity of firms. They are measured by

$$S_c \equiv \frac{C(Y_1, 0) + C(0, Y_2) - C(Y_1, Y_2)}{C(Y_1, Y_2)} \quad (1)$$

What reasons account for such cost savings to multioutput

production? For Marshall this saving results from the automatic production of an additional commodity with the production of the first. This is the classic case of the joint product and can be observed in the case of mutton and wool, sugar and bagasse and beef and hides. The logistics of production, therefore, enables single firm production to be cheaper, ceteris paribus, than multifirm activity.

Weldon, on the other hand, dealt with multi-output firms consciously producing a set of commodities that are not automatically related via technology. Here such scope economies arose as a result of indivisibilities of factors in place as well as market potential. One therefore observes savings due to efficient use of managers, workers and elimination of transport costs. The latter factor being identified as important, creating an incentive for multioutput production while providing the impetus for multiplant operations as well. His analysis thus extends to plant subadditivity as opposed to firm subadditivity alone.

More recent literature begins with the premise that, since all firms are basically multiproduct in character and hence scale economies cannot be rigorously defined, there exists the need to invent another construct that

will encompass both economies due to size (Product Specific Scale economies and Decreasing Average Incremental cost) as well as savings due to composition. A combination of the two will provide some measure of Ray Economies of Scale -- a proxy variable of scale in the multi-output firm.

The mechanics of this insight will be explored in a little detail to uncover any strengths, flaws or weaknesses that such a measurement approach will entail. First scale economies in the single product case can be defined as

$$S = AC/MC = \frac{C(Y)/Y}{(dC/dY)}$$

and if S>1 then total cost rises less than in proportion to changes in output and the firm experiences increasing returns to scale.

In the multi-output case, where fixed factors are shared, one relies on the Ray Economies of Scale which is a concept (similar to Ray Average Cost) "indicating the behaviour of costs as the production levels of a given bundle of outputs change proportionately; that is, the composition of output is assumed to remain fixed while its scale is permitted to vary."[4]

i.e. Ray Economies of Scale are given by

$$Sn(y) = C(y)/y \cdot \nabla C(y) \equiv C(y) \Big/ \sum_{i=1}^{n} y_i C_i(y),$$

where $C_i(y) \equiv \partial C(y)/\partial y_i$.

$$Sn(y) = \frac{C(Y)}{\sum y_i MC_i}$$

or

$$\frac{C(Y)}{y\,\partial c/\partial y_1 + y_2\,\partial c/\partial y_2 \ldots + y_n\,\partial c/\partial y_n} \qquad \text{eq 2}$$

This is just the relationship of a change in total cost to an expansion of all outputs by a scale factor t, with y representing the composite commodity where output composition is fixed but scale varies, so that $y = ty^\circ$ "This follows because the quantity of this composite commodity embodied in y is t, and because the marginal cost of the composite commodity is $dc(ty^\circ)/dt = \sum c_i(ty^\circ)y^\circ$. Consequently, Sn(y) can be interpreted as the elasticity of the output of the relevant composite commodity with respect to the cost needed to produce it."[8]

But this is weak in the sense that if the output proportions are rigid there is only limited capture of the essential concept of economies of diversification. The choice of weights is therefore highly arbitrary and the true effects of economies of scale in the multi-output unit can be seriously distorted if as output proportions alter there exists the potential for cost reduction. Expanding output while holding proportions fixed (i.e. restraining production to just one ray) could

thus involve a needless increase in costs by the producer.

In addition, accurately calculating multiproduct economies of scale one must not only account for scope economies but product-specific scale economies as well. We have previously seen (chapter 1) that product specific scale economies for good 1 are measured by

$$S1 = \frac{AIC_1(y)}{MC_1}$$

$$= \frac{1}{y_1} \quad \frac{(c(y_1, y_2, \ldots y_n) - c(0, y_2, \ldots y_n)}{\partial c / \partial y_1}$$

Algebraically manipulating the expressions given in equations (1) and (2) for ray economies of scale and economies of scope, Bailey and friedlaender have demonstrated that "the measure of multiproduct scale economies can be related to the measures of product specific scale economies and scope economies (by) .....

economies $S_{12} = \frac{wS_1 + (1-w)S_2}{1 - Sc}$ = multiproduct scale

where $w = \frac{Y_1 \partial c / \partial Y_1}{Y_1 \partial c / Y_1 + Y_2 \partial c / \partial y_2}$

(subscripts 1 and 2 represents two separate goods)

which roughly represents the share of the variable cost of production incurred for product 1. Thus the overall

degree of scale economies for both products is a weighted average of the degrees of scale economies pertinent to products 1 and 2, magnified by economies of scope through the factor 1 / 1-Sc".

Where one faces a cost function that reveals scope economies but exhibits a marked absence of product specific economies of scale one could still experience cost savings due to diversification. Any such observation may prompt one to infer trans-ray convexity where the savings of joint production are great in relation to increasing returns to scale in the single existing lines. As such the existence of trans-ray convexity resulting primarily from decreasing ray average costs up to the relevant levels will ensure that a cost function is subadditive which is the necessary conditions for the existence of a natural monopoly.

Unlike the earlier writers the later scholars concentrate on algebraic manipulation neglecting a thorough exploration of the causes of scope economies. They just postulate that the occurrence of joint production creates cost reductions. One will thus be tempted to ask if any firm can diversify to any limit and experience reductions in costs. Clearly there are limits to diversification and so the causes of scope economies

must be identified to determine if indeed it is a greater barrier to entry than scale.

Nonetheless there has been recognition of this limitation and some descriptive explanations for the sources of scope have been proposed by these later writers.

First the notion of the classical mutton and wool combination makes its reappearance and scope arises due to joint production or automatic production processes.

In addition items such as the re-use of vital inputs such as the production of paper and books provide another reason. Other examples occur in the case of transport industries networking and route combinations which facilitate maximum use of existing facilities (e.g. the use of aircraft and larger planes along the varied routes). In addition, where information is needed by related firms but the market is a poor transmitter, firms are forced to merge so internalising the information processes.

But these are all variants of the most important characteristic of all. This is the indivisibility of fixed inputs in the production of Y that like a public

good becomes costlessly available to the production of x_1, x_2 ...x_n. It is this existence of latent excess capacity and the limitations of demand on Y that enables the multiproduct firm to reap the so-called economies of scope.

Any firms producing x alone will be faced with an average cost much higher than that which produces x with little or no overhead extra cost. Whereas for the former x must bear the brunt of fixed cost for the latter x is available at only the use of variable inputs. It is therefore crucial to identify this character of multi-output production and the comparative advantage of producing simultaneously. To market one commodity may require a fleet of trucks, managers, workers, etc. and when demand is satisfied these fixed factors can no longer be profitably used. They may be working then to full capacity in the production of y but then demand surges for x a product that can be produced at little effort with the plant of y . The incremental cost of x therefore is small, with capital in place, and the managers who work effectively for 5 hours per day when forced to work for six though imposing little exertion upon themselves enables the firm to utilise its idle spare capacity optimally and efficiently.

This argument clearly is a pitch for reconsidering the importance of economies of scale and it is tending to emphasize that in the final analysis it is scale economies (arising from fixed and sunk cost) not scope that is the barrier to entry. The incumbent is able to experience scope savings precisely because of his fixed capital and were they not of such a size as to possess latent excess capacity then it is very unlikely that scope economies will rise.

By scaling down operations from 1000 widgets to ten widgets using a ten-widget plant size, there will exist little excess capacity to enable producton of blue, red and green widgets. It is only when we need to produce 900 blue widgets and this requires a plant size (considering human capital as well) of 1000 widgets that one experiences the latent excess capacity. For even though there is little real excess capacity -- such a large demand or scale of production of blue widgets will require specialized fixed inputs which at this scale of operation may not be optimally utilised. Thus specialist machines, managers, transport equipment and material that is feasible to produce and market 900 widgets because they are more than required and because they are indivisible, will be costlessly available to produce 10 red widgets. The existence of fixed costs and the

consequent emergence of latent excess capacity is thus a primary element in understanding the idiosyncracies of multi-output activities from a purely technical perspective.

This notion, however, only accounts for plant subadditivity which requires that there be at least some indivisibilities in production. The related concept of firm subaddivity has been casually mentioned to emphasize the importance of the efficiency of the firm over the market.

Again in the context of the multiproduct firm this may be an invalid dichotomy where it can be argued that the firm engages primarily in production whereas the market deals with prices sending out the signals needed for the firm to engage in multi-output production. In this approach the separation of the firm and the market are unjustified especially where they both interact doing different tasks. The firm engages in multiproduct activities not because of its efficiency over the market process but because the market instructs it to do so promising in effect an excess of revenues over incremental costs. In the absence of such market information it can be argued that the firm could profitably produce many different commodities but is

constrained from doing so by a lack of information. Any discussion concerning the efficiency of the firm over the market process giving rise to multiproduct firms must thus explicitly consider the informational role of the market and the productive role of the firm.

Concentrating on the productive role of the firm, it is thus imperative to determine the extent to which scope benefits are effective entry deterrents. As Schmalensee' has argued, in the cereals industry an entrant must be prepared to service entire market demand by entering six different brands on the market. This gives the impression that scope economies are the effective barrier to entry. But is it really scope of operations that causes the barrier? Again no definitive explanation has been forwarded. Can it not be argued that the size of plant required to produce one brand is such that the incremental cost of producing five other brands are all diminished because of the existence of plant in place. If such fixed costs do indeed enable a producer to manufacture many commodities, then the true barrier to entry will not be the degree of scope economies but rather its source -- the existence of overhead costs and latent excess capacity.

To say thus that scope economies prevent entry is to

say in effect that firms producing four or five products regardless of their returns to scale is invulnerable to entry. Any firm, it is here argued, which produces a multiplicity of commodities but produces with a sub-optimal sized plant is vulnerable to a more efficient entrant who using scale optimally re-arranges production until he reaches the point of minimum ray average cost. For here our entrant can enter, steal the market and lure customers away at a lower price.

Having made the case of emphasizing the importance of scale as the barrier to entry, it now introduces the question of why this should be the overall deterring factor. Scale arguably is important since to produce efficiently there is need to use larger and larger plants of a size that create complementarities in production, with such benefits being absent at smaller plant sizes. Reaping the benefits of declining average incremental costs (resulting from the large scale plants) means that an incumbent firm can capture a significantly larger share of the market. Any firm not so endowed will be unable to sell the identical products at remunerative prices. As such, it is the advantage of being the incumbent combined with a large scale of operations that provides barriers to entry. Any potential entrant not only faces higher unit cost (in the absence of

complementarities) but also must contend with the incumbent keeping his level of outputs unchanged, lowering prices and driving him out of the market.

From this discussion of economies of scope and scale we now focus attention on the contestable market to examine the extent to which this construct will enable all benefits of the multiproduct firms, natural monopoly and first best pricing to accrue in such an environment.

NOTES & REFERENCES TO CHAPTER4

1. Elizabeth Bailey and Ann Friedlander. (1982) Market Structure and Multiproduct Industries, Journal of Economic Literature, Vol. XX, September, p. 1025.

2. Ibid., p. 1028.

3. Francis M. Boddy. (1940) The influence of costs on production and price policy in a joint product industry, Ph.D. dissertation, University of Minnesota, as referred to by Bailey and Friedlander, op. cit.

   Alfred Marshall quoting Professor Dewsnup in AER Supplement, 1914, p. 89, "Things should be described as joint products when their total cost of production by a single plant are less than the sum of the costs of their production by separate plants" suggests that this definition is less general than his own but that it is convenient for some special uses. Marshall, Principles, Book V, Chapter VI, p. 322, 8th ed., McMillan.

4. Bailey and Friedlander, op. cit., p. 1029.

5. Contestable Market, loc. cit. p.51

6. Bailey and Friedlaender op. cit., p. 1032

7. Richard Schmalensee. (1978) "Entry Deterrence in the Ready to Eat Breakfast Cereal Industry," Bell Journal of Economics, Autumn, 9 (2), pp. 305-27.

CHAPTER 5

5.1 The Theory of Contestability

> "A perfectly contestable economic market is defined to be one into which entry is completely free, from which exit is costless, in which entrants and incumbents compete on completely symmetric terms, and entry is not impeded by fear of retaliatory price alterations."[1]

The characteristics of such a market thus are highly restrictive and in its present form cannot be posited as a description of reality. We will see presently that by modifying some of these assumptions, contestability may prove to be a useful description of industrial behaviour in advanced economies. But first the idea of a contestable market must be examined to investigate its use as a benchmark of comparison for actual performance.

As in the definition we see that

a) Entry and Exit are easy and costless -- this assumes no legal barriers to entry, n o inhibitions on the

cost side and equal availability of technology.

b) Information is a free and public good.

c) Sunk costs are absent i.e. exit is basically costless.

d) Sellers recognise their power in the market.

e) A market with more than one seller is characterised by marginal cost pricing and production in a region of constant returns to scale.

f) Consumers react instantaneously to a change in price. (There is a perfectly elastic demand curve up to some region and a kink thereafter), while entrants produce immediately upon the creation of a profitable opportunity. Incumbents react after a time lag long enough for entrants to reap profits and costlessly exit.

g) Aggressive entrants are always present in the wings awaiting their entry cue.

h) Price is the only argument in the demand function (this is not a specific assumption, but one that is implied throughout).

Such freedom of entry, ease of exit and perfectly horizontal demand curves is enough to ensure both internal and pricing efficiency. Because any profitable opportunity regardless of magnitude invites entry, an incumbent is constrained to act as if he were a perfect

competitor. It is mainly the elimination of entry barriers and the stimulation of potential competition which provide welfare maximising results.

Among these are the absence of internal inefficiency (X-inefficiency); the enforcement of marginal cost pricing, or Ramsey optimal pricing in a monopoly (as dictated by the weak invisible hand), absence of Averch Johnson effects, no cross subsidy or predatory prices and an efficient industry structure, one which produces total demand at least cost.

These felicitous conditions are a result of the absence of sunk cost and barriers to entry, in addition to equal access to technology. It is this equal access coupled with fungible equipment that allows potential entrants to be credible. These conditions are enough to ensure that the large number requirements of perfect competition are no longer necessary to secure the optimal results associated with competitive markets.

As a result of this free exit criterion even natural oligopolies must price competitively. If they did not, potential entrants would replicate their production (same technology), sell at a lower price (elastic demand) and exit with surplus profits. Similarly any upward

deviation from marginal cost pricing would invoke the same reaction. It is not necessary for entrants in the wings to actually enter (though they must on occasions to prove their credibility); all that is required is their presence -- this is enough.

It is easy to criticise the theory on the basis of its assumptions, as we later argue. However, it is our objective now to determine the extent to which the theory in and of itself possesses limitations. In the following sections the integration of consumer behaviour will be attempted to demonstrate that even a perfectly contestable market could yield suboptimal results. In addition, the concept of time plays a central role in the later analysis. We see too that inherent in the assumptions is the concept of price sustainability, but, as will be argued, there are advantages to being an incumbent in such a market and the incumbent himself could pose a credible threat to entrants by keeping output fixed. An inelastic demand curve is also considered, existence of which yields suicidal results for any prospective entrants. Finally the contestable market in a multiproduct environment is explicitly considered and an amendment regarding economies of scope is outlined. This amendment supports the notion that under certain conditions competition among the few in the

absence of collusion is enough to guarantee the results of a contestable market. A large pool of entrants on the wings is not a crucial requirement for contestable results nor for the existence problem i.e. the existence of a sustainable configuration. We shall also see the importance of a contestable market in precipitating efficiency by allowing only least cost producers to emerge who will utilise and exploit all multiproduct economies and yet price at such a level that overall optimality is obtained.

Attention will now be focused on some conditions that will prevent the emergence of contestable markets.

5.2 Fixed vs. Sunk Cost in Contestable Markets[2]

A distinction is made in the new literature between fixed and sunk costs. Traditionally it was felt that large fixed costs (i.e. costs incurred when no output is produced and results from decisions made in the past) were the barrier to entry and any enterprise requiring such fixed outlays were guaranteed some barriers to entry. Here, however, fixed costs are no longer associated with entry barriers and while large fixed

costs may enable a natural monopoly to be sustainable (due to Decreasing Average Incremental Cost) entry can only be barred if the monopolist produces and prices efficiently i.e. such costs need not be associated with welfare losses.

Expected welfare losses, on the other hand, arise from the existence of sunk cost i.e. outlays that must be made by entrants but not simultaneously by incumbents. It is these sunk costs that render free entry and costless exit difficult and which ultimately cause markets to be non-contestable.

"Long run fixed costs are those costs that are not reduced even in the long run by decreases in output so long as production is not discontinued altogether. But they can be eliminated in the long run by total cessation of production."[3]

"Sunk costs, on the other hand, are those costs that (in some short or intermediate run) cannot be eliminated, even by total cessation of production. As such once committed, sunk costs are no longer a portion of the opportunity cost of production."[4]

A sunk cost can therefore be identified as product

specific capital equipment. There is very little alternative use for such capital if the entrant decides to leave the industry. The roadbed of a rail company is a sunk cost since alternative uses are basically non-existent. A truck or an aircraft operating a particular route is a fixed cost that is not sunk. It is fixed in the sense that a certain amount of equipment is required before even one unit of output is produced. But it is not sunk, since if one particular market were to become unprofitable, operators can always shift towards another market or line of activity altogether (airlines may change routes, or alter operation practices, e.g. operate only charters, luxury service, etc.); trucks could likewise service different markets, transport different commodities, etc.

Where capital is fungible, i.e. where alternative uses prevail, there is no fear of loss on exit. Sunk costs are necessarily fixed (i.e. they are immobile fixed costs) but fixed costs need not be sunk. This nuance is extremely important in contestable markets and it provides the mechanism whereby exit barriers are no longer entry barriers.[8] For not only must it be easy for firms to enter, it must be just as simple for them to exit when the profits decline. In the presence of sunk costs profitable opportunities will not be exploited as

entrants may fear that the discounted profits of entry is less than the cost of entry and the loss of exit. One can thus argue that in reality, sunk costs are so prevalent that nearly all markets are non-contestable.

However, a prevailing theme in this paper is that entrants themselves almost always are established in their own line of operations, and if the market presents a profitable opportunity, then these entrants must only utilise their excess capacity in place (as a result of scope economies in joint product firms), enter using their trademark as a means of hurdling advertising barriers, and exit freely when the profits diminish. Such a scenario is not only more realistic but also more logical in the context of a developed economy. Here sunk cost is still important as a barrier to entry, but the possession of latent capacity in the established entrant's plant ensures that exit is almost costless. This concept of latent capacity will occupy our attention in a subsequent section.

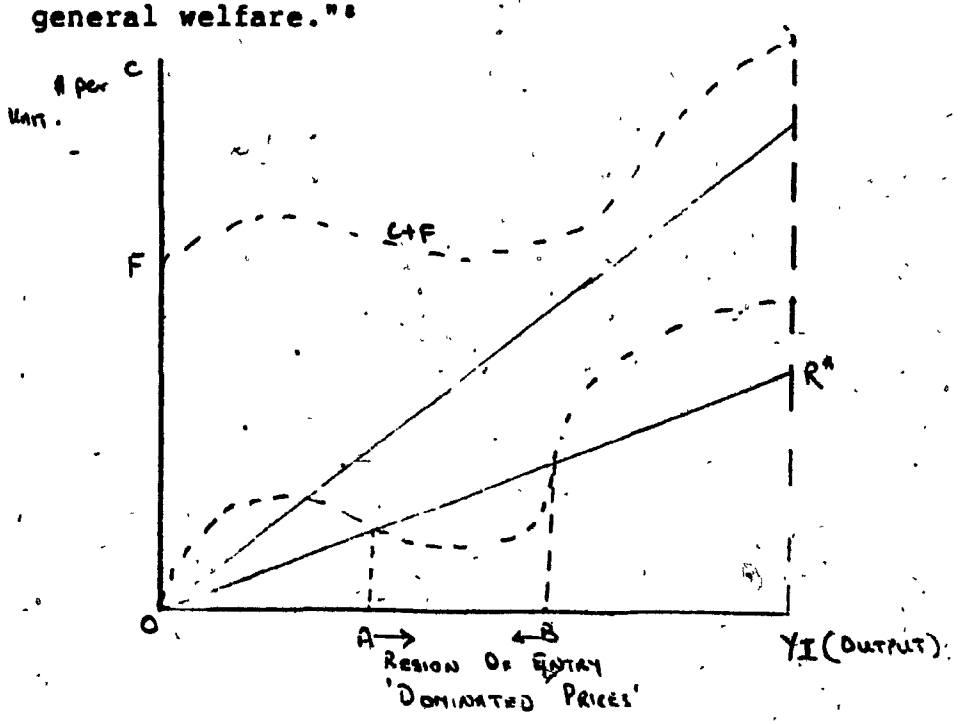
We see then that if there do exist barriers to entry (legal or otherwise) such barriers will of necessity produce an overall welfare loss[6] since incumbents fearing no entry will be induced to price at a level that reduces consumers surplus. Since fixed costs and scale economies

do not, *a priori* introduce barriers (according to Baumol *et al*), policy will have to be stressed on sunk costs characterized by irrevocable expenses incurred by the entrant who has no guarantee of a salvage compensation upon exit.

It is crucial now to ask how fixed costs may preclude entry (via sustainable prices) but yet result in no welfare losses. Sustainable prices, as we have seen in our previous discussions, exist when a particular price-quantity vector emerges at which entrants are not induced to invade a market. Such fixed costs facilitate sustainable prices when the cost function is subadditive in the sense that only a particular quantity of output will permit cost mimimisation. But subadditivity is no guarantee for sustainability -- incumbents must, in addition, not produce in the region of increasing cost, must restrict themselves to least ray average cost and reap all economies of scope and joint production (cost complementarities or trans ray convexities). If these conditions are fulfilled and because fixed costs "offer an advantage to the incumbent only to the extent that his output is greater, and this permits him to spread his costs more thinly than the entrant can"[7] then all advantages of multi-output production will be reaped and our incumbent(s) will only deter entry in a perfectly

contestable market if they price at the point where the revenue hyperplane is just tangent to the point of average cost. He may be protected by fixed costs and his price output vector is sustainable but his is not a blissful equilibrium. He must always operate as if he were constrained by perfect competition in a perfectly contestable market with fixed but no sunk cost.

"The availability of sustainable prices does permit the incumbent to avoid entry. But he can do so only by offering the public the very same benefits that actual competition would otherwise have brought with it. With entry barriers super-normal profits, inefficiencies, cross subsidies, and non-optimal prices all become possible. But in a contestable market which is perfectly consistent with the presence of fixed costs that are not sunk, matters change drastically and government intervention can contribute far less if anything to the general welfare."[8]



In the diagram A - B represents the region of dominated prices where an entrant can produce less than the Y_I (current output level of incumbent) and still be able to reap a profit. The introduction of fixed cost, however, shifts both the cost function and the revenue function upwards. There is no price p* at which an entrant can offer less than Y_I and hope to make a profit. Why is this? Precisely because the introduction of large fixed costs allows for a subadditive cost function (in single product firm it yields economies of scale) so that in the absence of dominated prices, product specific returns to scale and non least ray average cost -- the revenue hyperplane nowhere intersects the cost surface enabling fixed cost via subadditivy to produce optimal output Y* with no threat of wasteful entry.

We now relax some of the rigid conditions required for contestability and modify some key assumptions in an attempt to improve the usefulness of this potentially powerful concept.

5.3 Contestable Markets and Economies of Scope An Amendment

The concept of sustainability in a contestable market is framed in logical time. It is a position of rest with no motive for entrepreneurs to alter their current plans and no incentives for entrants to be aggressive. This equilibrium condition, however, requires the presence of potential entry to ensure that incumbents behave as if they were constrained by an invisible hand.

But such potential entry can only exist if the entrants can establish credibility by possessing actual plant and equipment which may be used to produce the goods of an industry when its price is raised. It is not enough to postulate that entrants can purchase plant of the shelf for every conceivable type of activity in the economy. To do so will implicitly impose the existence of an enormous amount of idle plant waiting in reserve -- to be used as a threat. This is both empirically and theoretically unreasonable, for it means that owners of equipment must continuously stock the latest plant and have it readily available for sale to prospective entrants.

An inconsistency now arises since the entrant must pay a price that will reimburse the owner of capital goods for all his development costs plus the cost incurred in holding in stock the newest technology and destroying outdated ones. If such a market for capital goods did exist and if an entrant could purchase the plant right off the shelf and begin to produce immediately, his total cost would be higher because the price he paid would have to cover the cost of the equipment being idle. No such cost is incurred by the incumbent.

In order for the theory of contestability to be valid, there is a need for this super market of investment goods ready and waiting at a higher price (being idle while the market is in equilibrium) or else requiring that potential entrants themselves possess the latest equipment.

If the latter is the case there will exist the requirement of jelly capital that can be used to produce any particular type of good. If the market for blue widgets were to become profitable, the entrant should be able to transform his capital to blue widget production at no disadvantage in cost or time.

This unrealistic nature of investment as currently proposed limits the value and applicability of the theory of contestability. The assumption of jelly capital and easy and instantaneous access to the latest equipment is totally unrealistic. That investment takes time to be delivered is not at all considered. Since the process of plant acquisition and use is so crucial to establishing the results of contestability, it will be worthwhile to consider the effects of relaxing these rigid, unrealistic and constraining assumptions and approach the problem from a different angle to determine the extent to which a reasonable modification can emerge.

A market may be de facto contestable at time t and there may exist willing entrants but because the cost of holding plant idle in an uncertain world is very high these prospective entrants may at t-1 be reluctant to sink their liquid funds into actual plant and equipment. When the opportunity arises at t there may very well exist an upsurge in the demand for physical plant that could only be delivered at t+1 at which time the incumbents will have had time to re-adjust their prices leaving entrants with plant that yields no profits. The excess capacity at t+1 will so depress prices that what was previously a sustainable vector will with some small deviation on the part of incumbents lead to a tremendous

loss in overall industry profits at that later time.[18]

Knowledge of such a tendency for suicidal entry when investment goods are delivered at a later date may prevent entrants from ordering equipment at t because they expect that the profitable opportunity at t will not continue until t+1. This possibility of loss in holding physical capital anticipating a profitable entry opportunity (obsolence, interest foregone, idle labour), together with the prospective loss associated with the time nature of investment, is enough to seriously undermine the theory of entry as postulated by Baumol et al.

Potential competition is the critical element in contestability, yet the factors making for such competition are usually non existent. Free entry and costless exit with the possibility of resale of plant after the profits have disappeared exist only hypothetically. Such a restriction implies idle equipment at all stages of production -- the owner of capital goods industries must own idle equipment to which the entrant has easy access. Alternatively the entrant must possess idle capacity anticipating a profit in another industry. If there exist one hundred industries and fifty disciplinary potential entrants whose only job

is to look out for opportunities and invade markets with mercenary entrepreneurs then one can immediately observe the implication of this for idle capacity.

These fifty entrants must have access to an unlimited amount of credit to purchase equipment and to compensate workers who are not producing but who must be retained in anticipation of an opportunity (plants need some labour to operate). There is thus idle labour, equipment and anxious entrepreneur-entrants waiting for a profitable opportunity to arise.

If the incumbent were now to increase his price by x% so inviting entry, one must of necessity ask whether the entrant can with this price rise cover all his costs incurred by idly waiting in the wings for an extended period.

Potential entry is not costless and this knowledge is possessed both by the prospective potential entrant and the incumbent. The latter could frequently contrive situations where prices are such as to attract entry but actual entry is impossible since the time and cost involved is such as to make it not worth the entrants while. One must now differentiate between potential competition that is desired but ineffective and potential

competition that is possible and effective. Entrepreneurs may like to enter when the price rises by x% but entry is not possible since they lack plant and equipment and are unwilling to incur costs by holding idle plant and equipment. Here potential competition is ineffective.

The separation between the two types of potential competition serves to illustrate why a seemingly contestable market could possess incumbents pricing above the welfare maximising level. Desired entry can be prevalent but ineffective, while possible potential entry is effective but costly. If the cost of possible potential entry is greater than the profits expected upon entry even after the incumbent raises his price; one will observe a situation of pseudo contestability -- profitable opportunities will go by without entry.

The observations that have so far been outlined do not discredit the concept of contestability. The analysis can escape the problems here identified by relaxing some of the assumptions and bringing it closer to reality. We need a method that will eliminate the costly process of idle entrepreneurs waiting in the wings.

And here one may introduce the notion of "latent excess capacity." Latent excess capacity is said to exist if a firm producing commodity x could with its same plant produce a different good y and profitably market this commodity if its price were to be increased by established producers.

In the presence of such latent excess capacity there no longer exists the need for potential entrants with physical plant which are incurring costs to them every hour. Firms already established in a different line of business with sunk capital in place could with the possession of such capacity produce and market a commodity y with very little transition costs. It is primarily the fungible nature of its fixed equipment that facilitate such a possibility.

One can imagine two firms A and B with A producing x and B producing y (Ax, By). If B increases its price of y then this will invite entry from A if after transition costs are met A can still make a profit at the new price of y. If B subsequently reduces his price to the competitive level then A can always cease production of y, continue in his old activity of producing x (alone) and incur no loss in trying to sell product specific fixed costs. The market will be truly contestable for

the only barrier to entry is the modification and transition costs (retooling, etc.) incurred by A when he modifies his plant. Now there no longer arises the need to have large numbers of potential entrants with idle equipment incurring costs of all sorts and unmeasurable waste in the economy.

5.3.2 Latent Excess Capacity and Economies of Scope

It is easy to extend the notion of latent excess capacity to explain the existence of economies of scope in the multiproduct firm. How does such capacity arise and how is this related to scope economies? This investigation into the sources of scope is a useful exercise for it points to the reasons that facilitate the evolution of multiproduct firms.

The first source one can identify is the logistics of a production process. This is one form of embodied scope economies. The production of beef necessarily entails the production of hides. A mechanical relationship in the case of joint products ensures that a multiproduct firm will emerge.

A second source is technical conditions. Technical economies of scope are those savings of joint production due primarily to the technical relations prevailing in a particular industry. Technical change both embodied and disembodied may present conditions favourable to joint production. Here no automatic production process is involved. The production of one good cheapens the production of the other. Such an extra line necessitates no extra burden upon the firm but rather uses its facilities in place, in production of the first to produce the second.

Other technical results may follow in that producing y, the established firm realises some input that may be used to produce z. Under ordinary circumstances these inputs might be discarded but were it technically feasible to produce z and were a market to exist then the firm would find it in its best interest to undertake that production. An example of this source of scope is perhaps the production of sugar which yields molasses, a necessary input into rum production.

Yet a third source of scope is economic factors. It may be easier to use one plant to produce product x, y and z for a particular market rather than use separate plants and transport these products to the point of

consumption. Additional costs of joint production here are more than compensated by the savings obtained due to transport, storage insurance, etc.

Scope economies facilitate the profitability of producing more than one good. Latent excess capacity on the other hand provides a credible threat to another firm. By producing x the capacity emerges with which to cheaply produce y and z and to market them at prices that are lower than those of independent firms. This latent excess capacity arises from the fixed inputs in place for the production of x. It is worthwhile to note that a single and multiproduct firm can both possess latent excess capacity.

A single product firm possesses such capacity if it knows that it could modify its equipment and profitably produce another firm's product. It may also possess capabilities of economies of scope but since it has not experimented with producing more than one product such knowledge is absent and though it could produce y and z and reap cost savings it is not aware of such opportunites.

The multiproduct firm knows if it possesses scope economies since such economies are realised ex-post.

This process of realisation of its potential is due to its indivisibilities in its fixed costs and also to the public goods nature of both fixed and sunk costs. Indivisible machinery, once purchased for producing x, has capacity that is costlessly (cheaply) available to produce y. This is not all, however. Such a multiproduct firm, in addition to enjoying scope economies, can also possess latent capacity. Its capital in place and its large sunk costs in producing x, y, and z is available to produce a product q if its price were to be increased. It does not currently produce q because the modification costs of doing so plus the extra burden upon its staff and other facilities is greater than the total revenue it could hope to receive.

We can identify some sources of latent excess capacity as resting with 1) managerial use; 2) workers use; 3) plant use; 4) storage, transport, and inventory savings; and research and development.

Managerial Savings: To produce x alone requires a set of managers at each stage of the process. But this does not mean that they are fully utilised. Their full use is obtained only when they expend all energies to exploit every profitable opportunity. The same can be said of workers savings.

Plant Savings: As far as plant savings are concerned, it is assumed here that the plant possesses latent excess capacity both in the embodied and disembodied form. It is embodied if the plant itself must undergo only minor changes to produce another good. It is disembodied if by a different form of organisation, management or a reorganisation in general -- the firm is faced with the possibility of producing more than one good when the market opportunity arises. Opportunities provided both by the physical (technological) characteristics of the plant and by the reorganisation possibilities means that a firm could produce different products at little extra cost and do so in a short space of time.

Such possibilities no doubt are made easy by the existence of research facilities which serve to point out those profitable areas lurking within existing plants. These, then, are some of the sources of scope economies that will induce a firm to become a multiproduct outfit. Latent excess capacity thus does two things -- it facilitates the emergence of scope economies and so induces multiproduction and in addition it provides the needed mechanism for a contestable market. In the presence of such a phenomenon it is now possible for a multiproduct firm to reap all its available opportunities

as well as act as a forceful and credible disciplinarian on other firms.

It is, as was previously pointed out, the latent excess capacity that enables a series of products to cost less when produced together than when produced separately. This same latent excess capacity can now be used as a threat to entry. At any one time there is a vector of products x, y, and z which are produced using the profitable part of such capacity. When the price of q rises, it now becomes feasible to utilise the unprofitable latent excess capacity and to engage in its production.

Likewise it is also possible that firm B which produces p, q and r could, if the price of x rises, use its previously unprofitable capacity and invade A's market. Such possession of this property by many firms will act as a forceful incentive to competitive behaviour and lead to efficiency in resource use.

One could also link the unsustainability of a natural monopoly to the existence of scope economies and indivisibilities in production. Possessing sunk facilities implies substantial excess capacity creating the possibility of producing more than one good. But

these economies of joint production do not necessarily mean that the monopolist possesses an absolute advantage in each of x, y and z. It is possible that an entrant will steal any one of those goods if there are product-specific economies of scale.

The forces of technology, industrial concentration and indivisibilities facilitate a viable multiproduct concern. But it could lose a substantial market to an entrant precisely because of a subsequent change in technology which will allow for product specific economies of scale. The possession of scope economies with its sources -- the latent excess capacity --- in an environment of evolution, adaptation and change is thus no guarantee of freedom from entry. Thus the natural monopoly not only can be unsustainable in the short period (as shown by Panzar & Willig) but over the long period as well.

5.3.3 The multiproduct firm in the absence of scope

So far we have implicitly assumed that the firm produces a vector of outputs due to complementarities of one form or another. We must now consider the firm

producing a set of completely unrelated products. One has in mind a firm producing machinery, industrial chemicals, textiles and food products. This is the megacorporation that finds itself integrated into a wide array of activities.

One assumption here is that the production of machinery does not provide a mechanism that easily facilitates the production of industrial chemicals. These various branches of activities are so separate that the unit under consideration may be seen as comprising a series of single product firms under one umbrella. This view would be mistaken for even where the products are completely unrelated as far as production is concerned, they may not be unrelated when other considerations are taken into account. Here one can identify the advantages of inventory, transport, marketing, distribution, and overall coordination.

Though there is a different plant to produce chemicals from animal feed there would exist a huge sales staff that would be commonly available to all products. In addition, the R&D facilities, storage space, and transport facilities, planning, personnel and risk reduction are all shared by the various products. Calculating the average cost of each good is impossible

but since so many goods share common facilities, a competitor planning to produce only one product in the absence of strong produet specific economies of scale may not be able to compete. The multiproduct firm, therefore, achieves "economies of utilisation of overheads". These utilisation economies cannot be reaped by producing only a single product, for any larger increases in output may so depress price that overall production becomes unremunerative. Had single output production been feasible the firm would have remained with it, avoiding the additional work of overseeing many products. It must be remembered that the firm is a multiproduct firm because it is in its best interest to be so. On the one hand, this decision maximises profits, but it also satisfies the desire of managers for power. In a corporate economy where ownership is divorced from management the production of many goods conforms to the discretionary nature of managers. Now managers can be separated whenever the corporate bureaucracy is threatened with uprisings, mutiny and upheavals and conflicts.

By splitting managers into many divisions with many products it may thus be possible to retain the aggressive character so typical of true entrepreneurial behaviour. Having more ladders to climb, the corporate managers who

would languish in a huge megacorporation would when divided possess a greater incentive to at least become manager of his division. The multiproduct firm not only reduces internal conflict but promotes aggressive behaviour needed to survive.

Another important reason for the emergence of these firms is the risk factor. In bad years a firm can maintain itself if some of its products yield profits while others are a drain on revenues. By producing a set of products that may be unrelated a corporation by spreading risk is much more secure than its single product counterpart. Such an advantage will place it in a better position to secure financing for expansion and for long term growth. While this advantage cannot be subsumed under the term scope economies, it nonetheless provides a rationale for the multiproduct firm that may be quite important in an economy faced with uncertainty.

Of course, there exists too the possibility that even if there are no visible complementarities in production, a firm will produce more than one good if there exists a market for this product and if no other firm seems willing, able or aware of the possibilities. This advantage in a capitalist economy will ostensibly be capitalised quickly by a firm that does possess cost

complementarities or which can produce this good at a lower cost than the incumbent.

We therefore see that even in the case of completely unrelated goods with no economies of scope the firm producing more than one product will find itself in an advantageous position. Whether it can always maintain this advantage is as we have seen dependent on numerous factors.

5.4 Internal Industry Structure

An important claim of the theory of contestability is its provision of an explanation of industry structure. This is really the upheaval in industrial organisation and the scenario is as follows.

One first assumes the existence of potential competition where the threat of entry is sufficient to guarantee outcomes hitherto associated with only a competitive market. It is this environment of potential competition that ultimately leads to the most efficient structure. A particular industry at the beginning is populated by many firms each supplying a portion of the

market. Technology, however, ensures not only that large scales of operation occur but also that large scope economies are reaped. It is postulated that both these economies facilitate the emergence of multiproduct firms serving large segments of the market. Such firms arise only because of the efficiency of the competitive process together with the nature of their technologies (as observed from the cost function).

A multiproduct firm in a contestable market reaping all economies of joint production, scope, scale, cost complementarities (and perhaps even pecuniary economies) is able to achieve results in input use superior to single product firms. Given an exogenous demand curve (curves), potential competition will ensure that only the most efficient configuration arises. At a demand of 100x, 150y, 150z -- where the minimum point on a U shaped average cost surve is achieved at 10x, 15y, 15z (assuming same plant and technology, access to inputs, quality of manpower, and same R&D potential) then the entire industry demand will be determined by 10 firms each with identical sized plants. Structure is then seen to be determined endogenously.

While one will again attack the assumptions which are excessively strong, it is useful to note that the

approach may sometimes approximate reality. By the nature of technologies and scope of operations only four firms may efficiently serve market demand for automobiles, aircraft or other large items. However, if one accepts the argument of Galbraith[11] where these large corporations may significantly influence consumer demand, the mechanical determination of industry structure will invariably alter.

The thesis[12] that it is technology that shapes consumer demand is well taken but to assume demand as entirely exogenous is not. While the former hypothesis may be more applicable in the long run, consumer demand may fluctuate considerably in the short run. Thus consumer demands are shaped by recessions, inflation, changing tastes, size and composition of the labour force and many other factors beside price. A change in consumer demand from 100x, 150y, 150z to 200x, 300y, 300z may cause only four or five firms to remain in the industry or may even allow for single product firms, if product specific economies of scale outweigh the benefits of multi-output production. One must then question the causality of the process. Does the exogenous change in demand cause the alteration of structure, or is such an alteration primarily due to the technology best suited for this new demand. To say that structure in a

contestable market is entirely internally determined while a reasonable hypothesis is clearly not an unequivocal fact.

Leaving aside this troubling integration of demand, we come to the so-called 'existence problem'. Consumer demand at the contestable price may be, say, 600 units per week at price p*. The point of minimum average cost (Least Ray Average Cost in Multiproduct Case) is 110 units per week so that demand is satisfied by 600/110 i.e. 5.6 firms so that there clearly cannot be an equilibrium here. Either five or six firms must serve the market. If demand remains at 600 units per week then either some firms will be forced to produce on the rising portion of their average cost curves (inviting entry in a contestable market) or some consumers will go without the good in question. The notion of a flat bottom average cost where marginal cost reaches a minimum but remains constant over a specified region has been proposed as one solution, and the erection of barriers to entry to prevent wasteful entry is seen as another.

This paper has argued that where "latent excess capacity" exists in established firms, such capacity can be used in a contestable market to produce any residual demand not serviced by the current producers. Thus were

five firms to produce 550x, then another established firm engaged in producing y need only utilise its excess capacity and produce another fifty units of x since a profitable opportunity exists. This is, of course, accepting that the 'new entrant' already has capital in place and that if he decides on an exit his actual loss will be negligible having invested no product specific sunk costs for that purpose.

However this simple explanation of structure being internally determined (via technology) abstracts from the problems of conduct which characterised earlier writings. There is no need for games or responsive action (potential competition and contestability take care of that), price is determined at the point of least ray average cost and demand is satisfied (barring the existence problem). The main catch, however, is that the firms in each industry produce homogeneous goods ..... goods that are indistinguishable from each other. While this may be the case of telecommunications, transport, electricity, and other service industries -- Manufacturing business in the advanced economy is characterised by product differentiation, rendering the assumption that demand is a function of price alone, null and void. Conduct in a multiproduct, multifirm market is then not as simple as one would predict from the theory

of contestability.

This is not to assert that the internal determination of structure has no place in the literature. On the contrary, this writer believes that such a theory is crucial to understanding the dynamics of the industrial economy. However, while contestability and endogenous structure pertain to a longer time horizon, the literature of the older school deals with a shorter period. Contestability provides the clue for industrial concentration over time -- via technology, reaping of economies of large scale and in shaping consumer demand.

Oligopoly theory with its series of games and strategic interactions, on the other hand, is a short term phenomenon. We can infer then, that contestability tells us what will happen in the long period once there is free and unfettered competition in the short period -- the latter being characterised by a disequilibrium process. They are not mutually exclusive and from a global perspective one can see that contestability is just an extention of a series of short period models. The one fits neatly into the other.

In addition to the existence problem of a

sustainable industry structure there also exists a dilemma on the intertemporal allocation of resources. Here what one basically has in mind is a firm operating at time t but with an expectation of demand growth to t+20 years for instance. A rational entrepreneur will construct a plant with the capacity to service demand for the next twenty years providing of course that the present costs of the plant now is less than the discounted profits to be earned by being in a position to serve any increase in demand. Of necessity then and to facilitate planning a certain amount of fixed investment is required and the corresponding excess capacity catering to demand growth entails an outlay which makes the average cost of producing at t+1...19, greater than if an entrant can come in every year with a plant size sufficient to service prevailing demand and then exit with a self-depreciating plant or resell such plant on the secondary market.

This problem occupies chapter thirteen of the new book Contestable Markets and the Theory of Industry Structure but one must question the logic of the argument more so since we are dealing with a contestable market.

An incumbent firm will be foolish to establish a plant with a long life span if it knows that there is the

possibility of entrants in t+1...t19 producing the same product but not burdened by carrying excess capacity. (Note this is real excess capacity as opposed to latent excess capacity.) If an entrant can come in every year and reap a profit then the implication for consumer behaviour is such that demand is only a function of price -- not continuity of service, stability of output supply, consistent performance, etc. In a contestable market where such information is known and where consumers are not concerned with non-price attributes, our incumbent will do well to satisfy consumer demand in just the manner that consumers wish. Indeed, if he establishes a plant with a life of twenty years, he will in effect be producing a product that no one desires. This is an allocative inefficiency.Here consumers are not interested in continuity of service, they are only interested in a lower price which the entrant provides.

On the other hand, in a perfectly contestable market where consumer behaviour is again included, one could see that the inter-temporal misallocation of resources becomes a non issue. Here the entrepreneur knows his customers and their preferences. Utilising such information he can now build his durable plant to take advantage of scale economies in construction as well as fulfill demand in the future. Consumers concerned about

stability and continuity of service will not patronise a fly-by-night entrant who offers a lower price for a short time period; the latter in effect provides an inferior good to his potential customers. Thus the optimal allocation of resources will prevail and even if the incumbent is willing to price at an excess profit level, he is constrained from doing so by entrants already established and reputed for their own performance in their own separate sphere of operations. The unlikely result of the intertemporal failure of the invisible hand must now be closely examined. "The analysis employs only two assumptions -- first, that there are decreasing unit costs in the construction of durable sunk capacity and second, that in a market in which the demand curve is shifting outward with the passage of time (i.e. demand is growing), while it pays to build excess capacity in anticipation of larger future sales, it pays to do so for only some limited period ahead...."[13]

Thus one observes that a plant established in t producing y can produce in t+1 the same volume of output operating at efficient capacity. But in t+1 demand grows by ΔY so that total demand is now

$Y + \Delta Y$ in t+1

This causes the unsustainability of the incumbent at t+1. Why? The reason is provided in the first assumption. We

know that the per unit cost of building the excess capacity ΔY at t+1 is greater than if one builds a separate plant to produce not only ΔY but Y as well i.e. $[\Delta Y + Y]$ Because of economies of scale in construction one finds

$$\frac{K(\Delta Y)}{\Delta Y} > \frac{K(y + \Delta Y)}{y + \Delta Y}$$

At t+1 there thus exists an incumbent who has a plant producing y in both periods pricing at P_1 and P_2 in t and t+1. Demand increases to $Y_1 + \Delta Y_1$ in t+1 and by assumption the cost of producing y_1 with the old plant and ΔY_1 with the new plant is higher than if $(y_1 + \Delta Y_1)$ are produced together. As such, an entrant can produce at t+1 the entire market demand charging a lower price since he experiences scale economies in the larger plant.

But why does the incumbent anticipating a rise in demand not build a larger plant from the beginning. The answer lies in the second assumption for it is assumed that the cost of excess capacity is great, since while the factory is not producing, the entrepreneur will bear maintenance cost, interest payments and other burdens of carrying unused capacity. The conclusion thus is that the firm cannot quote a price at t which will enable it to be sustainable over time. As such it will possess no incentives to undertake long-lived investments eventhough

at time t such investments appear to be in society's interest.

The question to be addressed now is whether this phenomenon represents a failure of the invisible hand. To answer this one must search for the cause of this disturbing anomaly.

The incumbent knows that if he has spare capacity at t to produce $y_1 + \Delta Y_1$ at t+1 he will sustain losses due to carrying costs. He, therefore, invests in plant size y to avoid losses of excess capacity.

If an entrant appears at t+1 with a plant $[Y + \Delta Y]$ and experiences a lower average cost, so driving out the established firm, one can conclude that the incumbent has made a mistake, he should at t have invested in the bigger plant. Had he done so then entry would have been prevented and though he pays the carrying cost of capacity he saves himself from extinction. For when the entrant produces $Y + \Delta Y$ at a lower average cost, the incumbent is eliminated. That the latter did not behave in this manner would suggest that he did not foresee events clearly and investing in plant of capacity Y without considering any future increase in demand results from a lack of foresight and, more importantly, from

disappointed expectations.

The second reason for the result lies in the fact that the incumbent's plant is technically speaking obsolete. It becomes obsolete because developments in demand indicate that a larger plant size is requred. This result is inevitable, for over time although machinery is intact, it is not as productive as newer vintages. Older models, although they still produce, may have to be scrapped. If one allows the incumbent to operate with his old plant, then the economy is denied that right to utilise the best equipment available at lower per unit cost. In a free market, individual entrepreneurs will be forced out and newer ones will emerge. It is certain that private costs are involved but social cost will be minimised only if the market is allowed to dictate events. Far from being a failure of the invisible hand, then, the intertemporal allocation of resources in a contestable market is one perfectly consistent with efficiency in resource use.

We focus now on situations where a market may appear to be contestable but due to certain conditions will not realise the expected results.

5.5 Consumer behaviour in a contestable market

One of the weaker aspects of the theory of contestability is its treatment of demand. There exists no formal integration of the theory of consumer behaviour and the theory of the firm. Prices are taken as given, exogenously determined but at the same time closely tied to the long run point of minimum average cost. There exists no mechanism in the contestable framework to ensure that price is identical to the point of minimum average cost except if one implicitly assumes that the sole determinant of demand is price. Only in this environment will the demand function be

$$D = F(P)$$

and as we have shown, although the downward sloping demand curve persists, the fact that price is the only argument in the demand function ensures that a sustainable vector in a multi-output, multifirm configuration must price at marginal cost. There exists as a result an infinitely elastic demand curve facing each firm. Any upward deviation from average and marginal cost pricing will cause demand to fall entirely to the lower priced entrant up to the region of market saturation at price equals marginal cost equals minimum point of average cost or ray average cost.

In this section we explore consumer demand in a plausible real world setting with a view to subjecting the theory to this very crucial aspect of reality. Goods produced in a sustainable configuration must be sold and the amount of goods produced in the entire economy must be just that amount that consumers wish to buy. Consumers must also possess income with which to buy these goods. Any recession or slackening of demand means that, although firms are producing efficiently and pricing at least cost reaping all scope economies of joint production, if producers cannot sell their output, some factors of production will be released, labour will become abundant and firms may become non-feasible. Whereas at a demand level of 100x the sustainable configuration existed, a fall in demand to 40x may be enough to ensure that perhaps the market is large enough to support at most one single firm. Industry structure as such is not entirely internally determined and exogenous changes in demand could either make a natural monopoly or render the market competitive.

It is not the objective here to investigate the general equilibrium nature of a contestable market, but rather to integrate consumer demand to see if it affects in any way the robustness of the theory.

Consumers are assumed to be expenditure minimisers and utility maximisers. They further demand not goods but the characteristics[14] of goods which is used as an input into their utility functions which produces satisfaction. Two goods may appear to be identical but they will possess different characteristics. Such characteristics we shall see, play an important role in the analysis.

An established incumbent will already have captured a large segment of his customers by differentiating the characteristics of the goods he sells. An automobile producer will provide repair service, continuity, reliability, security and will sell to customers not only a car, but all the ancillary services that go with it. In order to do so, however, a large amount of funds are required to establish these very important differentiations. These are the sunk costs that an entrant with similar plant will have to face. And because they are borne by the entrant but not simultaneously by the incumbent, effectively increasing the average cost of the former, they constitute real barriers to entry.

The incumbent(s) will collect rent, due to their privileged position, until the entrant can either sink

just as much capital into developing his characteristics or invest heavily in plant, in effect burning his bridges so signalling to incumbents that he intends to stay; even though short period price is less than his average cost.

If, however, consumers are motivated only by price and a large fringe consumer group exists, the entrant may face no such disadvantage. Where, on the other hand, service, quality, consistency and the myriad non price factors play a role in shaping consumers' demand, one can infer that there is no mechanism by which the sustainable configuration suggested by contestability will emerge. It may be least costly to use three firms to supply entire demand, but because consumers are conditioned by non price considerations -- an incumbent pricing at a much higher level than potential competitors is still able to deter entry even if he is not technologically efficient.

One could conceivably argue on the other hand that a more efficient entrant could contract with customers before entry. But the problem here is that the cost of contracting with each potential customer may exceed the cost of entry or of establishing sunk cost in reputation. In addition, customers, if they are rational, will be hesitant to contract with an entrant, especially if they

anticipate price warfare as a result of entry or if the industry is changing rapidly.

Only if an entrant were to emerge with a technology that is everywhere superior to the established firms' (whose plant is now obsolete but not depreciated) could the price reduction overcome the characteristic effects. Even here, however, consumers must be sure of the credibility and continuity of the entrant. No entrant can expect to sell electricity at a fraction of incumbents' current price if consumers are not sure that he will continue providing this service on a long term basis. Such continuity of service is but one argument in the demand function, it is a characteristic required for the purchase of durable goods using as an input the product in question. Contestability theory does not provide any answers to structure determination in the presence of non-price determinants of demand.

At this point, it is useful to retrace our steps to place demand in its proper setting. Markets are perfectly contestable, characterized by free entry and costless exit, with the salvage value of invested capital close to its initial purchase price. Current industry structure is inefficient and entrants can thus enter with little or no disadvantages. There even exists an equally

skilled pool of labour for use (reserve army of unemployed) willing to work at a competitive wage, i.e. wages equal marginal productivity in this completely frictionless world. Likewise the rewards to other factors are equalised in this world of perfect information. Time adjustments are negligible and equilibrium (the state of rest) is not a transient phenomenon but rather the norm. In this world production is efficient and only efficient firms emerge. We now introduce consumers into the picture. They are consumers per se and although they may be producers in other sectors, in one market they are considered as consumers. Since information is free and costless, these consumers are aware that the incumbent is inefficient and that entry will lower prices.

However, they do not know how credible the entrant will be. The incumbent may be unwieldy and expensive but unless the entrant is known and established in another line of business, he may be unable to woo the customers of the incumbent with his lower price. Consumers pay the higher price not because of any affinity for the incumbent but because they are familiar with his history and with the quality. Even with basically undifferentiated products, e.g. telecommunications services, an incumbent will have an advantage over the

entrant by his service and reliability. Customers plan expenditure in the belief that certain inputs will be readily available (the purchase of TV sets expecting a steady supply of electricity). If they believe that their long term interests are best served by the more expensive incumbent, even a contestable market with no sunk cost will generate an alternative structure -- unless entrants themselves are willing to undertake such programs.

Cautious consumers will enable the incumbent to earn profits greater than zero ($\pi>0$) and yet survive, while fly by night operators whose entry is imminent may do little harm to the incumbent. A credible entrant, however, already established and firmly reputed in another industry will definitely impose a real threat to the incumbent Why? Simply because this new entrant brings with him all his accumulated expenditures on reputation to bear on the incumbent's market. There exists now efficient use of sunk investments (investment in shaping characteristics) and established entrants face no barriers to entry. Here a coup by one's peers is more dangerous than spontaneous guerilla warfare and is enough to motivate efficient performance. How many companies neglect the activities of minute firms and yet become very worried at any sign of a corporate takeover or a

fellow corporation aggressively marketing one of their products. The fear is further compounded when the entrant can produce with a superior plant or via technical change utilise its own plant to produce the incumbent's line. Here the altered theory (i.e. latent excess capacity) can be quite useful in explaining efficient behaviour.

Nevertheless when subjected to the test of demand, one observes that factors other than price are of crucial importance and it is precisely because of the existence of these rigidities that cannot be assumed away, that a market can be perfectly contestable and yet not be consistent with overall efficiency.

5.6 Potential competition and incumbent power[18]

One of the primary assumptions in the analysis is that of an unresponsive incumbent and an aggressive entrant. This assumption can be easily attacked and is very 'vulnerable' especially where the world of reality is the world of corporations not on the wings, but existing as separate entities with their own long term plans.

Nevertheless it is assumed that incumbents do not react by altering prices upon entry. A time period must elapse for the entrant to produce output, sell at a profit and costlessly exit. This is a very strong assumption but it is argued here that even if incumbents fail to respond, by their very incumbent status, they can successfully deter entry. This is the advantage of being the established firm with a favourable reputation. In addition, the established firm can deter entry in a contestable market even by pricing at monopoly levels and undertaking no visible entry deterring policies.

Let us postulate that the incumbent firm produces 1000 units per week and earns profits on this output. (YI = output of the incumbent.)

i.e. $P_I Y_I > C_I Y_I$

$P_I Y_I - C_I Y_I = \overline{\pi}_I$

$P > MC$

The market is perfectly contestable, i.e., entrants can enter with little fear of retaliatory price cuts and no loss in sunk investment. They enter with a plant size of minimum efficient scale capable of producing 1000 units per week. But now the entrant begins production of his 1000 units at the beginning of week 1. This is to be sold at the end of the week.

i.e. YI = 1000 units

Ye = 1000 units; Ye = output of entrant

The entrant as we have seen cannot respond instantaneously but must produce over the course of this week his 1000 units. The argument here is that if the incumbent keeps his quantity fixed there are now two thousand units of output at the end of that week and price must fall. Depending on the elasticity of demand, the entrant stands to lose if the incumbent behaves in this way. The latter then even if he operates in a perfectly contestable market, by threatening to keep his output level fixed (no price retaliation) is enough to deter any entrant from what might be considered suicidal entry.

If demand is perfectly inelastic they both lose i.e. they cannot cover average cost. Even in a case of mildly inelastic demand, total revenues generated by a fall in price will be insufficient to support both firms. A market may be price sustainable but quantity rigidity and the nature of the demand curve will render a previously contestable market non operational.

It can also be pointed out that the partial equilibrium approach renders the theory null in some cases. Incumbents already are equipped with plant and personnel. However, in a perfectly contestable market

with a multitude of entrants in the wings, the assumption is that they are able to rent capital cheaply and quickly as well as dip into the reserve army of unemployed for their complement of skilled workers. The model thus is not a full employment one and the existence of available factors and unemployed resources means that national income is not as large as it could be. As such the unemployed workers (also on call with the entrant) may not possess income to demand the goods produced by a an entrant, and while firms may produce on their least point of Ray Average Cost, the output they produce may not be sold. The point here is that it is inconsistent to postulate instant purchases in an economy with unemployment. For the theory implicitly assumes unemployment since such unemployment is necessary for entry (as a complement to capital however defined).

Of course if one were to alter the assumption slightly and introduce competition among the few then the reserve army will disappear altogether. These firms already in operation need only re-allocate some of their own factors and there is no need for unemployment in the entire economy, or numerous entrants waiting on the wings with unemployed manpower most of the time.

5.7 Time and contestability

As the reader will recognise the theory is essentially mechanical. It completely neglects changes through time. The main tool of analysis i.e. the average cost or ray average cost is ambiguous. Is one dealing in the short period when capacity is fixed and sunk or the long period when established firms can exit if they wish? In working through the book (contestable markets), one thus does not know if it is short period average cost which must be covered and which determines industry structure or if it is an average cost consistent with changes in technology.

We have also observed that there is the unreasonable restriction that incumbents must hold their prices fixed at pre-entry levels while entrants can invade a market, instantaneously produce, and market a product, hastily exiting before the incumbent has any time to react. There clearly exists then the highly irregular idea that one can use such instantaneous time horizons to understand the determination of industry structure. Sunk costs were seen as the only deterring factor to efficiency in production, pricing and entry deterrence. It will be argued here that a market could be devoid of

sunk costs and yet because of the time element possess within it massive barriers to entry.

The key to understanding the role of time is to accept that production itself consumes time -- in gathering inputs, transport, organisation, actual production, distribution, etc. Only in cases such as electricity and telecommunications may one encounter instant production and instant consumption.

In many other industries, however, production is time consuming. It is not enough to say that the point of least ray average cost is 5x, 4y and 3z without specifying the time frame in which we work -- a day, a week, a month? Suppose by operating a plant efficiently all cost savings are reaped (with no x inefficiency, slack, etc). The rate of production is 5000x, 4000y, 3000z per month and this output is produced not in any fixed day but is a flow produced in varying quantities over time, with demand itself fluctuating over the period of the month.

But the incumbent firm prices where P>MC and at this price even with entry barriers an entrant can make a profit and exit as quickly as he entered with little loss of initial investment. But because he cannot produce

(x,y,z) instantly and because he must take time to organise his inputs and process his intermediate stages and because the point of minimum average cost is (5000x, 4000y, 3000z) per month our entrant must weigh the cost of operating for an entire month in order to reap the cost complementarities. He cannot operate for a day, a week or three weeks since the theory assumes that both incumbent and entrants alike possess the same technology. It is thus argued that even though the market is perfectly contestable i.e. an entrant can enter and exit at the minimum loss before the incumbent can react -- the element of time in production which will constrain any entrant, is enough to act as a powerful and important barrier to entry. Even though sunk costs can be eliminated, real production time cannot, and a market can appear to be perfectly contestable with costless and reversable entry and yet with the inclusion of time the market possesses invisible and important barriers to entry, just as important as the presence of sunk costs. The period of one month is more than enough for an incumbent to realise that he can react if entry occurs and because an entrant is aware of this, the incumbent can continue to produce and price inefficiently even where the market is contestable and sunk costs are absent.

The reader will recognize that this attack arises primarily because the theory consists of the mechanical extension through time of a set of pre-arranged options. This implies the use of the methodology of mechanical time sequentiality. Following this, any deviation in price will invite entry ensuring that the mechanism of the process dicates realized events.[16]

Such behaviour programmed by the mechanical extensions of industrial behaviour indicates that there exist no fears, uncertainty about the future, animal spirits, waves of spontaneous optimism and pessimism and no entrepreneurial temporising. The only argument in the investment function is current profitability. There exist no instance where the past, present and future are qualitatively different, where entrepreneurs learn from the cumulative episodes of their experiences, and where potential investors are able to weigh past knowledge and utilise this cumulative information to project into the future.

In short, there is no use of historical time sequentiality which implicitly incorporates behavior amidst an atmosphere of uncertainty. It is potential competition which moves a configuration from one sustainable equilibrium to another and thus represents

the force moving the process along the path. It implies a logical extension of sustainable equilibria via the forces of potential competition. This is logical time sequentiality where the forces moving a system from one equilibrium to another are stressed. The movement between these two equilibrium points characterised by a mechanical process implies mechanical time analysis.

The endogenous explanation of industry structure utilises these two processes. On the one hand there exist the forces of potential competition which ensure the emergence of efficent monopolies, duo-polies, oligopolies, etc.; and on the other there arises a mechanical relationship between technology, the point of minimum average cost and industry demand.

The former is characteristic of a logical process, the latter a mechanical sequence. But the neglect of the real forces at work in calendar time in an actual business environment characterised by waves of optimism and pessimism, has seriously undermined the logic of what is clearly an elegant theory.

That each moment is different from the next and that investment decisions are so volatile and prone to panic and uncertainty indicates that any theory of industrial

organization must incorporate some historical explanation of behaviour. This is only to ensure that it is people not machines that ultimately dictate the course of events.

NOTES & REFERENCES TO CHAPTER5

1. W. Baumol et. al. "Contestable Markets," loc. cit., p. 349.

2. W.J. Baumol and R. D. Willig. (1981) "Fixed Cost, Sunk Cost, Entry Barriers and Sustainability of Monopoly," Quarterly Journal of Economics, 96, August, pp. 405-431.

3. Ibid., p. 406.

4. Contestable Markets, Loc. cit., p. 280.

5. Eaton, B.C. and Lipsey, R.G. (1981). "Capital Commitment and Entry Equilibrium," Bell Journal of Economics, 12, Autumn 1981, pp. 593-604

6. C.C. Von Weizaker. (1980a) "A Welfare Analysis of Barriers to Entry," The Bell Journal of Economics, 11, No. 2, Autumn, pp. 399-420.

7. Baumol and Willig, op. cit., p. 417.

8. Ibid., p. 420.

9. Ibid., p. 410.

10. Kalecki, Michal (1935). " A Macrodynamic Theory of the Business Cycle," Econometrica, 3, 1935, pp. 327-44

11. J.K. Galbraith. (1968) New Industrial State, Signet Books, New York.

12. Spence A.M. (1983). loc. cit.

13. W.J. Baumol and R.D. Willig. (1981a) "Intertemporal failures of the invisible hand: Theory and implications for international market dominance," Indian Economic Journal, Vol. XIV, nos. 1 & 2.

14. K. Lancaster. (1966) "A New Approach to Consumer Theory," Journal of Political Economy, Vol. 74, No. 2, pp. 132-157.

15. Similar argument presented by William Brock. (1983) "A Review of Contestable Markets," Journal of Political Economy, December, Vol. 91, no. 6, pp. 1055-1067.

16. Valeria Termini. "Logical, Historical and Mechanical Time in Economics, Economic Notes, Vol. 10, No. 3, Sienna, Italy.

CHAPTER 6

6.1 Contestable Markets - The Link to Multiproduct Firms and Policy Implications

The time has now come to link the theory of the multiproduct firm to the notion of contestability (as presented by Baumol et al). Throughout this exercise primarily two market structures have been emphasized, oligopoly and natural monopoly. Little attention has been paid to the multiproduct duopoly or the classical competitive firm. The reason as we shall see is dependent upon the fact that industrial structure depends upon the cost function and it is this phenomenon which, if used to explain one market structure, can be easily extended to study the intermediate cases.

One thus observes the following trend of thought. In the single product case with declining average costs and economies of scale a natural monopoly will emerge as a result of the competitive process. This will be the most efficient industry structure in terms of production, and a single producer will minimise use of resources to produce market demand. The constraint on behaviour here again is the presence of potential competition. In the

absence of sunk costs, this natural monoply will be constrained to earn only normal profits if it hopes to remain in business. Two part tariffs and non-linear pricing schedules are not enough to maintain sustainability.

In the multiproduct natural monopoly case the cost function again is the determinant of structure. But complications arise in the presence of scope, decreasing average incremental cost, product specific returns to scale and demand substitutes. A sustainable natural monopolist is one who deters entry and yet is financially feasible. The weak invisible hand theorem implies that under some constraints on costs and demand this price vector will deter entry.

Where demand is such that one firm alone cannot produce at minimum ray average cost, then a few firms may become the cost minimising configuration. Faced with similar sized plants, with each producing at least ray average cost, reaping all multiproduct economies, the nature of costs will again determine industry structure. Here, however, one encounters the existence problem in that the market may be such that no integral number of firms each producing at minimum ray average cost will satisfy demand. The concept of flat bottom average costs

have been suggested (a region of constant marginal cost) by Baumol et al, while the erection of entry barriers to solve this problem has been postulated by others.[1]

Not withstanding this difficulty it has been proposed by the authors that a multiproduct firm in a contestable market can only be sustainable if it produces and prices at a level that minimise costs and offers no incentive for entry. The structures just described arise from the nature of contestability.

A monopolist may produce efficiently but is constrained from pricing at the monopoly level due to the existence of potential entry. This threat may occur not only in one product line but in each and every product that he produces. If he raises the price of x, an entrant will invade his market price at a lower level and exit when the incumbent readjusts.

Similarly there is an absence of games and strategic behaviour in an oligopolistic environment. Such interaction is eliminated once entrants can exit costlessly. However, product differentiation, brand loyalty, advertising, and other such characteristics so typical of oligopoly are assumed away. What the authors have in mind one can imagine, are industries like

telecommunications, transport, electricity and such services which are basically undifferentiated.

The powers of the market mechanism with potential credible entry together with an elastic demand curve facing each seller are the invisible hand of the theory. Where the market is contestable, incumbents must produce their vector of outputs at minimum average cost, leaving no profitable entry plan available. Pricing must also be at the level of marginal cost. A P>MC invites entry by an entrepreneur who can produce at least ray average cost and charge a P*<P and due to the elasticity of demand earn a positive profit. A price lower than marginal cost means that an entrant can restrict output to the point where his MC equals this lower price and successfully enter. The threat of entry is enough to ensure pricing at the optimal levels except in the case of natural monopoly where one is constrained by the elasticity of demand and by the need to break even.

A market that is perfectly contestable will yield results that are obtained in the perfectly competitive case. But now there is no longer the need for large numbers of sellers to act as a pre-requisite to efficiency. The presence of potential competition is enough to guarantee efficient behaviour (analagous to

that in perfect competition) in natural monopolies, and to reduce the strategic interactions so typical of oligopolistic models. One may mistake such a sweeping conclusion as a policy of laissez-faire but close examination will reveal that it is not. It will be seen that in many instances even contestable markets fail to provide a social optimum thus necessitating some form of regulation.

In the first case where the cost function is subadditive but the industry structure is unsustainable in that no entry preventing price-output configuration exists then there is the need for some protection from the forces of the market.

Again concerning the existence problem with the absence of flat bottomed average cost curves, no one sustainable structure will emerge and this may require some form of entry barriers to protect firms already in the industry from wasteful entry. Aside from these inherent difficulties the theory of public finance has outlined market failures which though they are important are not explicitly considered by the theory and thus provides no policy directive except that of government intervention.

These include externalities in consumption and production where one party gains or loses from the activity of the other with no mechanism to enforce a reduction of detrimental activities or an increase of desirable services. Such external behaviour will require some form of outside interference in the absence of any possibility of bargaining between actors and the affected parties. Even if such bargaining were possible, the equity consideration will be violated if affected parties do not gain any of the increase in output that the delinquent activity supplies.[2]

In addition to external effects the problem of public goods with the impossibility of exclusion provide yet another area of intervention. Where a public good is costlessly available to one person once it has already been supplied to another and where the fortunate individual is under no compulsion to pay there emerges the classic free rider problem.[3] This is one problem with which the theory cannot deal since no one will be willing to pay for a good that they can freely obtain by understating their preferences.

But perhaps most crucial of all is the inability to deal with social goals in consumption, i.e. those ideals that for better or worse enter into the social welfare

function of any society. Such goals may be the provision of universal telephone services or the provision of postal services for the entire country or electricity to each and every house from the remote settlements of the backwoods to the most populated areas.

Such undertakings entail a fair amount of loss as the unremunerative services must be subsidised by those more profitable or from general tax revenues. Contestable markets will allow production only on the profitable routes. On these routes there will be efficiency in resource use and potential entry will enforce efficient behaviour. However, residents in remote communities will have to pay a price much higher than they may be able to afford. The requirement of contestability whilst facilitating efficiency fails totally in welfare considerations or social policy. For example, the basic question of how to ensure universal access whilst gaining the benefits of market competition in telecommunications in a country like Canada cannot be answered.

Nevertheless Baumol et al have argued that the case of natural monopoly as a market failure can easily be discarded, with only an unsustainable natural monopoly posing any problems. Recall that such unsustainability

emerges primarily from satisfying total demand for allits products in the presence of product specific economies of scale. To overcome this problem the idea of imposing the same market supply requirements upon the entrants is seen as a possible solution. Furthermore to prevent monopoly pricing under protection the concept of the quasi permanence of price reduction is put forward where incumbents if they lower their price (to drive an entrant away) will be forced by statute to keep it at this level for an extended period.

Some forms of market failure such as inefficiency in the supply of products (producing within the production possibility frontier) cross subsidisation, monopoly pricing and inefficient structure are all discussed[4] but the basic problems of combining efficiency and equity or distributing the rewards of efficiency so that there is no need for equity considerations is omitted. Such an omission, it might be observed, must not be a discredit to the work. These wider issues will always challenge the minds of researchers. However, even though the market process can work to reduce waste in inputs and maximise output, the problems here discussed must always be addressed to determine the extent to which the proposed theory provides a basis for solving crucial and pressing problems.

NOTES & REFERENCES TO CHAPTER6

1. Ten Raa, T. (1980), " A Theory of Value and Industry Structure", Ph.D. dessertation, New York University, 1980 as referred to by Baumol et al Contestable Market op. cit.

2. Ronald Coase. (1960) "The Problems of Social Cost," Journal of Law and Economics, Vol. 3 (October), pp. 1-44.

3. For fuller discussions of this problem see Robin Boadway, Public Sector Economics, Chapter 4, Winthrop Publishers, Inc., Cambridge, Mass.

4. Contestable Markets, loc. cit., chapter 12.

CONCLUSIONS

Most of the ideas put forward by the new approach to industrial organisation have not yet been fully explored. However, any attempt at investigating the characteristics of the multiproduct firm will undoubtedly benefit from the armoury of powerful tools and analytical concepts which have emerged. Terms like Ray Average Cost, economies of scope, average incremental costs, product specific scale economies and trans ray convexity provide the framework within which one can identify the special characteristics of the multiproduct firm.

It is these ideas, too, which will serve as the link in bridging that gap which currently exists between industrial organisation and microeconomic theory and which will help focus attention on the possible benefits which accompany size. But while a multiproduct firm may be technically superior to its competitors, it does not mean that such efficiency will translate itself into lower prices.

Only in the contestable market will such a firm be forced to become internally efficient in addition to pricing at marginal cost (the case of natural monopoly

excepted). But while the theory of the multiproduct firm is analytically acceptable, that of the contestable market is not so easy to accept. The stringent requirements of information and absence of sunk cost are enough to invalidate the theory as a fair approximation of reality. This is not however to diminish the usefulness of the approach but only to indicate its limitations.

Where the approach does excel, however, is in its use as a theortical construct capable of yielding results not formerly provided within the analytical paradigms provided by the theories of perfect competition , oligopoly, duopoly and monopoly. Some of the newer results are those concerning the unsustainability of the natural monopoly, the internal explanation of industry structure and the requirement of ease of exit. The latter result provides the conditions where efficiency no longer requires the existence of large numbers of firms in a particular market. Other results such as the importance of scope vs. scale (though controversial) and the subadditive cost function could only have emerged by explicitly considering the mechanics of the multiproduct firm.

In opting for theoretical rigour and analytical

elegance, however, much has been sacrificed in abstracting greatly from reality. The reactions and games so typical of older oligopoly theories are now assumed away by introducing a homogenous product. There is thus no "scope" for demand manipulation by advertising or loyalty effects. Incumbent power becomes negligible since the entrepreneur is imprisoned by the forces of technical change and potential competition. As an individual he has no ability to protect his position. Thus the theory framed primarily in mechanical sequences avoids situations based in historical time where the past, present and future are qualitatively different and where economic agents must continually modify their behaviour by learning from past experiences. Thus entrepreneurs are not allowed to hold excess capacity anticipating a boom for this means that a higher-than-usual average cost forces him out of the market. Uncertainty, fears, disappointed plans, animal spirits, expectations, unequal access to capital markets, the use of the political system and indeed those myriad factors which enter into the decision process of firms becomes lost in the hypothetical world of a contestable market.

The actual economy is characterised by the immobility of factors of production, sunk costs,

information gaps and a degree of inertia. Firms operating in a concentrated environment are able to protect themselves by building barriers to entry and practising markup pricing. While some industries such as the telecommunications and computer industries may appear to approximate (in some measure) contestable markets, many other industries are not so placed. Thus the theory itself cannot claim to be universal in practice, though in fact it can often be used as a theoretical construct which will determine if any changes are an improvement, in relation to what would have happened had the environment been contestable.

As far as the theory of the multiproduct firm is concerned, one can only conclude that it provides a wealth of instruments capable of further assisting industrial theory in particular and micro theory in general. Now that the tools are available the path has been cleared so that theory as a whole can begin to be based on economies characterised by large firms engaged in many operations. Such theory will not only provide ideas on the forces that tend towards concentration, but also on the possibility of efficiency gains or losses in the absence of contestable markets. In addition, the mechanism by which these gains in technical efficiency are transmitted to final consumers may prove to be more

challenging than the gains in efficiency itself.

Ultimately the success of the theory will depend on the extent to which it can lead to a re-appraisal of orthodox theories of the firm. The objective is of course, to shift emphasis away from the single product unit and to focus on the multiproduct megacorporations which dominate the advanced industrial economies.

BIBLIOGRAPHY

Arrow, K.J., Chenery, H.B., Minhas, B., and Solow, R.M. (1961). "Capital Labor Substitution and Economic Efficiency," The Review of Economics and Statistics, 43, No. 3, August 1961, pp. 225-250.

Averch, H. and Johnson, L. (1962). "Behavior of the Firm under Regulatory Constraint," American Economic Review, 52, No. 5, December 1962, pp. 1052-1063.

Asimakopulos, A. (1978). An Introduction to Economic Theory - Microeconomics, Oxford University Press, 1978.

Bailey, E.E. (1981). "Contestability and the Design of Regulatory and Antitrust Policy," American Economic Review, 71, No. 2, May 1981, pp. 178-183.

Bain, J.S. (1954). "Economies of Scale, Concentration and Entry," American Economic Review, 44, March 1954, pp. 15-39.

Bain, J.S. (1956). Barriers to New Competition, Cambridge, Mass., Harvard University Press, 1956.

Baseman, K.C. (1981). "Cross Subsidies and Open Entry in Regulated Markets," in G. Fromm, ed., Studies in Public Regulation, Cambridge, Mass., MIT University Press, 1981, pp. 329-360.

Baumol, W.J. et al. (1982). Contestable Markets and the Theory of Industry Structure, Harcourt Brace Jovanovich, New York (NY), 1982.

Baumol, W.J. et al. (1983). "Contestable Markets: An Uprising in the Theory of Industry Structure: Reply," American Economic Review, June 1983, 73, pp. 491-96.

Baumol, W.J. (1977). "On the Proper Cost Test for Natural Monopoly in a Multiproduct Industry," American Economic Review, 67, December 1977, pp. 809-822.

Baumol, W.J. (1979a). "Quasi Optimality: The price we must pay for a price system," Journal of

Political Economy, 87, No. 3, 1979, pp. 578-599.

Baumol, William J. (1982). "Contestable Markets: An Uprising in the Theory of Industry Structure," American Economic Review, March 1982, 72, pp. 1-15.

Baumol, W.J. and Bradford, D.E. (1970). "Optimal Departures from Marginal Cost Pricing," American Economic Review, 60, June 1970, pp. 265-283.

Baumol, W.J. and Fischer, D. (1978). "Cost Minimising Number of Firms and Determination of Industry Structure," Quarterly Journal of Economics, 92, August 1978, pp. 439-467.

Baumol, W.J. and Willig, R.D. (1981a). "Intertemporal Failures of the Invisible Hand: Theory and Implications for International Market Dominance," Indian Economic Journal, 1981.

Baumol, W.J. and Willig, R.D. (1981b). "Fixed Cost, Sunk Cost, Entry Barriers and Sustainability of Monopoly," Quarterly Journal of Economics, 95, August 1981, pp. 405-431.

Baumol, W.J. (1977). Economic Theory and Operations Analysis, 4th ed., Prentice Hall Int. Inc., London, 1977.

Boadway, Robin (1979). Public Sector Economics, Winthrop Publishers, Inc., Cambridge, Mass., 1979.

Brock, William A. (1983). "Contestable Markets and the Theory of Industry Structure: A Review Article," Journal of Political Economy, December 1983, 91, pp. 1055-66.

Carlson, Sune. (1937). A Study in the Pure Theory of Production, London.

Caves, R.E. and Porter, M.E. (1977). "From Entry Barriers to Mobility Barriers," Quarterly Journal of Economics, 91, May 1977, pp. 241-261.

Caves, D.W., Christensen, L.K. and Tretheway, M.W. (1980). "Flexible Cost Functions for Multiproduct Firms," Review of Economics and Statistics, 62, No. 3, August 1980, pp.

477-491.

Clemens, E. (1950). "Price Discrimination and the Multiproduct Firm," Review of Economic Studies, 19, No. 48, 1950-1951, pp. 1-11.

Coase, R. (1960). "The Problem of Social Cost," Journal of Land and Economics, Vol. 3 (October 1960), pp. 1-44.

Coase, R. (1970). "The theory of Public Utility Pricing and its Application", Bell Journal of Economics 1, Spring 1970, pp.113-129.

Cohen, K., Cyert, R., Theory of the Firm - Resource Allocation in a Market Economy, 2nd edition, Prentice Hall, Inc., Englewood Cliffs, N.J., 1975.

Demsetz, H. (1968). "Why Regulate Utilities?" Journal of Law and Economics, 11, April, 1968, pp. 55-65.

Demsetz, H. (1978). "Joint Supply and Price Determination," Journal of Law and Economics, 16, No. 2, October 1973, pp. 389-405.

Demsetz, H. (1982). "Barriers to Entry," American Economic Review, 72, March 1982, pp. 47-57.

Dewsnup, E.R. (1914). "Railway Rate Making," American Economic Review, 4 (Suppl.), 1914, pp. 81-100.

Dupuit, Jules (1844). "On the measurement of the utility of public works," in Arrow and Scitovsky (eds.), Readings in Welfare Economics (American Economic Association, Vol. XIII, Richard D. Irwin Inc., Homewood, Ill.), 1969.

Dixit, A. (1980). "The Role of Investment in Entry Deterrence," The Economic Journal, 90, March 1980, pp. 95-106.

Eaton, B.C. and Lipsey, R.G. (1981). "Capital Commitment and Entry Equilibrium," Bell Journal of Economics, 12, Autumn 1981, pp. 593-604.

Faulhaber, G. (1975b). "Cross Subsidization: Pricing in Public Enterprise," American Economic Review, 65, December 1975, pp. 966-977.

Friedman, J.W. (1977). Oligopoly and the Theory of

Games, Amsterdam: North Holland, 1977.

Fuss, M.A. and McFadden, D. (eds.) (1978). Production Economics: A Dual Approach to Theory and Applications, Amsterdam and New York: North holland, 1978.

Fuss, M.A. and Waverman, L. (1981), "The Regulation of Telecommunications in Canada," Economic Council of Canada, February 1981.

Friedlaender, A. and Bailey, E. (1982). "Market Structure and Multiproduct Industries," Journal of Economic Literature, 20 September 1982, pp. 1024-1048.

Fromm, Gary (eds.) (1981), Studies in Public Regulation, London, MIT Press.

Galbraith, J.K. (1968) The New Industrial State, Signet Books, New York.

Gold, B. (1981). "Changing Perspectives on Size, Scale, and Returns. An Interpretive Survey," Journal of Economic Literature, 19, March 1981, pp. 5-33.

Green, Christopher (1980). Canadian Industrial Organization and Policy, Toronto: McGraw-Hill Ryerson, 1980.

Hicks, J.R. (1935). "Annual Survey of Economic Theory - Monopoly," Econometrica, 3, 1935, pp. 1-20.

Hicks, J.R. (1939). Value and Capital - An Inquiry into some fundamental Principles of Economic Theory, Oxford University Press, London.

Hotelling, H. (1938). "The General Welfare in Relation to Problems of Taxation and Railway and Utility Rates," Econometrica, 6, 1938, pp. 242-269.

Joskow, P.L. and Noll, R. (1981). "Regulation in Theory and Practice: An Overviews," in G. Fromm (ed.), Studies in Public Regulation, Cambridge, Mass.: MIT University Press, 1981, pp. 1-65.

Kalecki, Michal (1935). "A Macrodynamic Theory of the Business Cycle," Econometrica, 3, 1935, pp. 327-44.

Koutsoyiannis, A. (1979). Modern Micro Economics, London, The Macmillan Press, 1979.

Kousoyiannis, A. (1982). Non Price-Decisions - the Firm in a Modern Context. London: Macmillan, 1982.

Lancaster, K. (1968). Mathematical Economics, New York: Macmillan, 1968.

Lancaster, K. (1966). "A New Approach to Consumer Theory," Journal of Political Economy, Vol. 74 (1966), pp. 132-157.

Liebenstein, Harvey (1966). "Allocative Efficiency vs. X-efficiency," American Economic Review, 56, June 1966, pp. 392-415.

Littlechild (1970). "Marginal Cost Pricing with Joint Cost," Economic Journal, 80, June 1970, jpp. 223-235.

Marshall, A. (1925). Principles of Economics, London and New York: Macmillan, 1925.

Miller, R.L. (1982). Intermediate Microeconomics, New York: McGraw Hill, 1982.

McClosky, D. (1982). Applied Theory of Price, New York and London: Macmillan, 1982.

Modigliani, F. (1958). "New Developments on the Oligopoly Front," Journal of Political Economy, 66, June 1958, pp. 215-232.

Panzar, J.C. (1981). "Cross Subsidies and Open Entry in Regulated Markets - Comment," in G. Fromm, ed., Studies in Public Regulation, Cambridge, Mass., MIT University Press, 1981, pp. 365-369.

Panzar, J.C. and Willig, R.D. (1977a). "Free Entry and the Sustainability of Natural Monopoly," Bell Journal of Economics, 8, Spring 1977, pp. 1-22.

Panzar, J.C. and Willig, R.D. (1977b). "Economies of Scale in Multi-Output Production," Quarterly Journal of Economics, 91, August 1977, pp. 481-494.

Panzar, J.C. and Willig, R.D. (1981). "Economies of Scope," American Econoic Review, 71, No. 2, May 1981, pp. 268-272.

Pfouts, R. (1961). "Theory of Cost and Production in the Multiproduct Firm," Econometrica, 29 (4), October 1961, pp. 650-659.

Pfouts, R. (1964). "Multiproduct firms vs. Single Product Firms: The Theory of Costs and Production, Metroeconomica, vol. 16, pp. 51-66.

Ramsey, F.P. (1927). "A Contribution to the Theory of Taxation," Economic Journal, 37, March 1927, pp. 47-61.

Robertson, D.H. (1957) Lectures on Economic Principles, London, Staples Press, Vol. 1.

Salop, S.C., (ed.) (1979). "Strategic Entry Deterrence," American Economic Review, 69, May 1979, pp. 335-338.

Samuelson, P.A. (1963). Foundations of Economic Analysis, 2nd ed., Cambridge, Mass.: Harvard University Press, 1963.

Samuelson, P.A. (1964). "The Pure Theory of Public Expenditure," Review of Economic Studies, 36, No. 4, November 1964, pp. 387-389.

Schumpeter, J.A. (1947). Capitalism, Socialism and Democracy, London: Allen and Unwin, 1947.

Scherer, F.M. (1980). Industrial Market Structure and Economic Performance, 2nd ed., Chicago: Rand McNally, 1980.

Schmalensee, R. (1978). "Entry Deterrence in the RTE Cereal Industry," Bell Journal of Economics, 9, Autumn 1978, pp. 305-327.

Sharkey, W.W. (1982). The Theory of Natural Monopoly, Cambridge, U.K.: Cambridge University Press, 1982.

Sharkey, W.W. and Telser, L.G. (1978). "Supportable Cost Functions for the Multiproduct Firm," Journal of Economic Theory, 18, No. 1, June 1978, pp. 23-37.

Schwartz, Marius and Reynolds, Robert J. (1983). "Contestable Markets: An Uprising in the Theory of Industry Structure: Comment," American Economic Review, June 1983, 73, pp. 488-90.

Sheppard, W.G. (1984). "Contestability vs. Competition," American Economic Review, September 1984, 24, pp. 572-587.

Spence, A.M. (1983). "Contestable Markets and the Theory of Industry Structure: A Review Article," Journal of Economic Literature, September 1983, 21, pp. 981-90.

Stackelburg Heinrich Von (1952). The Theory of the Market Economy, William Hodge, London, Translated by A.T. Peacock

Stigler, G. (1957). "Perfect Competition, Historically Contemplated," Journal of Political Economy, 65, No. 1, February 1957, pp. 1-17.

Stiglitz, J. (1981). "Potential Competition May Lower Welfare," American Economic Review, 71, May 1981, pp. 184-189.

Taussig, F.W. (1913). "Railway Rates and Joint Costs," Quarterly Journal of Economics, 27, No. 4, August 1913, pp. 692-694.

Termini, V. "Logical, Historical and Mechanical Time in Economics," Economic Notes, Vol. 10, No. 3, Sienna, Italy.

Turvey, Ralph. (1982) Book Reviews, Studies in Public Regulation, Economic Journal, Vol. 92, pp. 731-732.

Uzawa, H. (1964). "Duality Principles in the Theory of Cost and Production," International Economic Review, 5, May 1964, pp. 216-220.

Varian, H.R. (1978). Micro Economic Analysis, New York: W.W. Norton & Co., Inc., 1978.

Von Weizsacker, C.C. (1980). "A Welfare Analysis of Barriers to Entry," The Bell Journal of Economics, 11, No. 2, Autumn 1980, pp. 399-420.

Weldon, J.C. (1948). "The Multiproduct Firm," Canadian Journal of Economics and Political Science, vol. 14, May 1948, pp.176-190.

Willig, R.D. (1979a). "Multiproduct Technology and Market Structure," American Economic Review, 69, No. 2, May 1979, pp. 346-351.